Graduate Theses, Dissertations, and Problem Reports

1999

Explaining age-related problem -solving differences on concept identification problems as a function of problem content, strategies, and stereotypes

Dave Buck
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Buck, Dave, "Explaining age-related problem -solving differences on concept identification problems as a function of problem content, strategies, and stereotypes" (1999). *Graduate Theses, Dissertations, and Problem Reports*. 1053.

<https://researchrepository.wvu.edu/etd/1053>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

Explaining Age-related Problem-solving Differences on Concept Identification
Problems as a Function of Problem Content, Strategies, and Stereotypes

Dave Buck

Dissertation submitted to the
College of Arts and Sciences
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Developmental Psychology

Hayne W. Reese, Ph.D., Chair
Stanley H. Cohen, Ph.D
JoNell Strough
Barry A. Edelstein, Ph.D
Richard T. Walls, Ph.D

Department of Psychology

Morgantown, West Virginia
1999

Keywords: Problem-solving, aging, concept identification, stereotypes, strategies

Abstract

Explaining Age-related Problem-solving Differences on Concept Identification Problems as a Function of Problem Content, Strategies, and Stereotypes

Dave Buck

The purposes of this study were to: (a) determine whether older adults would use logic-based strategies on a concept identification task when the stereotyped content of the task was changed to “experience free” (b) determine whether older adults would use logic-based strategies when the stereotyped content was left intact after completing an “experience free” version of the concept identification task, and (c) explore relations of potential causal mechanisms of strategy use for older adults such as problem interpretation or past experience. Thirty-two older adults ($M = 70.0$ years) and 32 younger adults ($M = 21.2$) solved two versions of the Pictures task (adapted from the Accountant problem, Laipple 1991). Multiway frequency analyses and follow-up nonparametric tests revealed younger adults used significantly more logic-based strategies than older adults for the version with stereotype content (i.e., the goal the task was to find the surgeon). No significant age differences were revealed when the content was “experience free” (i.e., find the person born in March). There were also no significant age differences on the version with stereotyped content when the participants first completed the “experience-free” version. Contrary to results of other studies, interpretation of the problem was not correlated with strategy use. Participants’ references to stereotypes were correlated with type of strategies used but this relation was greater for younger adults than older adults. Older adults are posited to use experience-based strategies when the task contents allow for such a strategy and use this strategy more automatically, whereas, younger adults are posited to use logic-based strategies more often regardless of task content. Construct validity of concept identification tasks and other cognitive tests may be suspect if the task content is not examined for possible stereotypes.

Table of Contents

Introduction	1
Status of Current Literature	2
Traditional Definitions of Problem	3
Deficit-Difference Distinction	5
Problem Solving and Older Adults	6
Competence-Performance Distinction	7
An Alternative Theoretical Framework for Explaining Age Differences	8
Task Interpretation	9
Generalization Versus Interpretation	10
Protocol Analysis	11
Intervention Studies	12
Statement of the Problem	15
Method	16
Participants	16
Materials	17
Consent form	17
Demographics questionnaire	17
Thinking aloud instructions and practice problem	17
Pictures task	17
Surgeon version	19
Birthday version	19
Post-task questionnaire	19
General intelligence indexes	20
Procedures	20
Strategy Scoring	22
Standard classification of strategies	22
Consistent constraint seeking	23
Reference to stereotype	24
Reference to experience	24
Task interpretation	25
Hypotheses	26
Surgeon Version versus Birthday Version	26
Receiving Birthday Version First	27
Results	27
Demographic Variables	27
Age and Sex Differences in Health, Vocabulary, and Digit Symbol Scores	29
Overall Strategies	30
Planned Comparisons	31
Rationale for planned comparison approach	31
Strategy type as dichotomous dependent variable	31
Description of planned-comparison analyses	32
First Planned Comparison	33

Predominant constraint seeking	33
Consistent constraint seeking	34
Second Planned Comparison	35
Predominant constraint seeking	35
Consistent constraint seeking	36
Third Planned Comparison	37
Predominant constraint seeking	37
Consistent constraint seeking	38
Omnibus Analyses	39
Birthday First versus Surgeon First	40
Predominant constraint seeking	40
Consistent constraint seeking	40
Surgeon First versus Surgeon Second	41
Predominant constraint seeking	41
Consistent constraint seeking	41
Birthday First versus Birthday Second	41
Predominant constraint seeking	41
Consistent constraint seeking	41
Omnibus-F Test	41
Latin-square Design	43
Interpreting the position effect	45
Interaction involving sex	46
Reference to Stereotype	47
Reference to Experience	48
Reference to Stimuli	48
Task Interpretation	50
Photoset effects	50
Discussion	53
Problem Content as a Causal Factor	55
Construct Validity	56
Problem Interpretation	57
Limitations	58
Directions and Implications for Future Research	59
General Comments	60
References	62
Appendix A	66
Appendix B	67
Appendix C	68
Appendix D	70
Appendix E	72
Appendix F	73
Appendix G	74
Appendix H	77
Appendix I	81

Appendix J	83
Vita	86

List of Tables

Table 1	<u>Definitions of “Problem”</u>	4
Table 2	<u>Four Orders of Version and Stimulus Set Presentation</u>	22
Table 3	<u>Demographics and WAIS-3 Subtests: Means and Standard Deviations</u>	29
Table 4	<u>Frequency of Strategies Used by Older and Younger Adults for Each Order of Presentation</u>	30
Table 5	<u>Analysis of Variance for Factorial and Latin Square Designs: Predominant Constraint Seeking</u>	42
Table 6	<u>Analysis of Variance for Factorial and Latin Square Designs: Consistent Constraint Seeking</u>	43
Table 7	<u>Frequency of Constraint Seeking Displayed in Factorial and Latin-square Designs</u> .	45
Table 8	<u>Frequency of Questions Using Specific Stimuli to Eliminate Pictures</u>	49
Table 9	<u>Photoset and Version for Participants Who Asked about Hair Length on Second Question</u>	52
Table 10	<u>Best Fit Models and G^2 for Logit Analyses of the Omnibus Tests</u>	82

Introduction

[The quality of problem-solving research among the aged] is strained because, although older adults have been shown to respond differently from young adults in many problem situations, few significant steps have been taken toward understanding the nature of the differences. (Reese & Rodeheaver, 1985, pp. 495-496)

Although researchers have reliably found that older adults' problem-solving performance differs from that of younger adults, attempts to explain these research findings have been limited. The aim of this study is twofold: (a) to present a framework and methodology useful for determining why older adults' problem solving differs from younger adults, and (b) to test whether older adults employ strategies used by younger adults if a task is changed such that past experience or general knowledge no longer appears useful.

In the following sections of this introduction, traditional definitions of problem are provided and problem-solving research involving older adults is summarized. Prior to summarizing this research, a distinction is made between performance on problems that have descriptively different solutions and problems that have normatively different solutions (i.e., difference versus deficit distinction). This distinction affects how results of problem solving studies comparing older and younger adults are interpreted. Next, the concern with inferring older adults' actual abilities by evaluating their performance (i.e., competence-performance distinction) is discussed. An empirical method for determining the true problem-solving abilities of older adults on a given problem is then described.

Status of Current Literature

Problem solving is a rich and complex topic that has been examined within a variety of theoretical frameworks. The study of developmental changes in problem solving across adulthood has, however, not been couched in any theoretical framework (Reese & Rodeheaver, 1985). Rather, research findings on problem solving and aging have been mainly descriptive, quasi-experimental, and correlational. The overall findings clearly demonstrate that older adults perform differently on most problems compared to younger adults (Arenberg, 1974; Denney, 1982, 1990; Reese & Rodeheaver, 1985).

The viewpoint that problem-solving performance lies on a normative continuum of bad to good, worse to better, or incorrect to correct has limited the range of hypotheses produced to explain the age differences. Although some problems do indeed fall along such a continuum, many do not.

As summarized in following sections, researchers have found many areas in which older adults do not perform the same as younger adults. In the normative view of problem solving, these age-related differences are understood to be deficits (i.e., older adults are said to perform worse than younger adults). It is argued in this paper that the normative view has hindered researchers from examining the true cause of age-related problem-solving differences, at least for some types of problems. A more promising theoretical framework includes the assumptions that: (a) older adults' strategies and solutions evolve and adapt over time due to changes in everyday problem solving and (b) problem-solving behavior in laboratory problems is generalized from everyday life situations especially when the content of laboratory problems is similar to everyday life situations.

Traditional Definitions of Problem

The problem-solving literature contains very few definitions of problem. Most researchers seem to assume that problem is a clearly understood concept and needs no explicit definition. Some researchers have provided definitions (Duncker, 1945; Reitman, 1964; Skinner, 1966). Others have paraphrased these definitions but in doing so, have created slightly different definitions (Chase & Bjarnadottir, 1992; Garnham & Oakhill, 1994; Reese, 1994; Reese & Rodeheaver, 1985; Scheerer, 1967). Table 1 lists a few definitions found in the literature.

The definitions in Table 1 can be loosely categorized into two types: (a) a problem occurs when an organism is not in a “goal state”, and (b) a problem is when an organism’s response does not produce the established reinforcer. The first type of definition is useful when problem interpretation is being addressed. Problem interpretation and goal state are discussed in the following section. The second type of definition is consistent with a behavioral perspective. This definition is also consistent with the theoretical framework adopted in the present study to explain age-related problem-solving differences. That is, an environmental change occurs as people age which causes different behaviors (viz., problem-solving strategies) to be used to produce solutions (viz., reinforcers). The problems used by researchers evoke these different behaviors from older adults in the laboratory. In the experimental setting, these behaviors lead to different solutions. These different solutions, as well as the strategies being employed, are interpreted by many researchers as incorrect.

Table 1

Definitions of “Problem”

Author	A problem is . . .
Duncker (1945)	when a living organism has a goal, but does not know how this goal is to be reached. (p. 1)
Newell, Shaw, & Simon (1960) ^a	a situation where a problem solver desires some outcome or state of affairs that he does not immediately know how to attain. (p. 257)
Reitman (1964)	defined by such a vector [A, B, $\Rightarrow$] taken together with its associated requirement. (p. 289) Let A and B stand for initial and terminal states or objects, respectively, and let the sign $\Rightarrow$ denote a process, program, or sequence of operations. (p. 284)
Reitman (1965)	when [a system] has or has been given a description of something but does not yet have anything that satisfies the description. (p. 126)
Skinner (1966)	a question for which there is at the moment no answer. (p. 225)
Skinner (1966)	[when there is] no response available which satisfies a given set of contingencies. Or [when] competing responses may be evoked--among them emotional changes which weaken the very response upon which reinforcement is contingent or destroy the power of the reinforcer. Or [when] the contingencies may be satisfied only by a sequence or chain of responses, early members of which are too remote from the terminal reinforcers to be strongly affected by it until conditioned reinforcers have been set up. (p. 226)
Skinner (1966)	a specific set of contingencies of reinforcement for which a response of appropriate topography is to be found (p. 249)
Skinner (1966)	when a strongly reinforced response has deferred aversive consequences or when immediate aversive consequences conflict with reinforcers (p. 250)
Scheerer (1967) ^b	when the goal that is sought is not directly attainable by the performance of a simple act available in the animal's repertory (p. 27)
Reese & Rodeheaver (1985)	transformation of a situation from some initial state to some other state (p. 474)

Table 1 continued next page.

Table 1 continued . . .

Garnham & Oakhill (1994)	a problem has three crucial elements: (1) a starting state, (2) a goal state, and (3) a set of processes (usually called <u>operators</u>) that can transform one state into another. (p. 200)
Reese (1994)	how to get to a goal (p. 199)
Reese (1994)	[when] (a) the organism is in a state that is different from a desired or desirable state called the <u>goal state</u> ; and (b) coming to be in the goal state is being blocked. (p. 201)

Note: ^aQuote appeared in Reitman (1965, p. 129). Original source: Report on a general problem solving program. International Conference on Information Processing. Paris: UNESCO.

^b Scheerer (1967) attributed this definition to Thorndike (1898) but this definition does not appear in Thorndike's work.

Before summarizing research on problem solving and older adults, special attention must be given to some of the wording within this literature. The wording used by most problem-solving researchers hinders alternative interpretations as to why developmental changes occur. The following section is intended to expose the bias and provide tools for seeing beyond it.

Deficit-Difference Distinction

Several cognitive functions, including some areas of problem solving, can be viewed to be on a normative continuum from bad to good, worse to better, or incorrect to correct. Memory, speed of processing, reading, math, following directions, and discrimination learning, are just a few areas that fall along such a continuum. For example, being able to remember more items on a grocery list is better than being able to remember fewer items.

Some problem solving also falls along a normative continuum. Problems in which a specific end state has been expressed to a solver, and the solver has interpreted this end state the

way the experimenter intended, can result in correct or incorrect solutions. That is, if the designated end state was not achieved, the solution was incorrect. In such cases, there is good and bad problem solving. Not all problem solving, however, should be viewed as being on a normative continuum. Rather, solutions to some problems may simply be descriptively different (not incorrect versus correct). For example, “calling a repair man” and “checking the fuse box” are different solutions for what to do if a refrigerator was broken. Neither is incorrect, nor is one more correct than the other.

All too often this distinction between normative and descriptive differences in problem solving has not been made by researchers. Rather, the terms ‘correct’ and ‘incorrect’ are widely used for categorizing solutions within the problem-solving literature without explanation or rationale-- as though all problem solving fell along a normative continuum.

The deficit-difference distinction and related concerns were pointed out several decades ago by other researchers in this area (e.g., Labouvie-Vief, 1977). Unfortunately, no adequate theoretical framework has been suggested in which aging changes would be best understood as predictable differences rather than deficits.

Problem Solving and Older Adults

Researchers have generally shown that older adults are less likely than younger adults to solve a wide range of traditional problems in the intended ways (e.g., Piagetian tasks, concept learning tasks, verbal and nonverbal search tasks, verbal reasoning tasks, visual and spatial tasks, and creativity tests; see Denney, 1990 for a thorough review).

For problems such as linear syllogisms and other deductive reasoning problems, studies have reliably demonstrated what performance changes occurred as people aged, and also which

cognitive resource factors reliably affected performance (Light, Zelinski, & Moore, 1982; Salthouse, Legg, Palmon, & Mitchell, 1990; Wright, 1981). For example, linear syllogisms require solvers to recall information and use working memory to obtain a solution. Cognitive abilities such as recall and working memory decline as people age, thus affecting older adults' performance on problems that require these abilities.

Not all problems impose memory demands or working memory requirements. For example, tasks with few steps where all necessary stimuli are always available to the solver do not require significant amounts of memory resources. These types of problems are the focus of this study.

Competence-Performance Distinction

Based on the results of problem-solving research on older adults alone, it is unclear whether older adults lose certain abilities which, in turn, prevents them from solving problems the way younger adults do, or older adults can solve these problems the way younger adults do but something about the specific problem context does not promote such behavior. This is known as the competence-performance distinction.

Reese and Rodeheaver (1985) defined competence as what people can do under ideal conditions and performance as what people do under actual conditions. The issue of competence versus performance is always relevant in the study of cognitive aging, but it is of particular interest in problem-solving research. Finding that older adults did not perform the same as younger adults does not necessarily justify concluding that they are unable to perform the same as younger adults. It is possible that older adults performed differently simply because they used different strategies. If so, research is needed to determine whether older adults can perform like

younger adults under different conditions, and to determine what in particular made the older adults perform differently in that context.

An Alternative Theoretical Framework for Explaining Age Differences

Age differences in problem solving can be understood from the principle of generalization. That is, strategies and other problem-solving behavior are generalized responses to the novel problem-solving situation in the laboratory. Problem-solving research involves presenting participants with problems that are rarely identical to those experienced in everyday life, with the exception of some of the studies examining everyday problem solving (e.g., see Sinnott, 1989). A safe assumption, however, is that the problems or the instructions or some parts of the context used in the majority of studies likely have some feature that is similar to what participants have experienced before. This is likely for two reasons. One is that researchers usually construct their problems to be representative or correlated with some life event. The second reason is that these problems tend to evoke some kind of strategies and solutions. If the participant has never been exposed to the specific problem context, responses to the experimental problem are likely generalized from problems experienced in their everyday life. Participants' strategies and subsequent solutions are thus understood by discovering the problems from which the current responses have been generalized. These problems or the strategies used to solve them, may be different for younger and older adults.

As people grow older, the nature of some problems and ramifications for problem solving change. For example Berg, Klaczynski, Calderone, & Strough (1994) asked older and younger adults to report recent personal problems. Older adults' most salient problems were almost all family and health related. Younger adults' problems, however, were widely varied with no

dominant categories standing out. Other factors may also change as people age, response time may become less important, novel solutions might not be required as often, and past experiences may be helpful and require less effort for solving problems more often than when people were younger. Changes in environmental contexts that vary with age promote (and sometimes require) the development of different strategies and solutions for everyday problems. If laboratory problems are vague or ill defined, one would expect participants' everyday problem solving strategies to generalize to the laboratory test situation. With everyday problem solving differing across age groups, different generalized responses are expected.

Task Interpretation

Some researchers have examined performance differences based on participants' interpretations of the problems (for reviews see Berg et al., 1994; Laipple, 1991). Laipple (1991), for example, coded participants' verbal protocols for interpretations of problem tasks and found that older adults interpreted problems as being about personal experience while younger adults interpreted problems as being abstract and that solutions differed predictably as a function of problem interpretation. Interpretation of a problem has been considered a possible causal mechanism influencing problem-solving behavior (Berg, et al., 1994; Gilhooly, 1996; Laipple, 1991).

Problem interpretation refers to the participant's mental reconstruction of the problem statement into a problem form. Reitman (1965) classified three components of a problem form: (1) initial state, (2) end state, and (3) a set of processes that may be used to produce the end state from the initial state. The experimenter may not always completely specify all components in the problem statement. If all three components are completely specified in the problem statement, the

problem is well-defined. When one or more components of the problem statement are vague or unspecified, the problem is ill-defined (Gilhooly, 1996). Ill-defined problems are more likely to produce various interpretations across solvers.

Laipple (1991) defined interpretation as "the participant's active construction of a task into a problem form" (p. 2). Operationally, interpretations were statements that participants made regarding "what the problem is", "what information they used to solve the problem", and statements they made while solving the problem and reconstructing the task (Laipple, 1991, p. 87). These statements were used to infer what participants were thinking.

From a behavioral perspective, participants' statements may be viewed at least two ways:

- (a) They represent or reflect covert behaviors (cognitions) that serve as establishing operations or setting occasions that make the problem context an evocative stimulus for certain problem-solving behaviors. That is, they reflect unobserved behaviors that affect overt problem solving.
- Or (b) they reflect covert behaviors that are evoked by the problem context and do not affect overt problem-solving behaviors. That is, the problem context directly evokes both the covert "interpretations" and the overt problem-solving behaviors with no causal link between these products.

Generalization Versus Interpretation

Generalization does not involve mediating variables. For example, imagine a situation in which for some person, in the presence of a crystal glass, picking up this glass and drinking resulted in tasting soda. Assume this antecedent-response-consequence relation was reliable. Now this person is in the presence of a ceramic mug she has never seen before. If she picks up this mug and drinks, this is referred to as a generalized response. The new antecedent shared

some physical feature with the original antecedent (e.g., they were both handheld containers) and the new stimulus evoked a generalized response. In this example, the reason that the generalized response (drinking from mug) occurred is explained or understood by knowing that person's learning history with the crystal glass.

Interpretation is similar to generalization but involves a hypothetical mediating step. Interpretations would not be used to explain the behavior from the above example (i.e., no one would say the person interpreted the mug to be the crystal glass). Rather, interpretation would be used to explain more complex situations. For example, a person may be asked to "help" his sister with her homework. That person may then think "help means to do the homework for her" or "help means to give feedback after she attempts the homework." These covert behaviors or thoughts are interpretations and are hypothesized to determine subsequent actions. The existence and causal effects of these interpretations can be inferred from introspected verbal reports of the individual or from observing the type of overt helping behavior emitted.

Protocol Analysis

Explicit examination of the learning history of the problem-solver in contexts relevant to the current problem has been ignored. This is not surprising because specific learning history of this nature is likely to be complex, extensive, and difficult to directly observe.

Because direct observation is not feasible, alternative methods for inferring learning history must be considered. Asking participants to think aloud while solving problems is one such method. Thinking aloud protocols are data that can be directly relevant to understanding the contingencies that now cause current problem-solving contexts to evoke the responses that they do. Thinking aloud protocols are often used to discern which strategies or cognitive processes

participants are using, what information is attended to, what inferences participants are making, and number of steps involved (for a review see Ericsson & Simon, 1996). When participants are asked to think aloud, they often mention past experiences regarding the problem context, which parts of the problem are affecting their behavior, rules that they are following, and why they are following these rules. If the functions of problem-solving behavior can be gleaned from the content of verbal protocols, future researchers could attempt to examine the learning history of other participants directly by studying or manipulating specific life events.

As mentioned previously, older adults may have learned different strategies because of environmental changes they experience as they age. Statements used to infer interpretations may also provide insights as to what environmental changes occurred for older adults that cause changes in problem solving. Thus, recording participants' interpretations is one step toward understanding problem-solving behavior. The next steps are analyzing these interpretations for hypotheses about past history and rules governing behavior, testing these hypotheses by experimental manipulation, and finally, using these findings to guide future research aimed at direct observation or manipulation of the development of these learned behaviors.

Intervention Studies

Specific interventions such as instructions, training, modeling, or changing the problem content may be used to see if they alter the evocative effects of the problem context. In all intervention studies, participants in the experimental group are presented with a treatment condition before receiving the problem under investigation. Instruction as an intervention may involve presentations of verbal descriptions of the strategies that lead to the intended solution. Training involves presenting participants with similar problems and feedback on their

performances. Modeling involves the experimenter using the strategies that result in the intended solution on similar problems.

Instructions, specific training, and modeling, are all interventions that lead to “success” for older adults in a range of problem-solving tasks (Denney, 1990). That is, older adults are more likely to emit the intended strategies after such interventions. These techniques, however, do not address whether older adults learned the problem-solving behaviors solely from the intervention or whether the intervention served as a cue for responding differently (i.e., prompted them to use another strategy they already knew). Two questions are left unanswered using these types of interventions. One is, why do older adults solve problems differently? (i.e., are differences due to cognitive declines or due to differences in learning history). The other question is, are the intended strategies part of their learning history? (i.e., can they demonstrate competence?).

The key to testing this question is whether a modified version of a task can produce different problem-solving behavior without instruction, training, or modeling. One way to conduct such a test is to: (a) find a problem where older adults respond differently than intended; (b) hypothesize which stimuli within the problem evoked the unintended strategy; (c) modify the problem structure to remove these stimuli; (d) present this modified problem to the participant first; and (e) then present the original form of the problem. If the participants now use the intended strategies for both versions of the problem, then these strategies were likely part of their learning history and not learned from the problem presentation. If a modified version of a problem leads to an intended solution, then competence has been determined. Additionally, the specific changes to the problem that cause differential responding may provide insight into

possible events in the learning histories of older adults that explain why problems are solved in certain ways and not others.

A task in which different participants tend to report varying interpretations (and emit different solutions) can be used to infer participants' learning histories. At the same time, elements of this task (e.g., task content) can then be systematically varied in attempts to evoke the intended solution. If exhaustive variations are made and the intended solution is never emitted by participants, the solution may not be part of the participants' learning histories. Participants may simply not be able to emit the response without instructions. If, however, a variation causes participants to emit the intended solution, the discriminative stimuli that cause differences in performance can be determined. The Accountant task used by Laipple (1991) is one such problem that evokes various interpretations across participants.

The Accountant problem is an adaption of the poisoned-meal problem described by Reese (1994, p. 242), which was an adaption of the original version by Arenberg (1968). Laipple (1991), Reese, Puckett, and Cohen (1990), Jurden, Reese, Cohen, and Puckett (1992), and others have used these types of problems to uncover concept-identification strategies of participants. In Laipple's version, participants are shown eight cartoon faces and told one is an accountant. The characteristics of these faces vary across several dimensions (wearing a hat or not wearing a hat; messy hair or straight hair or bald; mustache or no mustache; smile or no smile). Participants are told they may ask the experimenter three yes or no questions and then must state who is the accountant. One common strategy is called item checking (i.e., using each question to eliminate one face). Another common strategy is constraint seeking (i.e., eliminating exactly half of the faces with each question). Constraint seeking will always result in the accountant being found.

Laipple (1991) found that the Accountant task evoked item-checking responses in older adults more often than younger adults. He also found that older adults were more likely to refer to personal experience while solving the problem compared to younger adults. According to Laipple's (1991) study, 13 out of 16 older adults made "extensive" reference to experience during the Accountant problem compared to 5 out of 16 younger adults. This task is one that is likely to be evoking strategies generalized from problems in which past experience or stereotypes about occupations was useful for older adults.

Statement of the Problem

The purpose of the present study was to determine whether older adults can emit strategies used by younger adults and to find out why older adults emit the strategies they do. To do so, the content of a problem known to produce age differences must be changed so that past experience does not appear to be helpful for solving the problem.

The function of older adults' item checking strategy was unclear in Laipple's (1991) study and other studies because the dimensions in the Accountant problem were not related to past experiences with accountants in any obvious way. Therefore, Laipple's (1991) Accountant problem was modified for the present study and renamed the Pictures task. Actual photos of people of different ages and gender were used instead of cartoon drawings. Two versions were created. In one, the participants were instructed to figure out who is the surgeon or who was born in March. The rationale for switching from Accountant to surgeon is reflected by the difficulty most people have when trying to solve the following riddle:

A boy and his father were in an automobile accident. The father was killed instantly. The boy was alive but badly injured, and was rushed to the hospital and prepared for the

emergency surgery. But the surgeon, on seeing the boy, said, “I can’t operate on that boy. He’s my son.” How can that be? (Mook, 1996, p. 392)

Presumably, due to one’s past experience regarding surgeons, the common response for this riddle is that the surgeon is a male. Thus, the surgeon must be the father, hence the difficulty because the father died. For the same reason that this riddle does not evoke the response “the surgeon is the boy’s mother,” the surgeon version in the current study was expected to evoke some responses that were clearly based on past experience or stereotypes.

A second version was created to be “experience free”. Rather than look for a surgeon, participants were told to look for a person born in March. Presumably, no one believes that people in March have specific physical characteristics in common (i.e., no stereotypes exist about people born in March). Past experience should not be helpful in this version. Thus, older adults were hypothesized to switch to logic-based strategies (e.g., constraint seeking) like those used by younger adults. Thinking aloud protocols and responses to experimenter’s probes were hoped to demonstrate the function of participants’ questioning strategies.

One potential problem is that some participants might not verbalize their rationale behind their strategies, assuming there was one. Even in this case, in the stereotype version pair checking or item checking strategies that tend to eliminate the photos of the two older men and then the two younger men would support the suspicion that participants were responding according to stereotypes and past experience. Such performance trends would suggest that the responses were experienced-based even in the absence of any corroborating verbal protocols.

Method

Participants

Sixty-four participants were recruited in a South Dakota University town (32 older adults; 32 younger adults). Half of each age group was female. Younger participants were college students and ranged from 20 to 25 years of age ($M = 21.2$, $SD = 1.4$). Thirty of these younger participants received course credit for participating. Older participants ranged from 60 to 80 years of age ($M = 70.0$, $SD = 5.9$). All older participants were recruited from existing subject pools and through word of mouth techniques. All were community dwelling as apposed to assisted living and none reported taking medication that had memory side effects. No other screening criteria were used. No participants were paid to participate. Average ages of males and females did not differ significantly for both age groups, $F_s < 1$.

Materials

Consent form. All participants signed a consent form (see Appendix A). The main points of the consent form were presented orally. Participants were asked if they had any questions and were reminded that participation was voluntary.

Demographics questionnaire. A demographic questionnaire (see Appendix B) was presented after the consent form was signed. Participants reported their date of birth, years of education, perceived health, chronic medical problems, and prescribed medications, and reported whether they had received training or taken a course in formal logic.

Thinking aloud instructions and practice problem. Participants were read instructions about thinking aloud (see Appendix C). Participants practiced thinking aloud while solving a version of the balance scale problem (shown in Appendix C) adapted by Reese (1994) from Klahr & Siegler (1978).

Pictures task. Participants were shown one of two 8 1/2 x 11 pages with photographs of

eight persons' faces displayed as a 4 x 2 matrix (see Appendix D). Three classes of stimuli served as dimensions that were counterbalanced across pictures in the photosets: Age, sex, and background behind the person (white or brick). Age and sex were chosen because these dimensions were likely to be related to stereotypes of who would be a surgeon. The ages of younger people in the pictures ranged from 30 to 35 years old (bottom four pictures). Ages of the older people were between 47 and 55 (top four pictures). The first and third rows of photographs were male, and the second and fourth rows were female. Background was chosen as the third dimension because it was neutral with regard to surgeon stereotypes. The order of type of background (white or brick) alternated down each row of the photoset; white - brick, brick - white, white - brick, and brick - white.

These three counterbalanced dimensions were used to promote constraint seeking. That is, if participants asked about sex, age, or background, the question could potentially eliminate half of the photographs depending on how it was asked. Constraint seeking could occur without using these three dimensions; for example, participants could use position of pictures (i.e., left versus right; top versus bottom).

The initial photosets included some salient characteristics that were not counterbalanced (e.g., facial hair, smiling, glasses, and ties). These characteristics, because they were odd in number, might have led some participants to partial constraint seek rather than constraint seek. Thus, some photographs were retaken to obtain a more uniform photoset. Finding enough people and scheduling when pictures could be taken turned out to be very difficult and time consuming because of the desired physical characteristics and background constraints. Thus, rather than retake all the photographs, pictures were taken of enough people until photographs were

approximately uniform. The revised photosets used in the present study contained men with no facial hair, all wearing ties. Facial expression was more uniform across these photos as well. In one of the revised photosets, all of the photographs contained were of people wearing glasses (glasses photoset) and in the other photoset, no people wore glasses (no glasses photoset).

Glasses was indeed a salient characteristic for participants. Twenty-one participants (12 younger adults and 9 older adults) mentioned wanting to use glasses to narrow down the possibilities until they noticed that everyone was wearing glasses in that photoset. Reference to glasses occurred twice more often in the surgeon version ($N = 14$) than the birthday version ($N = 7$). In retrospect, glasses should have been used as a counterbalanced dimension.

Surgeon version. Participants were told that one of the people in the photoset was a surgeon, the experimenter knew who the surgeon was, the participant was asked to figure out who the surgeon was, and the participant was told to ask three questions before giving the answer. The three questions could be anything as long as the experimenter could answer them yes or no. No photograph was designated as the surgeon a priori. Rather, the experimenter answered the participants' yes/no questions in such a way that the participant could not determine the surgeon until three questions had been asked.

Birthday version. Participants were told that one of the people in the photoset was born in March. The rest of the instructions and procedures were identical to the surgeon version of the problem.

Post-task questionnaire. In a post-task questionnaire, shown in Appendix F, participants were asked what information they used to solve the problems and to rate how similar the problems were to others they had encountered. Participants were then asked to administer the

problems to the experimenter. This portion of the task was audio-taped.

General intelligence indexes. The Wechsler Adult Intelligence Scale - Third Edition (WAIS-III) digit symbol subtest (Wechsler, 1997) was used as a complementary measure of nonverbal fluid intelligence. The WAIS-III vocabulary subtest (Wechsler, 1997) was used as a measure of crystallized intelligence. The digit symbol test was given first. Standard scoring methods were used for both tests.

Procedures

The consent form was explained to participants and any questions were answered. After consent forms were signed, demographic information was obtained. Next, participants were read the instructions for thinking aloud. They were then given the practice problem and asked to think aloud. After the practice problem, participants were given the two versions of the picture task, one at a time. This part of the session was tape recorded. The order in which the surgeon and birthday versions was presented was counterbalanced across age and sex groups.

Participants were read the problem script (see Appendix E), in which the gist was: One person in the picture is a surgeon (or has a birthday in March) and they must find out who. The experimenter knows who the surgeon (or person with a March birthday) is and the participant may ask three yes or no questions before he or she makes a choice. The experimenter answered participants' questions so that the fewest pictures were eliminated. For example, if a participant's first question was whether the surgeon was in the top row the experimenter would answer no. This question eliminated two pictures and left six. If the participant's first question was whether the surgeon was in the top three rows, the experimenter would answer yes. This question also eliminated two pictures and left six. Although questions that eliminated exactly half of the

pictures could be answered yes or no and have the same effect, the experimenter always answered ‘no’ to these questions to maintain consistency.

If participants did not constraint seek on the first picture task, they were told that their choice of the target picture was incorrect (after all three questions were asked). If they used constraint seeking, they were told that their choice was correct. Telling them they were correct was moot at this point, of course, because constraint seeking eliminates seven pictures from being the target. All participants were told that their choice of the target picture on the second version was correct, regardless of which strategy they used. This was done to avoid participants feeling that they failed.

After both Picture task versions were completed, participants were given the post task questionnaire and then the WAIS-III subtests (digit symbol and vocabulary). Finally, participants were debriefed about the purposes of the study, the Pictures task was explained, and questions about the study were answered.

The design of the study was a repeated-measures Latin-square design: 2 (age group) x 2 (sex) x 2 (serial position) x 2 (version; Surgeon versus Birthday) x 2 (photoset; Glasses versus No glasses), with the last three variables within-subjects in a Latin square. The order in which the photosets were presented was counterbalanced across age groups, sex, and versions of the picture task. Half of the participants who received the surgeon version first received the glasses photoset for this task and half received the no glasses photoset for this task. This was done to analyze whether differences in photosets had an effect on participants’ responses independent of version of task. Table 2 shows the four orders in which task versions and photo sets were presented to participants.

Table 2

Four Orders of Version and Stimulus Set Presentation

Version presented first/Photoset	Version presented second/Photoset
1. Birthday/Glasses	Surgeon/No glasses
2. Birthday/No glasses	Surgeon/Glasses
3. Surgeon/Glasses	Birthday/No glasses
4. Surgeon/No glasses	Birthday/Glasses

Strategy Scoring

Participants' strategies were coded using two methods: (a) the standard classification method created by Reese, et al., (1990) for the Poison Meals problem and (b) consistent constraint seeking. The standard classification method was used so the present data could be compared to past research like Laipple's (1991) Accountant problem data. This method, however, does not differentiate between participants who constraint seek on some questions and participants who constraint seek on all questions. The later participants are more likely "truly" constraint seeking where as the former participants, may or may not be purposely using a constraint seeking strategy. Thus, categorizing consistent constraint seeking was used as an alternative method of measuring strategies. Statistical analyses were conducted twice using data produced from both methods of classifying strategies as dependent variables. Both methods are described below.

Standard classification of strategies. Participants asked three questions per task. Each question was scored according to how many pictures it eliminated (0, 1, 2, 3, or 4) and strategies were classified using a standard method based on these numbers. This method was applied to the

data obtained in the current study. Description of the strategy classification procedure can be found in Appendix G.

One strategy is to ask a question that involves selecting a characteristic that divides the choices by exactly half and asking whether the surgeon has that characteristic (e.g., Is the surgeon a male? Is the surgeon an old adult? Is the surgeon on the left side of the matrix?). This strategy has been labeled ‘focused constraint seeking’ (Laipple, 1991) or focusing (Reese, 1994), but is referred to as constraint seeking in the present study. Other strategies include constraining the search by eliminating not quite half of the people (partial or incomplete constraint seeking), eliminating two people with each question (pair or dimension checking), and eliminating only one person with each question (item checking). Asking a question that does not eliminate anybody is either a pseudo-strategy or an unidentifiable strategy.

Once each question was scored, the predominant or overall strategy was identified as the strategy that occurred at least two out of three times. When the strategy type of all three questions differed, the predominant strategy was classified as the strategy of the first question because this was the participant’s first reaction. Four participants (4 older and 1 younger) had three different types of strategies for one version and one older adult had three different types of strategies for both versions.

When the traditional classification method was used, overall strategy categorized as constraint seeking is referred to as predominant constraint seeking in this dissertation. This classification method was used in this dissertation so data could be compared to previous studies. However, a second, more conservative strategy classification was also used in this dissertation.

Consistent constraint seeking. The traditional method of classifying strategies may

inappropriately categorize a participant's strategy as constraint seeking when constraint seeking may not be the one intended. A participant who only sometimes eliminated half of the items may have been intending to simply eliminate clusters of items. That is, this participant may have purposely attempted to eliminate more than one item at a time without actually trying to eliminate exactly half of the items with every question but done so by accident. A participant who provides a questioning pattern in which participants eliminated four, two, and then one photo in questions 1, 2, and 3, is much more likely to have purposely used such a strategy. This strategy was coded as consistent constraint seeking. This strategy always leads to discovery of the target in this task. Using this classification method, all other strategies were coded as not consistent constraint seeking. Separate data analyses were conducted using predominant and consistent constraint seeking as dependent variables to determine whether different classification of strategies produced different results.

Reference to stereotype. Thinking aloud protocols for 63 participants (the tape recorder malfunctioned during one participant's session) were analyzed and coded for statements that referred to stereotypes about surgeons. For example; "The third guy on the left was too young"; "She would appear to be a nurse because of the way she's dressed"; "[Surgeons] spend so much time in the operating room . . . so a lot of them are fair skinned." Reference to stereotype was coded as a dichotomous variable (i.e., participants either referred to stereotypes or did not).

Reference to experience. Statements were also coded for reference to past experience. For example; "[I used] what I've encountered through certain experiences of what [surgeons] look like"; "I know a surgeon and he's a very placid person"; "Well I know there are a lot of women surgeons . . . when I had open heart surgery, the lady who did mine . . ." Reference to

experience was also coded as a dichotomous variable.

Task interpretation. Task interpretation was determined by using a method similar to the one outlined by Laipple (1991). Laipple measured interpretation two ways. First, concurrent interpretation was determined by examining participants' responses to the problems. Second, Laipple assessed the reconstruction of the task for interpretation by examining participant's responses to the instruction, "Pretend you are the experimenter and administer the task back to me but use your own words."

Only concurrent interpretation was coded in the present study. Reconstructed task interpretation protocols turned out to be too vague to infer interpretation using any criteria. Statements from participants' verbal protocols were examined with their actual questions (strategies) and experimenter's statements removed. The strategies were removed to avoid directly influencing the coding of interpretations. That is, participants' questions (strategy) to the Pictures task were removed from thinking aloud protocols for this assessment. Experimenter's statements were also removed. The remaining statements and participants' responses to the question, "what information were you using to solve these problems?" were used to score concurrent task interpretation. Recall that these 'interpretations' have been hypothesized to determine strategies (Laipple, 1991) and used to predict strategy use. Thus, strategies must not be allowed to influence the coding of these interpretations.

All remaining statements that referred to the goal of the problem were coded as either experienced-based, pragmatic, or logic-based interpretations. Experience-based interpretations contained statements about the goal or purpose being related to past experience or knowledge about surgeons or knowing what a surgeon looks like. Logic-based interpretations contained

statements about the goal or purpose being to narrow the possibilities with each question, or appearance of the people in photos and past experience being irrelevant for solving the problem. Pragmatic interpretations contained both types of statements (experience-based and logic-based). Appendix J shows the classification method used and examples of protocols.

Hypotheses

Age differences in other problem-solving studies are theorized to have occurred because older adults used experience-based strategies to solve problems more than younger adults and not because older adults cannot use the same strategies that younger adults use. Older adults were also theorized as being as capable as younger adults of using constraint-seeking strategies without specifically being instructed to do so. These theoretical assumptions led to the formulation of the hypotheses outlined below.

Surgeon Version versus Birthday Version

The content of the surgeon version was expected to evoke experience-based strategies in older adults more than younger adults. Changing the content to theoretically “experience-free” should hinder the use of experience-based strategies for older adults. If so, older participants should be more likely to use constraint seeking strategies in the neutral, birthday version causing age differences to disappear. If a constraint-seeking strategy is not part of participants’ learning histories, a guessing strategy should occur instead.

1. Fewer older adults who are given the surgeon version first will constraint seek compared to younger adults (viz., age effect on surgeon first).
2. No age differences will occur in frequency of constraint seeking for those who are given the birthday version first (viz., no age effect on birthday first).

Receiving Birthday Version First

The “experience-free” context of the birthday version should evoke constraint seeking in older adults and lead to finding the person born in March. If so, how will this experience affect strategy use on the surgeon version? The content of the surgeon problem may be difficult to inhibit for older adults and may cause them to use an experience-based strategy. If this is true, older adults who constraint seek on the birthday version will switch to an experience-based strategy on the surgeon version. Older adults, however, are not hypothesized to switch to an experience-based strategy once they have used constraint seeking. Rather, constraint seeking in the birthday version was hypothesized to transfer to the surgeon version for older adults.

3. More older adults will constraint seek on the surgeon version when given second compared to given first (viz., order effect on surgeon version for older adults).

4. Number of younger adults who constraint seek on the surgeon version will not differ as a function of order (viz., no order effect on surgeon version for younger adults).

Nonspecific carry-over effects or practice effects are not predicted. If there are no practice effects, the same number of participants should constraint seek on the neutral version regardless of whether they receive it first or second.

5. For both younger and older participants, the frequency of constraint seeking on the birthday version will not differ for those who were given it first compared to those who were given it second (viz., no order effect on birthday version for either age group).

Results

Demographic Variables

Prior to analysis, WAIS-3 Vocabulary, WAIS-3 Digit Symbol, and years of education

were examined for accuracy of data entry, missing values, and conformity to the assumptions of multivariate analysis. No missing values occurred. One value in years of education (a 78-year-old male with multiple degrees reported 12 years of education past high school) was a statistically significant univariate outlier ($z = 3.45$). This case was not eliminated, however, because this participant was not an outlier in Vocabulary or Digit Symbol tasks, and also because his years of education was not an outlier when examined within the older participants' distribution alone. No outliers occurred for Vocabulary or Digit Symbol.

Using Mahalanobis distance with $p < .001$ (alpha recommended by Tabachnick & Fidell, 1996), no cases were detected as multivariate outliers. Analyses of Vocabulary and Digit Symbol distributions did not uncover significant skewness or kurtosis. As expected, years of education, however, was significantly skewed. Positive skewness occurs in years of education distributions because all participants had a minimum of 12 years of education. After 12 years of education, as years of education increases, frequency of observations tapers off. An inverse transformation returned the distribution to normal (no significant skew or kurtosis). Thus, while tables displaying mean years of education reflect raw scores, multivariate and univariate analyses involving education utilized the transformed scores. Table 3 shows Demographic variables.

Table 3

Demographics and WAIS-3 Subtests: Means and Standard Deviations

Variable	Younger				Older			
	Female		Male		Female		Male	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>
Age	21.1	1.6	21.3	1.3	69.6	6.1	70.4	5.9
Health	8.3	1.2	8.5	0.8	8.7	1.3	7.9	1.8
Education	14.4	1.2	14.8	0.9	14.9	2.8	17.1	3.5
Vocabulary	42.5	7.7	40.3	7.5	53.0	8.7	52.4	7.4
Digit Symbol	92.1	12.7	81.6	13.0	59.3	12.2	59.3	12.2

Note: Health represents participants' subjective ratings on a scale from 1 to 10, where 10 was excellent health.

Age and Sex Differences in Health, Vocabulary, and Digit Symbol Scores

A 2 (age) x 2 (sex) multivariate analysis of variance (MANOVA) was conducted using WAIS-3 vocabulary, WAIS-3 Digit Symbol, and years of education (transformed to inverse scores) as dependent variables. The multivariate age x sex interaction was not statistically significant ($p = .192$). Multivariate F -tests for main effects of age and sex were significant (p 's < .01). Follow-up univariate tests revealed no significant differences in self-reported health ($F < 1$) or education, $F(1, 60) = 2.80$, $p = .100$, as a function of age. The main effect of age was significant for the other two dependent variables. Specifically, older adults scored significantly higher than younger adults on the vocabulary task, $F(1, 60) = 33.14$, $p < .001$, and significantly lower on the Digit Symbol task, $F(1, 60) = 63.90$, $p < .001$. Main effects of sex were not significant for Vocabulary and health (F 's < 1), but were significant for Digit Symbol and education. Specifically, females had significantly higher Digit Symbol scores than males, $F(1,$

60) = 5.82, $p = .019$, and had significantly fewer years of education, $F(1, 60) = 5.25$, $p = .026$.

Overall Strategies

Overall strategy was determined using the traditional classification method (see Appendix G). Possible strategies were: Unidentifiable, Item checking, Pair checking, Partial-constraint seeking, and Constraint seeking. Table 4 shows the frequencies of unidentifiable strategies, item checking, pair checking, partial constraint seeking, and constraint seeking. The frequencies of strategies were almost identical for both versions of the task, but not for the two age groups.

Table 4

Frequency of Strategies Used by Older and Younger Adults for Each Order of Presentation

Strategy	Received first		Received second	
	Birthday	Surgeon	Birthday	Surgeon
Younger adults				
Unidentifiable	2	2	0	0
Item checking	3	3	3	5
Pair checking	1	0	1	0
Partial constraint seeking	1	0	0	0
Predominant constraint seeking	9	11	12	11
Older adults				
Unidentifiable	2	0	1	2
Item checking	8	10	7	3
Pair checking	0	1	0	0
Partial-constraint seeking	0	1	0	0
Predominant constraint seeking	6	4	8	11

Planned Comparisons

Rationale for planned comparison approach. Five specific hypotheses were generated for this dissertation. Each specified presence (and direction if present) or absence of particular simple effects. Rather than employ an insensitive omnibus test often used to protect against type I errors by ‘data-mining’ researchers who have no specific hypotheses, a planned-comparison approach was utilized. A planned-comparisons approach also provides increased power for detecting effect sizes which may be moderate or small and interpretation tends to be more substantive and clear (Rosnow & Rosenthal, 1989).

Keppel (1991) warns, however, that testing simple effects without testing the interaction may not provide an unambiguous picture of the results. Thus, after the planned comparisons were conducted, omnibus tests were performed on the data. Results of these tests are reported to demonstrate the power (or lack of power) the current design may have to detect particular effects. That is, if the simple effects are found to be significant as hypothesized, but the omnibus interaction is not statistically significant, the magnitude of effects may be only moderate. In such a case, if the traditionally omnibus approach had been used, these simple effects would not be detected, thus committing a Type II error (failing to reject the Null hypothesis when the Null is false).

Strategy type as dichotomous dependent variable. Because of the low frequency of unidentifiable, item checking, pair checking, and partial constraint seeking strategies, these strategy types were collapsed and recoded as “not constraint seeking”. This yielded a dichotomous variable of not predominant constraint seeking or predominant constraint seeking. A second method for measuring strategy type distinguished between consistent constraint seeking

and not consistent constraint seeking. This also yielded a dichotomous dependent variable representing strategy type.

Other measurements of strategy type. Strategy type was also measured on a quantitative continuum of total items eliminated after three questions and on a continuum of no constraint seeking, some constraint seeking, and consistent constraint seeking. These interval data allowed for general linear model analyses to be conducted. Appendix H describes these alternative methods and results of subsequent analyses are displayed. Results did not differ from original nonparametric analyses described below but were included for the purpose of informing the reader of alternative methods for measuring strategy type.

Description of planned-comparison analyses. Multiway frequency analyses were conducted to test the five hypotheses of the present study using SPSS (ver. 9.0) Crosstabs function. Likelihood ratios were used instead of Pearson Chi-square tests because Likelihood ratios are less affected by small sample sizes (Howell, 1992). Each of these likelihood ratios was computed for a 2 x 2 table (factor x dependent variable) where constraint seeking or no constraint seeking represented the dichotomous dependent variable. Reported *p* values for these analyses are probabilities calculated from Fisher's Exact test. That is, the probabilities reported for these likelihood ratio values are the exact probabilities of the observed frequencies being as different or more different than they are from the expected frequencies. The usual *p* values reported for likelihood ratios and Chi-square analyses are actually estimates based on Monte Carlo studies of Chi-Square distributions. Pearson's *r* is reported as a measure of association and is equal to Eta, Phi, and Cramer's *V* in these analyses.

Each analysis was performed twice: once using the predominant constraint seeking as the

dependent variable and once using consistent constraint seeking as the dependent variable. After data analyses were complete, alternative methods for measuring strategy type were employed and further analyses conducted.

First Planned Comparison

Predominant constraint seeking. Likelihood ratio analyses were conducted to test for specific age effects on each version when given first. This analyses tested the hypotheses that fewer older than younger adults would use constraint seeking in the surgeon version but not the birthday version. There was no age effect on constraint seeking for the birthday version $\chi^2 = 1.13$, $p = .240$, $r = -.19$. There was, however, a significant effect of age on constraint seeking for the surgeon version, $\chi^2 = 6.15$, $p = .016$, $r = -.44$. Significantly fewer older adults ($N = 4$) used constraint seeking on the surgeon version than younger adults ($N = 11$) when this version was

Figure 1. Number of participants using predominant constraint seeking on birthday and surgeon versions received first.

given first. Younger adults were 6.6 times more likely to constraint seek on the surgeon first condition than older adults.

This odds ratio (i.e., 6.6 to one) was calculated by first calculating conditional odds for constraint seeking for younger adults and then for older adults in the surgeon first condition. Conditional odds are the ratio of occurrences to nonoccurrences. For

example, for the surgeon first condition, 11 younger adults used constraint seeking. Because there was a total of 16 younger adults in this condition, five younger adults did not constraint seek in this condition. Thus, the conditional odds for younger adults constraint seeking in the surgeon first condition are 11/5 (i.e., for every 11 younger adults that constraint seek, five do not). The same conditional odds for older adults are 4/12. Odds ratios are calculated by dividing the conditional odds for the young by the conditional odds for the old (11/5 divided by 4/12). The ratio between these two odds is 6.6 to one, meaning that younger adults are 6.6 times more likely to use constraint seeking than older adults.

Consistent constraint seeking. The same analysis was conducted using consistent constraint seeking as the dependent variable. Frequency of consistent constraint seeking did not differ as a function of age in the birthday first condition ($N_s = 6$). Similar to the first planned

Figure 2. Number of participants using consistent constraint seeking on birthday and surgeon versions received first.

comparison using the predominant constraint seeking, fewer older adults ($N = 2$) used consistent constraint seeking in the surgeon first condition than younger adults ($N = 7$, see Figure 2). This difference, however, was not significant, $\chi^2 = 3.87$, $p = .057$, $r = -.35$. Although expected frequencies for two cells were 4.5 in this 2 x 2 table, the Chi-square test is particularly robust when overall N is greater than 20 (Howell, 1992). Younger

adults were 5.4 times more likely to use consistent constraint seeking in the surgeon first condition than older adults.

Second Planned Comparison.

Predominant constraint seeking. Likelihood ratio analyses were conducted to test for effects of order in the surgeon task for each age group. This analysis tested the specific hypothesis that receiving the birthday version first increases the likelihood of constraint seeking on the surgeon problem for older adults. The same number of younger adults used constraint seeking in the surgeon version when it was given first or second ($N = 11$). However, significantly fewer older adults used constraint seeking when they were given the surgeon version first ($N = 4$) than second ($N = 11$), $\chi^2 = 6.15$, $p = .016$, $r = -.44$ (see Figure 3).

Figure 3. Number of participants using predominant constraint seeking on surgeon version.

Consistent constraint seeking. The same analysis was conducted using consistent constraint seeking as the dependent variable. The number of younger adults who used constraint seeking in the surgeon first condition ($N = 7$) did not significantly differ from younger adults receiving it second ($N = 10$), $\chi^2 = 1.13$, $p = .240$, $r = -.19$ (see Figure 4). However, significantly fewer older adults used constraint seeking when given the surgeon version first ($N = 2$) than second ($N = 8$), $\chi^2 = 5.24$, $p = .027$, $r = -.41$.

Figure 4. Number of participants using consistent constraint seeking on surgeon version.

Third Planned Comparison

Predominant constraint seeking. Likelihood ratio analyses were conducted to test for effects of order in the birthday version for each age group. This analysis tested the specific hypothesis that receiving the surgeon version first would not have an effect on frequency of constraint seeking in the birthday version. It also tested the assumption that the result of the second planned comparison was not due to practice effects. That is, if there were no general practice effects, receiving the birthday problem second should not lead to higher frequencies of constraint seeking overall. There was no order effect for younger, $\chi^2 = 1.25$, $p = .229$, $r = .20$, or older adults, $\chi^2 < 1$, $r = .13$. Figure 5 shows frequency of constraint seeking for the Birthday version.

Figure 5. Number of participants using predominant constraint seeking on birthday version.

Consistent constraint seeking. The same analysis was conducted using consistent constraint seeking as the dependent variable. There was no order effect for younger, $\chi^2 = 1.13$, $p = .240$, $r = .19$, or older adults, $\chi^2 < 1$, $r = -.07$. Figure 5 shows frequency of constraint seeking for the Birthday version.

Figure 6. Number of participants using consistent constraint seeking on birthday version.

Omnibus Analyses

Multiway frequency analyses. Multiway frequency analyses were conducted on the data used in each planned comparison in attempts to screen for significant main effects or interactions. Each was in the form of a 2 (age) x 2 (condition) x 2(strategy) logit analysis in which strategy (constraint seeking or not constraint seeking) served as the dependent variable. Logit refers to a log-linear analysis where one factor is treated as a dependent variable. As with the planned comparisons, the dependent variables was either predominant constraint seeking or complete constraint seeking. Condition referred to: surgeon first versus birthday first, surgeon first versus surgeon second, or birthday first versus birthday second. The logit analysis served the same function as an ANOVA. That is, interactions and main effects were tested for significance.

Traditionally, testing for effects is often the first step of a logit analysis. Significant effects are then used to form models of the observed data. These models are used to generate expected frequencies which are then compared to the observed frequencies for goodness of fit. Other approaches for using logit analyses include testing the fit of a model hypothesized a priori without testing for effects or simply testing effects. The approach used depends on the goals of the researcher. The logit analyses in the current study were used to test for significant interactions (and main effects from the ANOVA perspective) involving age, order or version and type of strategy the same way an ANOVA would be used. Best fitting models were also generated using a backward stepdown method for each logit analysis. Each model that best fit the data was no different from the significant effects reported for the respective planned-comparison logit analyses. For example, in the birthday first versus surgeon first logit analysis, the age x strategy interaction was significant and the model that best fit these data was the age x strategy

interaction. Thus, to avoid redundancy, these models were not reported in the following paragraphs but can be found in Appendix I.

Three interaction effects were tested in each logit analysis, (a) age x condition x strategy, (b) age x strategy, and (c) condition x strategy. These effects correspond to the age x condition interaction and the main effects of age and condition, in an ANOVA model respectively. In each analysis, the expected frequencies were greater than 5 for more than 80% of the cells and no cell frequency was zero. Component χ^2 s ($df = 1$, $N = 64$) are reported and reflect the partial likelihood ratio statistic.

Birthday First versus Surgeon First

Predominant constraint seeking. A 2 (age) x 2 (version) x 2(strategy) logit analysis was performed on data from surgeon and birthday versions when these tasks were given first. The age x version x strategy interaction was not significant, $\chi^2 = 1.12$, $p = .290$, nor was the version x strategy interaction, $\chi^2 = 0$. There was a significant age x strategy interaction, $\chi^2 = 6.38$, $p = .012$. Examination of the observed frequencies for the age x strategy term indicated that twenty younger adults used constraint seeking compared to 10 older adults. The age x strategy model best fit the observed data (see Appendix I) and represents a main effect of age on strategy type.

Consistent constraint seeking. The same analysis was conducted using consistent constraint seeking as the dependent variable. The age x version x strategy interaction was not significant, $\chi^2 = 2.24$, $p = .135$, nor was the main effect of version $\chi^2 = 0$. Unlike the analysis involving predominant constraint seeking, there was no significant age x strategy interaction, $\chi^2 = 1.80$, $p = .179$.

Surgeon First versus Surgeon Second

Predominant constraint seeking. A 2 (age) x 2 (order) x 2(strategy) logit analysis was performed on the data from the surgeon version only. The age x version x strategy interaction was not significant, $\chi^2 = 3.03$, $p = .082$, and the order x strategy and age x strategy interactions were not significant, both $\chi^2 = 3.34$, $p = .067$.

Consistent constraint seeking. The same analysis was conducted using consistent constraint seeking as the dependent variable. The age x version x strategy interaction was not significant, $\chi^2 = 1.09$, $p = .297$, and the age x strategy interaction was not significant, $\chi^2 = 3.46$, $p = .063$. Unlike the previous analysis, strategy classification method, the order x strategy was significant $\chi^2 = 5.56$, $p = .018$. The order x strategy model best fit the observed data (see Appendix I) and represents a main effect of order on strategy type.

Birthday First versus Birthday Second

Predominant constraint seeking. A 2 (age) x 2 (order) x 2(strategy) logit analysis was performed on data from the birthday version only. The age x version x strategy interaction was not significant, $\chi^2 < 1$, and the order x strategy interaction was not significant, $\chi^2 = 1.67$, $p = .197$. The age x strategy interaction was also not significant $\chi^2 = 3.20$, $p = .074$.

Consistent constraint seeking. A 2 (age) x 2 (order) x 2(strategy) logit analysis was performed on data from the birthday version only. The age x version x strategy interaction was not significant, $\chi^2 = 1.01$, $p = .315$, and the order x strategy interaction was not significant, $\chi^2 < 1$. The age x strategy interaction was also not significant $\chi^2 = 1.04$, $p = .307$.

Omnibus-F Test

A 2 (age group) x 2 (sex) x 2 (serial position) x 2 (version; Surgeon versus Birthday)

mixed-factorial analysis of variance (ANOVA) using predominant constraint seeking as the dependent variable was conducted. The purpose of this analysis was to test for main effects and interactions involving sex and to test for position effects in the Latin-square design. Photoset was treated as a random variable and not added as a fifth factor. The rationale for not including photoset was because based on inspection of the protocols, effects appeared to be negligible. The photosets and insignificant effects are described in detail later in this result section.

The main effect of sex and interactions involving sex were not significant. F ratios, p values, and effect sizes for the other main effects and interactions are shown in Table 5.

Table 5

Analysis of Variance for Factorial and Latin Square Designs: Predominant Constraint Seeking

Factorial Model			Latin-square Model	
Effect	$F(1, 56)$	p	Eta^2	Effect
Between-groups				
Age	3.75	.058	.063	Age
Order (subgroup)	.08	.783	.001	Version x Position
Age x Order (subgroup)	1.22	.273	.021	Age x Version x Position
Within-subject				
Version	.40	.530	.007	Version
Age x Version	.00	1.00	.000	Age x Version
Order (subgroup) x Version	14.40	.001	.205	Position
Age x Order (subgroup) x Version	3.60	.063	.060	Age x Position

The same mixed-factorial ANOVA was performed using consistent constraint seeking as the dependent variable. Results were almost identical. The order x version interaction was still

significant but was qualified, however, by a significant order x version x sex interaction, (see Table 6). The results of the two omnibus ANOVA's are interpreted in the following sections.

Table 6

Analysis of Variance for Factorial and Latin Square Designs: Consistent Constraint Seeking

Factorial Model			Latin-square Model	
Effect	F(1, 56)	p	Eta ²	Effect
Between-groups				
Age	2.41	.126	.041	Age
Order (subgroup)	.98	.327	.017	Version x Position
Age x Order (subgroup)	.98	.327	.017	Age x Version x Position
Within-subject				
Version	.09	.767	.002	Version
Age x Version	.78	.376	.014	Age x Version
Order (subgroup) x Version	10.72	.002	.161	Position
Age x Order (subgroup) x Version	.09	.767	.002	Age x Position
Sex x Order (subgroup) x Version	4.34	.042	.072	Sex x Position

Latin-square Design

A common use of a Latin-square design is for partitioning potential effects of nuisance variables equally across variables of interest. For example, an experimenter using a within-subject design might examine whether performance differs across two different tasks ("task" is the variable of interest). In such a design, the position in which a task is received (first or second) may have a general effect on performance (e.g., practice effects). By counterbalancing order in which tasks are received, variance due to such effects theoretically is spread equally across tasks

(but is not eliminated). That is, practice effects typically are assumed not to interact with task effects. This method also allows such potential effects to be analyzed and interpreted.

Because order of presentation of task version was counterbalanced, the order x version interaction was completely confounded with position. From the perspective of this Latin-square-design, position is a within-subject variable-- every participant receives two tasks; one first and one second. Table 7 contrasts the factorial design with the Latin-square design and displays corresponding frequencies for each cell. Note that task is a within-subject variable in both designs, which leads to difficulty in performing the statistical analysis using available software programs. Age, in this design, is considered a 'category' variable in Latin-square/factorial models (Reese, 1997).

None of the commonly used statistical software packages were able to conduct analyses on a Latin-square design (Reese, 1997). Fortunately, every main effect and interaction in a Latin-square model can be gleaned from the output of the factorial analysis (see Reese, 1997 for a thorough tutorial). For example, the version x order interaction in the factorial model is mathematically equivalent to the main effect of position in the Latin-square model (Reese, 1997).

Latin-square main effects and interactions appear in the last column in Tables 5 and 6 in correspondence with their equivalent factorial effects or interactions. The only significant finding was the position effect. Specifically, constraint seeking occurred significantly more often on tasks received second ($N = 42$) versus tasks received first ($N = 30$) regardless of version.

Table 7

Frequency of Constraint Seeking Displayed in Factorial and Latin-square Designs

		Version		Version		
Age	Variable	Birthday	Surgeon	Birthday	Surgeon	Total
Factorial model						
Young	Subgroup 1	1st position	2nd position	9	11	20
	Subgroup 2	2nd position	1st position	12	11	23
	Total			21	22	
old	Subgroup 1	1st position	2nd position	6	11	17
	Subgroup 2	2nd position	1st position	8	4	12
	Total			14	15	
Latin-square model						
Young	1st position	Subgroup 1	Subgroup 2	9	11	20
	2nd position	Subgroup 2	Subgroup 1	12	11	23
	Total			21	22	
Old	1st position	Subgroup 1	Subgroup 2	6	4	10
	2nd position	Subgroup 2	Subgroup 1	8	11	19
	Total			14	15	

Note. Subgroup is the term used for the counterbalancing variable when comparing factorial and Latin-square models (Reese, 1997). In this design, subgroup refers to order (subgroup 1 = birthday first; subgroup 2 = surgeon first) and is a between-groups variable. Note that position is a within-subject variable in the Latin-square model.

Interpreting the position effect. In Latin-square designs, position is often a variable that represents practice effects. That is, when participants are given a second task, performance may

differ from the first task simply because they learned from the first task or because they became more comfortable with the testing conditions and not because the second task was different from the first. Generally, position effect is a nuisance variable and makes interpretation of effects of task difficult. In the present study however, a position effect was predicted— but only for older adults on the surgeon version. That is, fewer older adults were predicted to use constraint seeking on the surgeon version when receiving it first than when receiving it second. No position effect was predicted for younger adults on either version nor for older adults on the birthday version.

This hypothesis would have been supported by a significant three-way interaction (age x version x position) in the Latin-square model. This interaction was mathematically equivalent to the age x order interaction in the factorial model (see Tables 5 and 6), and was not significant.

Interaction involving sex. The significant sex x order x version interaction is equivalent to the sex x position interaction in the Latin-square design (see Table 6). Figure 7 shows that the position effect can be attributed mostly to males in the surgeon condition (see last set of bars in Figure 7). Specifically, fewer males used consistent constraint seeking when given the surgeon version first ($\underline{N} = 3$) versus second ($\underline{N} = 10$).

Figure 7. Number of participants using consistent constraint seeking as a function of order across sex and version.

Reference to Stereotype

Thirteen younger adults and 20 older adults made references to stereotypes regarding surgeons (with no sex differences in either age group). Reference to stereotype was a good predictor of strategy type (predominant constraint seeking versus not predominant constraint seeking), $r(61) = -.51$, $p < .001$. However, performing correlational analyses on age groups separately revealed younger adults to be mostly responsible for this relation, $r(30) = -.69$, $p < .001$. The correlation for older adults was not significant, $r(29) = -.28$, $p = .134$.

This relation was similar using consistent constraint seeking as strategy type, $r(61) = -.56$, $p < .001$. Younger adults were also more responsible for this relation, $r(30) = -.63$, $p < .001$, although the correlation for older adults was significant, $r(29) = -.42$, $p = .020$.

Reference to Experience

Only four older adults (two men and two women) and two younger adults (both were males) made references to experience with surgeons. Both of the younger adults and two of the older adults item checked. The other two older adults used consistent constraint seeking and partial constraint seeking. Because of the low rate of occurrence no further analyses were conducted using reference to experience.

Reference to Stimuli

Frequencies of statements made in reference to particular stimuli in the pictures (e.g., age, gender, background, clothes, facial features, hair color) were tallied (see Table 8). The most frequent question was whether the target was a male. This question was asked in 40% of the 128 tasks and accounted for 29% of first questions. The second most common constraint seeking questions referred to hair (24.2%) and background (16.4%), but they were rarely asked as the first question (4.7%). Participants' second most frequent first question was whether the target was a female (8.5%).

Table 8

Frequency of Questions Using Specific Stimuli to Eliminate Pictures

Group	Stimuli used in questions					
	Male	Female	Background	Age	Hair	Earring
1st question						
Younger						
Females	17	3	1	0	0	0
Males	10	0	4	0	1	0
Older						
Females	4	8	0	1	0	0
Males	6	0	0	1	0	0
2nd question						
Younger						
Females	2	0	5	0	8 ^a	5
Males	5	5	0	1	1	4
Older						
Females	1	0	3	4	2	0
Males	5	1	1	2	2	1
3rd question						
Younger						
Females	0	0	4	2	8 ^a	2
Males	0	0	1	2	4	1
Older						
Females	1	1	2	0	3	0
Males	0	0	0	2	2	1
Total	51	18	21	15	31	14

^a Two participants referred to this stimulus more than once within the same task.

Task Interpretation

Concurrent task interpretation of the surgeon problem was scored for 63 participants. Twenty-two participants' protocols contained statements about narrowing down their choices or eliminating groups of photos without mentioning who was likely or unlikely to be a surgeon. These were coded as logic-based interpretations. Thirty participants' protocols contained statements about their experiences with surgeons or about trying to determine "who looks like a surgeon." These were coded as experience-based interpretations. Eleven participants' protocols contained both experience and logic types of statements. These were coded as pragmatic. Appendix J shows the coding method used for classifying interpretation and examples of statements from participants' protocols that specifically demonstrate how interpretation was coded.

Experience-based interpretations were coded as 0, pragmatic as 1, and logic-based as 2. This coding created a quantitative variable that rates degree of logic from low to high. There was no correlation between concurrent task interpretation and presence or absence of consistent constraint seeking in the surgeon version, $r(61) = -.02$, $p = .874$. A logit analysis using age group, order, interpretation, and strategy type as factors resulted in no model involving interpretation (either alone or with other factors) interacting with strategy type.

Photoset effects

The two photosets (glasses and no glasses) were designed to contain similar characteristics so as not to differentially affect problem solving. One unforeseen difference between sets was females' hair length. Because the most common first question ($N = 37$) was "Is the target a male?" photos of females often were left for the second question. In these cases, a

common second question was, “does the target have long/short hair?” The experimenter responded to this question by first asking “to which photos are you referring?” Differences in hair length of females in the photosets could have been a potential factor that affected constraint seeking.

Eleven participants used hair length as part of their second question (eight younger females, three younger males, one older female, and three older males). Two of these participants used hair length on both versions, making the total 13. Seven of these participants stated that the ratio of short hair to long hair was 50/50 (i.e., half of the women had short hair and half had long hair). That is, the question served as a consistent constraint seeking question. In all seven of these incidences, the photoset was ‘glasses’

Five participants (including one of the seven from above) indicated that the ratio of short hair to long hair was not 50/50 (i.e., some said three had short hair or three had long hair). One of these participant used hair length in both tasks making six questions total. Half of these six questions occurred in the glasses photoset and half in the no glasses photoset. Table 9 shows the breakdown of second questions involving hair length. Table 9 reveals no clear pattern of photoset effect or interaction between photoset and age, version, or sex based on hair length questions.

One negligible effect of photoset was that only one woman was wearing earrings in the glasses photoset and two women wore earrings in the no glasses photoset. Two participants (one younger male and one younger female) used earrings as a constraint seeking stimulus in the glasses photoset, which led to only partial constraint seeking.

Table 9

Photoset and Version for Participants Who Asked
about Hair Length on Second Question

Picture problem		Participant	
Photoset	version	Age group	Sex
Consistent constraint seeking			
glasses	surgeon	younger	female
glasses	surgeon	younger	female
glasses	surgeon	younger	female ^a
glasses	birthday	younger	female
glasses	birthday	younger	female
glasses	birthday	younger	male
glasses	birthday	older	male
Not consistent constraint seeking			
no glasses	birthday	younger	female ^a
no glasses	birthday	younger	female
no glasses	surgeon	younger	female ^b
glasses	birthday	younger	female ^b
glasses	birthday	older	female
glasses	birthday	older	male

Note: ^a same participant, ^b same participant

Discussion

Specific predictions about age differences between the two Picture task versions were expressed in hypotheses 1 and 2. Hypothesis 1 predicted age differences in the surgeon version when it was presented first. This hypothesis was based on previous research findings with the Accountant problem (Laipple, 1991) and Poison Meals problem (Jurden et al., 1992; Laipple 1991). Indeed, in the surgeon version, fewer older adults used constraint seeking than younger adults. Hypothesis 2 predicted that the “experience-free” content of the birthday version would result in disappearance of age differences. As predicted, older adults used constraint seeking just as often as younger adults in the birthday version. Together, these results suggest that age differences in concept identification problems: (a) are dependent upon whether task content is “experience free,” and (b) are not an issue of competence. Older adults are just as able as younger adults to use constraint seeking strategies without explicit training, instructions, or modeling when the problem content is changed to something that does not lend itself to using personal experience.

The fact that the least amount of constraint seeking occurred for older adults in the surgeon version is consistent with the assumption that the surgeon version promoted experience-based strategies whereas the birthday version did not. These results suggest that solutions produced by older adults are more likely than younger adults to be determined by factual information, rules of thumb, stereotypes, or personal experiences when possible.

Unfortunately, participants rarely referred to personal experiences with surgeons in this study. Future researchers should add prompts to encourage more statements about what affected participants’ strategies. For example, participants should be asked, “Why did you ask the

questions you did?”

Participants’ references to stereotypes were more abundant, and predicted strategy use. Stereotype is a not-always-true belief that members of a group share characteristics. Specifically, referring to a surgeon stereotype (e.g., “Surgeons tend to be male”; “Surgeons would look intelligent”) was correlated with using an item checking strategy. This suggests that item checking and other forms of non-constraint seeking strategies reflect hypothesis testing by the participant. That is, these participants could have been attempting to eliminate the most probable candidates with each question, or find confirming evidence about who they thought were the most probable candidates. However, the relation between stereotype referencing and strategy type was weaker for older adults. This is surprising given the assumption that older adults are more likely to rely on past experience. Reference to stereotypes per se, may not represent an underlying causal mechanism that determines strategy use; rather, reference to stereotypes may be evoked by the problem context and may not actually indicate that past experience is being used to solve the problem. The fact that stereotypes exist about surgeons and not people born in March still seems to be the best explanation for the results.

Hypotheses 3 and 4 dealt with the effect of receiving the surgeon version second. The fact that older adults were more likely to use constraint seeking on the surgeon version after they completed the birthday version demonstrates that older adults can use logic-based strategies even if the task promotes experience-based strategies. No specific instructions or prompts to ignore personal experience were necessary to cause older adults to constraint seek on the surgeon version just as often as younger adults. Rather, the presentation of the neutral, birthday version promoted constraint seeking which led to the successful selection of the target (viz., person born

in March). This strategy was then generalized to the surgeon version. Even if partial constraint seeking occurred on the birthday version, this was sometimes enough to evoke consistent constraint seeking on the surgeon version.

Younger adults' performance on the surgeon problem did not significantly differ as a function of order. This result suggests that because younger adults are less likely to use experience-based strategies, the birthday version first condition did not have an effect.

Hypothesis 5 dealt with demonstrating that general carry-over effects were not the cause of the results. There was no difference in strategy use on the birthday version when given first or second. Thus, the order effect appears to be specific for older adults solving the surgeon version and was caused by completing the birthday version first. However, the Latin-square analyses revealed a main effect of position suggesting an overall practice effect. Thus, a general practice effect may exist but the results related to the other four hypotheses suggest that the practice effect is small and mostly caused by the surgeon first condition in older adults.

Problem Content as a Causal Factor

What causes older adults who can use logic-based strategies, to use a strategy that is experience-based? An experience-based strategy (e.g., picking the most probable target as the solution or using rules of thumb) is probably adaptive in the everyday life of older adults. That is, such a strategy has led to successful problem solving. Older adults appear to use physical characteristics as clues for determining who is the target. The task instructions did not state that using physical characteristics was not allowed or that participants should not look for clues. Therefore, such a strategy was within the parameters of the problem (i.e., allowable) and not incorrect. In fact, this strategy could have led to correct selection of the surgeon in less than three

questions, had there actually been a surgeon. To say that this strategy is less efficient or effective is only to view the problem from the perspective of the experimenter who knows that no one is the surgeon.

These results explain differential performance between age groups found in the Accountant task (Laipple, 1991) and Poisoned Meals task (Laipple, 1991; Reese, et al., 1990). Had these studies involved tasks with targets devoid of potential stereotypes or had these participants been instructed not to pick out the likely target but to assume all items were equally likely targets, older adults might have used constraint seeking just as often as younger adults.

Using a logic-based strategy would require deliberate suppression of the typical experience-based response and switching to a strategy that is not commonly used. Given this assumption, the question becomes, what would make older adults use a logic-based strategy when the experience-based strategy is the dominant response? For the birthday version, an experience-based strategy was not evoked. This caused older adults to switch to strategies such as constraint seeking. Once the logic-based strategy was used on the birthday version, this response was generalized to the surgeon version. Use of this strategy was no longer effortful because it had just been used.

Construct Validity

The results of this study support a conclusion suggested by Laipple (1991) that concept-formation tasks (e.g., Accountant, House Plants, and Poison Meals problems) lack construct validity. That is, in their present form, they do not measure ability to use logic-based strategies for older adults. The present results suggest that if the content of the concept formation problems were changed to “experience-free”, older adults would likely switch to a logic-based strategy.

Indeed, when Laipple (1990) assessed participants' solutions relative to their definitions or interpretations of a concept formation problem (e.g., in the House Plants problem, some older participants stated they thought the problem was an assessment of their knowledge of plants), age differences in "relative correctness" disappeared. The birthday version in the present study does appear to be "experience free" because of the increase in constraint seeking demonstrated by older adults.

Problem Interpretation

The instructions of the Pictures task stated (a) that the experimenter knew which person was the target, (b) that the participant was to find out who the target is, and (c) that the participant could ask three yes or no questions before they had to indicate who was the target. From the protocols and from administering the tasks to the participants, the initial state and the goal state of the task appeared to be understood by all participants. Participants also understood that they solved the problem by correctly indicating the target photo and if they did not indicate the target, they did not solve the problem.

Problem interpretation has been one concept thought to explain age related problem-solving differences. Laipple (1991) found a strong correlation between interpretation and strategy use. Laipple found 'interpretation' even accounted for the variance due to age. This strong correlation may have been due, however, to a bias in how interpretations were coded. The questions (i.e., strategies) Laipple's participants used were not removed from the protocols when interpretation was coded (Laipple, 1999, personal communication). Thus, a participant's strategy likely affected the coding of that participant's interpretation. The method of coding interpretations used in the present study involved removing the participants' questions (i.e., their

strategies) from the protocols before coding. With this method, no significant correlation was found between problem interpretation and strategies. That is, participants' interpretations of the task, which were classified by the experimenter (without knowledge of type of strategy used) as logic-based, experience-based or pragmatic did not predict type of strategy. Interpretation, then, may not be a causal mechanism determining strategy use. The alternative hypothesis is that the assessment of interpretations was invalid in the present study. Appendix J shows the method used for classifying interpretation and examples of participants' protocols for readers to review as to their validity.

References to stereotypes could be another way to assess participants' interpretations. A statement such as "Surgeons tend to look intelligent" or "This guy looks too young to be a surgeon" may indicate an interpretation of the problem as one to be solved by using experience. However, because these statements occurred while participants were solving the problem, these statements may just be a verbal description of the strategy they are using (e.g., I should pick people who look intelligent, or I should eliminate the young people) and not an underlying hypothetical construct. Whether a reference to a stereotype is "an active reconstruction of the task" (i.e., interpretation) or "a process for getting from the initial state to the goal state" (i.e., strategy) becomes unclear at this point.

Limitations

The current experimental design had several limitations. First, participants, after first hearing the problem, should have been asked to restate it back to the experimenter, asked what the goal of the problem was, and asked what types of questions are 'legal' or 'OK' or to ask and what questions are not allowed. Responses to these probes would be more ideal for inferring

participants' interpretations.

Second, characteristics of people in the photosets must be more uniform. Although relatively few participants' asked questions using hair length and earrings, future researchers using larger sample sizes or different populations may encounter a photoset effect. Glasses versus no glasses would be a useful characteristic to divide a photoset.

Third, the photosets did contain salient features that promoted constraint seeking. That is, there were equal numbers of men and women in the photosets and "Is it a male?" was the most popular question. It is unknown whether sex being counterbalanced in the photosets made people ask about sex, or whether participants just tend to ask about sex. Also, the stereotype of surgeons being male could have led to constraint seeking because of the question it evokes.

Directions and Implications for Future Research

The present study demonstrated that problem content can account for age differences. Future researchers should examine the relation between age-related environmental changes and laboratory problems' content. Manipulating the environment may not be feasible but quasi-experimental designs that categorize older and younger adults by their environments would be useful. Older adults who are part of the academic environment (e.g., teachers, professors, students in their third or fourth year of college or graduate students) could be contrasted with those that are not. Older adults whose environments are most similar to younger adults would likely solve problems like younger adults. The academic context is not the only possible category but one that is easy to use as a matching variable. One problem is that some younger adults also use experience-based strategies and item check. The challenge for future researchers will be to find the common environmental context that causes problem solving differences.

Researchers who intend to measure reasoning abilities of older adults (e.g., ability to use logical strategies) must consider the content of their tasks. The content of tasks must be scrutinized as to whether past experience could be useful from some perspective. Researchers should attempt to understand the problems from the perspective of the participants. This could be done by comparing the content of the problem to everyday life contexts of older adults and pilot testing the problems to collect thinking aloud protocols.

Although the prompts for the participants to think aloud were the same as those used by Laipple (1991), they may not have been enough to elicit valid representations of interpretations. Interpretations, by definition, must occur before a strategy is used if they are to affect strategy use. Participants should be asked to interpret the problem before solving it (e.g., have participants administer instructions to the experimenter before completing problem, have participants state what they believe is the goal of the problem, have participants state examples of questions that can or cannot be asked, ask the participants if there are different ways to solve the problem).

General Comments

The general conclusion is that older adults are just as able as younger adults to use logic-based strategies. Researchers must learn about changing environmental aspects of aging adults to understand problem-solving tendencies of older adults in the lab. Rather than perceiving problem solutions to be correct or incorrect, unintended solutions should be perceived as generalized responses from problems experienced in everyday lives. Different solutions should not necessarily be interpreted as lack of competence. Rather, different solutions should be viewed as being caused by differences in recent and relevant learning histories of participants. Problem interpretation is a useful term to refer to solving problems in different ways than intended by the

researcher. Interpretation is also useful when referring to a participant having a different understanding of problem instructions or goal of the problem. However, the search for participants' "interpretations" as causal factors can be misleading. Even if interpretations are casual factors, they appear to be difficult to measure in an unbiased objective manner.

References

- Arenberg, D. (1968). Concept problem solving in young and old adults. Journal of Gerontology, 23, 279-282.
- Arenberg, D. (1974). A longitudinal study of problem solving in adults. Journal of Gerontology, 29, 650-658.
- Berg, C. A., Klaczynski, P. A., Calderone, K. S., & Strough, J. (1994) Adult age differences in cognitive strategies: Adaptive or deficient? In J. D. Sinnott (Ed.) Interdisciplinary handbook of adult lifespan (pp. 371-388). Learning, CT: Greenwood
- Chase, P. N., & Bjarnadottir G. S. (1992). Instructing variability: Some features of a problem-solving repertoire. In S. C. Hayes & L. J. Hayes (Eds.), Understanding verbal relations (pp. 181-193). Reno, NV: Context Press.
- Denney, N. W (1982). Aging and cognitive changes. In B. B. Wolman (Ed.) Handbook of developmental psychology (pp. 807-827) Englewood Cliffs , NJ: Prentice Hall.
- Denny, N. W. (1990). Adult age differences in traditional and practical problem solving. In E. A. Lovelace (Ed.) Aging and cognition: Mental processes, self awareness and interventions (pp. 329-349). North-Holland: Elsevier Science.
- Duncker, K. (1945). On problem solving. Psychological Monographs, 58, (whole no. 270) pp. 1-113.
- Ericsson, K. A., & Simon, H. A. (1996). Protocol analysis: Verbal reports as data (Rev. ed.). Cambridge: MIT press.
- Garnham, A. & Oakhill, J. (1994). Thinking and reasoning, Cambridge, MA; Blackwell.

- Gilhooly, K. J. (1996) Thinking: Directed, undirected and creative (3rd Ed.). London: Academic Press.
- Howell, D. C. (1992). Statistical methods for psychology (3rd Ed.). Boston: PWS-KENT.
- Jurden, F. H., Reese, H. W., Cohen, S. H., & Puckett, J. M. (1992) Cognitive resources in aging: decomposition of effects. Unpublished manuscript, West Virginia University.
- Klahr, D., & Siegler, R. S. (1978). The development of children's knowledge. In H. W. Reese & L. P. & Lipsitt (Eds.), Advances in child development and behavior (vol. 12, pp. 61-116). New York: Academic Press.
- Laipple, J. S. (1990). Solving everyday problems in young and old adulthood. Unpublished master's thesis, West Virginia University, Morgantown, WV.
- Laipple, J. S. (1991). Problem solving in young and old adulthood: The role of task interpretation. Unpublished doctoral dissertation, West Virginia University, Morgantown, WV.
- Labouvie-Vief, G. (1977). Adult cognitive development: In search of alternative interpretations. Merrill-Palmer Quarterly, 23, pp. 227-263.
- Mook, D. G. (1996). Motivation: The organization of action (2nd ed.) New York: W. W. Norton and Company
- Reese, H. W. (1994). Cognitive and behavioral approaches to problem solving. In S. C. Hayes, L. J. Hayes, M. Sato, & K. Ono (Eds.), Behavior analysis of language and cognition (pp. 197-258). Reno, NV: Context Press.

- Reese, H. W. (1997). Counterbalancing and other uses of repeated-measures Latin-square designs: Analyses and interpretations. Journal of Experimental Child Psychology, 64, 137-158.
- Reese, H. W., Puckett, J. M., & Cohen, S. H. (1990). Mechanisms underlying cognitive aging process. Final report, National Institute on Aging Grant No. R01AG06069, West Virginia University, Morgantown, WV.
- Reese, H. W., & Rodeheaver, D. (1985). Problem solving and complex decision making. In J. E. Birren & K. W. Schaie (Eds.), Handbook of the psychology of aging (2nd ed., pp. 474-499). New York: Van Nostrand Reinhold.
- Reitman, W. (1964). Heuristic decision procedures, open constraints, and the structure of ill defined problems. In M. W. Shelly II and G. L. Bryan (Eds.) Human judgments and optimality. (pp. 282-315). New York: Wiley and sons.
- Reitman, W. (1965). Cognition and thought: An information processing approach. New York: Wiley and Sons.
- Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44, 1276-1284.
- Scheerer, M. (1967). Problem-solving. In A. J. Riopelle (Ed.) Animal problem solving. (pp. 26-42). Baltimore: Penguin Books.
- Sinnott, J. D. (Ed.). (1989). Everyday problem solving: Theory and applications. New York: Praeger.
- Skinner B. F. (1966). An operant analysis of problem solving. In B. Kleinmuntz (Ed.), Problem solving: Research, method, theory (pp. 225-257). New York: Wiley.

Thorndike, E. L. (1898). Animal intelligence; an experimental study of the associative processes in animals. Psychological Review Monographs, Supplement 2, No. 4.

Tabachnick, B. G., & Fidell, L. S. (1996). Using multivariate statistics (3rd Ed.). New York: HarperCollins.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale - Third Edition. New York: Psychological Corporation.

Appendix A

CONSENT FORM: Problem solving research

Introduction. I, _____, have been invited to participate in this research study which has been explained to me by Dave Buck or another experimenter, _____, under Dave Buck's supervision. This research is being conducted for Mr. Buck's Dissertation.

Purpose of the Study. The purpose of this study is to learn more about how people solve different types of problems.

Description of Procedures. I will be asked some demographic questions, presented problems to solve and given a vocabulary and numeric task. Responses to the problems will be audio-taped. Approximately 64 people will participate in this study.

Risks and Discomforts. There are no known or expected risks or discomforts from participating in this study.

Benefits. I understand that this study is not expected to be of direct benefit to me, but the knowledge gained may be of benefit to others. Students participating in this research may receive course credit.

Contact Persons. For more information about this research, I can contact Dave Buck, at 423-692-6669 or Dr. Hayne Reese at 304-293-2001, ext. 643. For information regarding my rights as a research participant, I may contact the Executive Secretary of the Institutional Review Board at 304-293-7073.

Confidentiality. I understand that any information about me obtained as a result of my participation in this research will be kept as confidential as legally possible. In any publications that result from this research, neither my name nor any information from which I might be identified will be published without my consent. Audio tapes will be kept in a locked room. After typed transcripts are created, these tapes will be destroyed.

Voluntary Participation. I understand that participation in this study is voluntary. I understand that I am free to withdraw my consent to participate in this study at any time. Refusal to participate or withdrawal from the study will involve no penalty or loss of benefits. If I am a student, I understand that my class standing, grades, and status on an athletic team will not be affected by refusal to participate or my withdrawal from this study. I have been given the opportunity to ask questions about the research, and I have received answers concerning areas I did not understand..

Upon signing this form, I will receive a copy. I willingly consent to participate in this study.

Signature of Participant or Participant's Representative

Date

Signature of Investigator or Investigator's Representative

Date

Appendix B

Demographics Questionnaire for research on problem solving Subject # _____

(Experimenter asks participant questions in bold print and records responses on this sheet. Bold print indicates experimenter's questions.)

_____ 1) Sex: Female 1 Male 2 (circle)

Today's date: Month _____ Day _____ Year _____

What is your date of birth: Month _____ Day _____ Year _____

_____ 2) Age in years (subtract date of birth from today's date then round to nearest year)

_____ 3) **How many years of education have you received (highschool diploma = 12 years) (2 semesters = 1 year, don't count summers)**

_____ 4) **How would you rate your health, right now today, on a scale of 1 to 10 where 10 is excellent?**

_____ 5) **Is there any problem for which you are under the regular care of a physician?**

1 No 2 Yes If yes, which? _____

_____ 6) **Are there any prescriptions drugs that you take regularly?**

1 No 2 Yes If yes, which? _____

_____ 7) **Have you received training in formal logic? For example, have you taken Critical thinking , Symbolic logic, Mathematical logic, or Discrete Mathematics?**

1 No 2 Yes

Appendix C

Thinking-Out-Loud Procedures (Script adapted from Laipple (1991, Appendix D))

Instructions

“In this study, I would like you to think out loud as you are solving a set of problems and say whatever comes to your mind. Pretend I am not here and you are thinking out loud as you are solving the problems. Try to remember that I am interested in how you are getting to the solutions, in what you are actually thinking, and not just the solutions. It is important to me that you approach the problem the way you normally approach a problem, so try not to evaluate what you are saying— just say it.”

Reminders and Probes

Reminder: (after 5-6 second pause) “Keep talking.” Probe: If participant’s responses are ambiguous, nondirective probes will be used. For example, “Tell me more about why you said that?” “I’m not sure I understand, tell me what you mean.”

Balance Problem (adapted from Klahr and Siegler, 1978)

“These pictures show a balance scale with disks on it. All the disks [point to disks] weigh the same amount and the pegs they are on [point to pegs] are the same distance apart. [Show first balance scale] Imagine that this balance scale is being held steady. [pretend to hold the scale steady] What will happen when I let go? Think all your thoughts out loud” After the participant finishes, repeat this procedure with the rest of the balance scale pictures.

Participants will be reminded to think out loud during this task and will be praised for thinking-out-loud during and after their problem solving.

Stimuli for Balance Problem

Appendix D

Photo sets for Stereotype and Neutral Surgeon Problem

Appendix E

Task Instructions (adapted from Laipple (1991))

Surgeon version

I would like you to think out loud on this problem. One of these eight people is a surgeon, and it is your job to find out who is the surgeon. Now I happen to know which person is the surgeon, and I will give you an honest answer to any question you ask about these people, provided the question can be answered YES or NO. But you can only ask three questions

Here are the eight faces [present pictures as shown in Appendix D]. Your job is to try to find out which person is the surgeon. You can ask any three questions you want to, but I can only answer you YES or NO. Remember to think out loud.

[After answering the first question] You have two more questions left.

[After answering the second question] You have one more question left.

[After answering the third question] Which person is the surgeon?

Neutral version

I would like you to think out loud on this problem. One of these eight people was born in March, and it is your job to find out which one. Now I happen to know which person was born in March, and I will give you an honest answer to any question you ask about these people, provided the question can be answered YES or NO. But you can only ask three questions

Here are the eight faces [present pictures as shown in Appendix D]. Your job is to try to find out which person was born in March. You can ask any three questions you want to, but I can only answer you YES or NO. Remember to think out loud.

[Prompts after each question are same as surgeon except: Which person is born in March?]

Appendix F

Post-task questionnaire (adapted from Laipple, 1991)

After both versions are complete, ask the participant the following questions: [continue audio-taping verbal responses]

1. What information were you using to solve this problem?" (Any ambiguous questions will be followed by the prompt, "Tell me more.")

2. "On a scale from 1 to 7, how similar is this problem to any problem that you encountered before? --Where 1 is not very similar, 4 is somewhat similar and 7 is very similar." (If participant reports 'very similar' ask "What was the similar problem?")

3. Now I want you to administer the task to me. Here are the pictures I presented to you. Pretend you are the experimenter and administer the task to me. It is important that you administer the task in your own words, but try to keep the meaning of the task the same as when I gave it to you." (If participant gives instructions that are unclear ask, "Tell me more.")

Appendix G

Traditional Strategy Classification Method

(Based on classification criteria created by
Hayne Reese, 1999; personal communication).

Question 1

Participants can attempt to eliminate 0 to 4 pictures.

0 = Unidentifiable strategy (e.g., Did the surgeon go to a local medical school?)

1 = Item checking (e.g., Is this person the surgeon?)

2 = Pair checking (e.g., Is the surgeon one of these two? Is the surgeon an older male?)

3 = Partial constraint seeking (e.g., Is the surgeon one of these three people?)

4 = Constraint seeking (e.g., Is the surgeon a male? Is the surgeon on the right side?)

Note: Participants who attempt to eliminate 7, 6, or 5 pictures should be classified as item checking, pair checking, or partial constraint seeking, respectively. This is because the experimenter will answer these types of questions in the affirmative, thus eliminating 1, 2, or 3 pictures, respectively. For example, if a participant asked, “Is the surgeon not wearing any white?” on the first question, the experimenter should respond, “Yes”.

Strategy classification for Questions 2 and 3 depends on how many pictures are remaining after Question 1.

Question 2

8 remaining for Question 2: Classification same as for Question 1

7 remaining for Question 2:

0 = Unidentifiable strategy

1 = Item checking

2 = Pair checking

3 = Constraint seeking

4 = Constraint seeking

6 remaining for Question 2:

0 = Unidentifiable strategy

1 = Item checking

2 = Pair checking

3 = Constraint seeking

5 remaining for Question 2:

0 = Unidentifiable strategy

1 = Item checking

2 = Unidentifiable strategy (could be constraint seeking or pair checking)

4 remaining for Question 2:

0 = Unidentifiable strategy

1 = Item checking

2 = Constraint seeking

Note: Eliminating 2 is not pair checking in this case because 4 remaining for Question 2 must be due to constraint seeking on Question 1.

Question 3

8 through 5 remaining for Question 3: Classification same as for Question 2

4 remaining for Question 3

0 = Unidentifiable strategy

1 = Item checking

2 = Constraint seeking if Question 2 was constraint seeking; Pair checking if Questions 1 and 2 were pair checking.

3 remaining for Question 3

0 = Unidentifiable strategy

1 = Unidentifiable strategy (could be constraint seeking or item checking)

2 remaining for Question 3

0 = Unidentifiable strategy

1 = Constraint seeking

Appendix H

Data analysis using other methods for classifying strategy type

Total Number of Items Eliminated

Total number of items eliminated after three questions was calculated for each participant. The results of participants' strategies ranged from eliminating zero to eliminating seven items. Zero items eliminated reflects an uninterpretable strategy. Seven items eliminated reflects consistent constraint seeking. Other values may reflect a range of possibilities. For example, eliminating three items total could reflect (a) item checking on each question, (b) uninterpretable strategy on two questions and item checking on one question, or (c) Pair checking, item checking, and uninterpretable for the three questions, respectively. The sequence in which questions occurred is also not reflected by this scoring method. This method does, however, provide a quantitative range of items eliminated and odds of successful selection of the target. ANOVAs analogous to the original planned comparisons were conducted on these data.

Birthday first versus surgeon first. Planned comparisons ANOVAs using were conducted separately on data from the surgeon version and birthday version received first to test for age effects. Total number of items eliminated served as the dependent variable with age group as the independent variable in each analysis. Levine's test showed no significant heterogeneity of variance, $p > .05$. There was no significant difference in mean number of items eliminated on the surgeon version between younger adults ($\underline{M} = 5.25$, $\underline{SD} = 2.6$) and older adults ($\underline{M} = 4.00$, $\underline{SD} = 1.9$), $F(1, 30) = 2.48$, $p = .126$, $\omega^2 = .044$. There was also no significant difference in items eliminated on the birthday version between younger ($\underline{M} = 5.00$, $\underline{SD} = 2.5$) and older adults ($\underline{M} = 4.44$, $\underline{SD} = 2.5$), $F < 1$. Although the differences were in the hypothesized direction, the statistical

results differ from the results reported in the planned comparisons using the traditional classification method of strategy in that there was no significant age effect in the surgeon version.

The omnibus 2 (young versus old) x 2 (birthday versus surgeon) ANOVA was conducted. Levine's test found no significant heterogeneity of variance, $p > .05$. The ANOVA resulted in no significant age x version interaction or main effect of version, $F_s < 1$. The main effect of age was also not significant, $F(1, 60) = 2.38$, $p = .128$, $\omega^2 = .021$.

Surgeon first versus surgeon second. Planned comparisons were conducted separately on data from each age group using number of items eliminated on the surgeon version as the dependent variable and order of presentation as the independent variable. Levine's test showed no significant heterogeneity of variance, $p > .05$. Similar to the results reported in the original planned comparisons, older adults eliminated significantly fewer items who were given the surgeon version first ($M = 4.0$, $SD = 1.9$) than older adults given it second ($M = 5.40$, $SD = 2.0$), $F(1, 30) = 4.43$, $p = .044$, $\omega^2 = .097$. Younger adults did not differ in items eliminated between surgeon first ($M = 5.25$, $SD = 2.6$) and surgeon second ($M = 5.63$, $SD = 2.0$), $F < 1$.

The omnibus 2 (young versus old) x 2 (surgeon first versus second) ANOVA was conducted. Levine's test found no significant heterogeneity of variance, $p > .05$. The ANOVA resulted in no significant interaction, $F = 1$, and no significant main effect of order, $F(1, 60) = 2.91$, $p = .093$, $\omega^2 = .029$. The main effect of age was also not significant, $F(1, 60) = 1.83$, $p = .181$, $\omega^2 = .013$.

Birthday first versus birthday second. Planned comparisons were conducted on data from each age group separately with number of items eliminated on the birthday version as the dependent variable and order of presentation as the independent variable. Levine's test showed

no significant heterogeneity of variance for older adults, $p > .05$, but did show significant heterogeneity of variance for younger adults, $p = .012$. Thus, as recommended by Keppel (1991), a conservative alpha of .025 was adopted for the comparison involving the younger adults. Similar to the results reported in the original planned comparisons, older adults did not differ in items eliminated between the birthday first condition ($M = 4.44$, $SD = 2.5$) and the birthday second condition ($M = 4.56$, $SD = 2.2$), $F < 1$. Younger adults also did not differ in items eliminated between birthday first ($M = 5.00$, $SD = 2.5$) and birthday second ($M = 6.19$, $SD = 1.3$), $F(1, 30) = 2.91$, $p = .099$, $\omega^2 = .033$.

The omnibus 2 (young versus old) x 2 (birthday first versus second) ANOVA was conducted. Levine's test found significant heterogeneity of variance, $p = .004$. Thus, the alpha level was set at .025. The ANOVA resulted in no significant interaction, $F < 1$, and no significant main effect of order, $F(1, 60) = 1.49$, $p = .228$. The main effect of age was not significant at the adjusted alpha level, $F(1, 60) = 4.13$, $p = .047$, $\omega^2 = .047$.

Partial Constraint Seeking

Strategy type was also classified as either no constraint seeking or some constraint seeking. Participants' strategies that contained at least one constraint seeking question (as defined by the traditional classification method in Appendix G) were categorized as 'some constraint seeking'. All other strategies were categorized as no constraint seeking. The same multiway frequency analyses (logit) conducted in the original planned comparisons were conducted on these data. χ^2 s represent likelihood ratios and ps represent Fisher's Exact tests.

Birthday first versus surgeon first. Planned comparisons were conducted separately on data from the birthday and surgeon data using frequency of some constraint seeking versus no

constraint seeking as dependent variables and age group as the dependent variable. There was a significant effect of age on strategy use in the surgeon version. Specifically, on the surgeon version significantly fewer older adults used some constraint seeking ($\underline{N} = 7$) than younger adults ($\underline{N} = 13$), $\chi^2 = 4.97$, $p = .033$. There was no age effect in the birthday version. That is, there was no significant difference in number of older adults ($\underline{N} = 9$) and younger adults ($\underline{N} = 11$) using some constraint seeking, $\chi^2 < 1$.

Surgeon first versus surgeon second. Planned comparisons were conducted separately on data from each age group using frequency of some constraint seeking versus no constraint seeking on the surgeon version as the dependent variable and order of presentation as the independent variable. The number of older adults who used some constraint seeking when given the surgeon version first ($\underline{N} = 7$) compared to second ($\underline{N} = 12$), did not differ significantly, $\chi^2 = 3.31$, $p = .074$. Likewise, the number of younger adults who used some constraint seeking when they received the surgeon version first ($\underline{N} = 13$) did not differ significantly from those who received it second ($\underline{N} = 11$), $\chi^2 < 1$.

Surgeon first versus surgeon second. Planned comparisons were conducted separately on data from each age group using frequency of some constraint seeking versus no constraint seeking on the birthday version as the dependent variable and order of presentation as the independent variable. The number of older adults who used some constraint seeking when given the birthday version first ($\underline{N} = 9$) compared to second ($\underline{N} = 8$), did not differ significantly, $\chi^2 < 1$. Likewise, the number of younger adults who used some constraint seeking when they received the birthday version first ($\underline{N} = 11$) did not differ significantly from those who received it second ($\underline{N} = 14$), $\chi^2 = 1.69$, $p = .197$.

Appendix I

Loglinear models for Omnibus tests

Logit analyses commonly involve testing a model of the data with the observed data for goodness of fit. Models can be based on interactions, main effects or combinations of interactions and main effects. Expected frequencies are generated based on the hypothesized model and contrasted with the observed frequencies using a likelihood ratio statistic. Models that fit the data are those that produce expected frequencies that are not statistically different from the observed. That is, nonsignificant results indicate good fitting models. The likelihood ratio statistic that is used to test goodness of fit is labeled G^2 .

In the first two planned comparisons, three-way interaction model were used to generate expected frequencies and tested against observed frequencies. These models did not demonstrate good fit. Thus, a backward elimination method using SPSS version 9 was used to determine best fitting models for the data. The models that best fit each data set were the same as the results of the omnibus tests. Table 10 shows models with the best generating class and respective G^2 s. Recall that strategy type is considered the dependent variable. Thus, any variable that interacts with strategy can be conceptualized as a main effect. For example, the age x strategy interaction is a main effect of age on strategy type. Models that only include strategy indicate that frequency of strategy type (no constraint seeking versus constraint seeking) significantly differed from 50:50 but did not differ as a function of any other variable.

Table 10

Best Fit Models and G^2 for Logit Analyses of the Omnibus Tests

	Strategy classification method					
	Predominant			Consistent		
Planned comparison	Model	G^2	p	Model	G^2	p
Surgeon vs. birthday	age x strategy	1.12	.891	strategy ^a	4.68	.586
Surgeon 1st vs. surgeon 2nd	no factors	11.10	.134	order x strategy	4.55	.337
Birthday 1st vs birthday 2nd	no factors	5.44	.606	no factors	4.58	.711

^a Models that only include strategy indicate that frequency of strategy type (no constraint seeking versus constraint seeking) significantly differed from 50:50 but did not differ as a function of any other variable (i.e., age, order, and version).

Appendix J

Method for Classification of Interpretation: Surgeon problem only

Participants' audio-taped responses were transcribed into typed protocols and saved using a word processor. Next, several elements of the protocols were removed before interpretation coding took place. (1) responses to the birthday version, (2) experimenter's responses to the each of the three questions, and (3) participants' 'official' three questions. Responses to the birthday task were removed because interpretation of birthday task was not of interest for this study. The participants' three questions were removed because 'interpretations' have been hypothesized to control or affect strategy use. Strategies, in this study, were understood and defined by the type of three questions and the pattern of these questions that participants used. Thus, for the construct of interpretation not to be circular or confounded with strategy, the coder must be 'blind' as to which strategy (three questions) the participant used. That is, the coder must not be aware of the participant's three questions so classification of interpretation is deduced independent of strategy. Also, if interpretation truly affects strategy use, interpretation would come before strategies are produced. If strategies were used to classify interpretation, the construct of interpretation would be circular.

Participants' answers to the probe, "What information were you using to solve this problem?" were left in the protocols and were used for classifying interpretation. Ideally, these statements would not be used to classify interpretation because, often, the participants simply described the strategy they used. However, thinking aloud prior to the questions was often sparse and sometimes nonexistent. Thus, the statements about information used to solve the problem had to categorize interpretation be used in the current study. This, method likely inflated the

correlation between ‘interpretation’ and strategy use. However, no such correlation was found even though the retrospective reports were also used to classify interpretation. Future studies should include probes that occur before participants solve the problems. These probes should ask the participant to reconstruct the problem, state the rules of the problem (i.e., what is the solver allowed and not allowed to ask about), and state the goal of the problem. Protocols produced by such probes would be ideally suited for classifying a participants’ interpretation of the initial state, goal state, and parameters of the processes allowed to be used to get to the goal state from the initial state.

In the present study, participants’ reconstructions of the problems (which occurred after they solved the problem) were short and did not vary greatly. Joseph Laipple reviewed the reconstructions and indicated that they were too “sparse” to be used to classify interpretation (personal communication, 1999). Thus, reconstructions were removed from the protocols. The remaining partial protocols were read by the principal investigator and examined for the content. Statements about characteristics of people not mattering or that everyone has an equal chance of being a surgeon were coded as Logic-based interpretations. Examples include: “. . . narrowing down the odds”, “. . . cut down who it could be”, “. . . eliminate as many people as I can”, “Elimination. Cutting out some of the possibilities.”, “Well I figured that I could divide the group in half”, “. . . the quickest way to single out probably half the people”, “. . . narrow the people down into categories”, “I found out that worked narrowing it down the last way”, “The brick background, if it didn't have it then it cut it down in half then the equal proportion of women to men”, “lower it down as far as I could if you can cut it in half is the best you can usually do”, “I could ask if it was a male or female that eliminated two rows, and then I asked

one of the two rows and one of the two columns and that would narrow it down”, “You can’t look at somebody and tell if they’re a doctor “, “there are a few different ways I could go about eliminating”, “. . . take half of the people off the list”, and “I’ve got to figure out how to eliminate all but one person”.

Protocols containing statements mentioning who was likely or unlikely to be a surgeon were coded as experienced-based interpretations. Examples include, “This fellow, he looks pretty serious. Doctors are serious I guess.”, “I think this fellow looks too young”, “I could not identify who may be the physician in that bunch”, Surgeon’s are kind of egocentric, as far as his looks.”, “I think, I don’t think my intuitive way sort of feel that she would have the spunk to be a surgeon”, “I think they all appear to be able to be surgeons”, “If I could see their hands it would show uh sometimes you can go by dirty fingernails”, “I was just trying to look and see if there’s a surgeon type”, “Well I arrived at the fact that I had absolutely no clue there was no pattern whatsoever and so the only thing I could do was take a wild ass stab”, “Well my first thought was which man is it, but of course, it could be a woman.”, unless they’re in proper clothing”, “I’ve got instinct you know, whatever reason.”, and “It would have been easier if you had shown us their hands.”

Protocols that contained both types of statements were coded as pragmatic interpretations.

Vita

Born September 8th, 1969

Birthplace Madison, Wisconsin

Degrees:

B.A. Psychology major/Philosophy minor Northern Michigan University, 1992

M.A. Developmental psychology West Virginia University, 1996

Ph.D. Developmental psychology West Virginia University, 1999