12-2010

Complex Decision Support for Older Adults: Effects of Information Visualization on Decision Performance

Margaux Price
Clemson University, margauxmae@gmail.com

COMPLEX DECISION SUPPORT FOR OLDER ADULTS: EFFECTS OF INFORMATION VISUALIZATION ON DECISION PERFORMANCE

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Applied Psychology

---

by
Margaux M. Price
December 2010

Submitted to:
Dr. Richard Pak, Committee Chair
Dr. Leo Gugerty
Dr. Christopher Pagano

ABSTRACT

Older adults are faced with complex decision tasks that impose high working memory demands. A representative task is choosing a prescription drug plan from a multitude of options that must be evaluated along many factors. The combined effect of the quantity of complex information, and reduced working memory capacity puts older adults at a disadvantage. However, research with younger adults suggests that the working memory burden of decision tasks can be reduced using well-designed, graphical decision aids (i.e., environmental supports). The current study examined the use of environmental supports to support complex decision-making for older adults. Two experiments were conducted; experiment 1 assessed two information visualizations (color and size) on their ability to minimize the working memory demands of the task. Results from experiment 1 suggest that the color information visualization does in fact minimize working memory demand by replacing cognitive comparisons with perceptual comparisons. The second experiment validated the efficacy of the color information visualization in an older adult group. Findings suggest that the use of color to visualize information can successfully ameliorate working memory demand for direct comparisons, but not for complex integration tasks. Finally, the results suggest that information visualizations that rely on perceptual abilities rather than cognitive abilities may help improve older adults' decision making accuracy.

ACKNOWLEDGEMENTS

I would like to first thank my parents for all of their support and encouragement, without it this wouldn’t have been possible. I am forever thankful for my little brother, Chase, for always motivating me to leave big footprints for him to fill. I am also grateful for all of the sacrifices my parents, Chase, and Emma have made for me.

I would also like to thank my advisor and committee chair, Dr. Richard Pak for his mentoring, patience and guidance. In addition, I would like to thank my committee members, Dr. Leo Gugerty and Dr. Christopher Pagano, for all of their contributions to this project and for helping me become a better researcher. I would also like to acknowledge Dr. Eric Muth for his continued guidance throughout my time at Clemson.

I am indebted to my officemates and friends, Nicole Fink and Jeremy Mendel for their willingness to help me, their encouragement, positivity, and hope. Finally, I’d like to thank my ESPY family and my friends for always being there for me and for keeping me grounded.

TABLE OF CONTENTS

Table of Contents (continued)

LIST OF TABLES

LIST OF FIGURES

List of Figures (continued)

INTRODUCTION

Today, older adult consumers are faced with an overwhelming number of options when it comes to making healthcare decisions. Choosing a prescription drug plan exemplifies how daunting some healthcare decisions can be. A search for prescription drug plans within a given zip-code using the Medicare.gov website returns a list of over 40 plans. Comparing and contrasting so many options is a complex task especially for older adults. The task of choosing a Medicare drug plan on the web is affected by issues such as the design (usability) of the site, the knowledge and experience of the user, and the user's cognitive abilities. Even if older adults are able to successfully navigate the site, choosing the *optimal* plan requires thinking about each plan's long-term costs. The older user must think about how the initial cost increases over time. This is not a straightforward task since the long-term costs are affected by other factors such as gap coverage, budgeted allowance for medical care, and out of pocket expense should they exceed their coverage.

A recent usability evaluation of the Medicare website showed that older adults were unable to successfully choose a prescription drug plan for a given medication regimen (Czaja, Sharit, & Nair, 2008). Example problems were general difficulty with navigating the site, frustration, and the inability to locate desired information (Czaja, Sharit, & Nair). Insurance and medical jargon (e.g., "gap coverage", drug sharing, etc) may have further complicated the task (see Appendix A for definitions). Comprehension of jargon and relating qualitative values (e.g., satisfaction ratings) to quantitative values (e.g., dollars) all requires reasoning ability. In sum, the seemingly simple task of choosing an optimal plan is one that potentially places very heavy demands on working memory and attention.

Choosing an appropriate drug plan is a complex task (see Appendix B). Differentially scaled factors must be considered to choose a plan that is optimal for an individual. For example, the *monthly premium* is how much a person will pay monthly, while the *annual deductible* is the amount that must be paid before coverage begins. Finding the yearly cost of a plan requires multiplying the monthly premium by 12 months, adding the annual deductible, and remembering this number so it can be compared to the other 40+ plan choices. Factors such as satisfaction ratings (based on a 5-point scale) or drug cost sharing (expressed as a percentage or dollar amount) are in units that are not directly comparable. Thus, each of these values (total cost, satisfaction rating, and drug sharing percentages) must be remembered separately for accurate comparisons between plans. Decision makers may not do all of the calculations by hand and may write down some information regarding the plans that are viable options. However, even if they are able to eliminate half of the plans (i.e., reduce from 40+ to 20) and compose a list of the viable options, both older and younger adults may still have a difficult time choosing the most optimal plan (Tanius, et al., 2009).

Not being able to choose the best drug plan can have negative consequences on an older adult's health and financial state (Hsu, et al., 2008). If the chosen plan does not provide sufficient coverage, an older adult may be forced to decide whether to continue with the medication regimen recommended by their doctor and incur out of pocket expenses, switch to cheaper medications, or take the health risks of discontinuing the regimen altogether. The plan with the most coverage may not be the best choice either because it may exceed the consumer's budget and create an unnecessary financial burden.

Trying to make optimal decisions in the face of uncertainty and with a large amount of inputs can be a very working memory-demanding task. *Working memory capacity* refers to the amount of information one can temporarily store and manipulate at any given time (Baddeley, 1986). The amount of information that must be stored or manipulated is the task's *working memory load*. If the task's working memory load exceeds one's working memory capacity, then task performance may be degraded or impossible. This capacity limit is central to one's ability to process information and thus make a decision.

The Information Processing Model of Decision Making

Making a decision is a multi-stepped, cognitively demanding task (see Figure 1). Choosing a prescription drug plan on the basis of cost first requires that the decider perceive the appropriate cues (monthly premiums, coverage in the gap), while ignoring irrelevant cues (Medicare ID numbers or contact information). After selectively attending to appropriate cues, the information is manipulated in working memory where hypotheses or potential outcomes are generated (e.g., plans with a low monthly premiums and low deductibles have less coverage). A more detailed account of the process is provided in the next section.

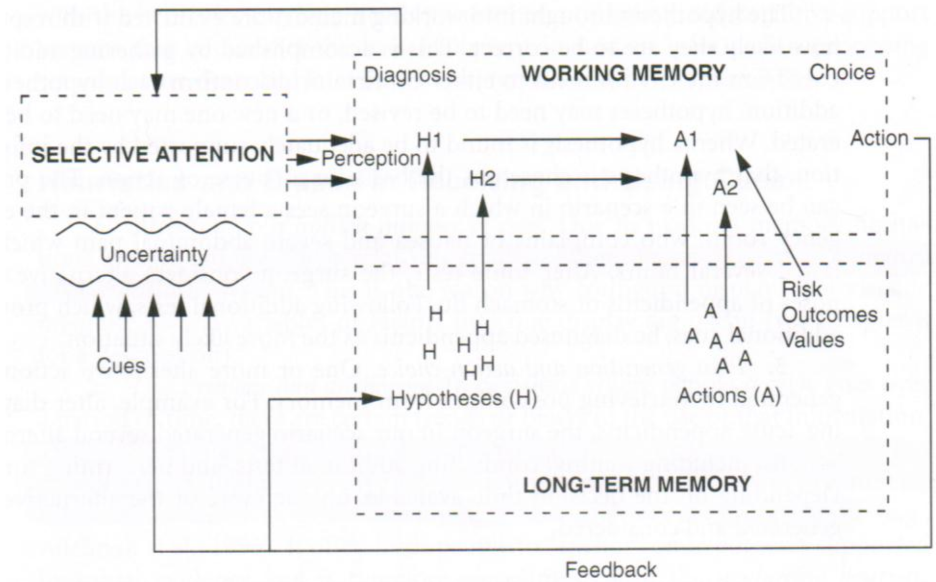


Figure 1. Information processing model of decision making taken from Wickens, 2004.

Step One: Cue Selection and Integration

In the first step, cues relevant to the decision are first perceived. Attentional limitations force the user to filter cues relevant to the decision goal from the irrelevant cues by selectively attending to only some of the information present. Cues may be selected based on their diagnosticity (amount of information the cue provides), reliability (trustworthiness of information), and salience (physical properties such as volume, color, and shape). For example, the salience of the cue (e.g., brightness, size, loudness) can, in some cases, override a cues' diagnosticity or reliability. Consider an example where information has high relevance but low salience (e.g., small size, low contrast) and thus fails to capture attention and is not available in working memory. Conversely, a cue with high salience will capture attention and enhance goal-driven tasks (e.g., finding the lowest monthly premium).

After cues are selected they are then integrated. Integration refers to how each piece of information (from all perceptual systems) is compared to other information in order to form a meaningful interpretation of the state of the system or environment. Although working memory limits the amount of information used to form this interpretation, information that shares similar perceptual or semantic features may be grouped together into object-like "chunks" or visual clusters that enable pattern recognition (Miller, 1956; Ratwani, Trafton, & Boehm-Davis, 2008). Information may be chunked together based on color, shape, meaning, spatial proximity or other properties (e.g., Gestalt principles) pre-attentively or automatically (without the need to selectively attend to each cue individually). This perceptual integration process may help facilitate later processing of more information with less effort. If information is perceived as part of an object (or chunk) rather than many separate objects it reduces the number of items that need to be held in working memory. Chunking also reduces the need for explicit cognitive integration - the effort applied to the formation of explicit groups. Reducing working memory demands at this step allows more resources to be devoted to processing in step 2 or 3 of the model.

Step Two: Generation of Hypotheses

Once appropriate cues have been attended to, they enter working memory where they can be manipulated. The individual will interpret this information, compare and contrast information (from step 1), and use experiences from long term memory (LTM) to predict potential outcomes of each decision option. This is another step that may be error-prone because it is dependent on the previous step (where optimal cues were not attended to) but also because irrelevant or incomplete information may be recalled from LTM. Using incomplete or irrelevant information from LTM (e.g., the plan your neighbor just bought) rather than evaluating all options may lead to a poor decision.

Retrieving information from LTM to assess the situation in WM also increases the task's working memory demand.   An example of this integration process can be found in task 3.1 of the task analysis (Appendix B). The decision maker draws on long term memory to remember current drug costs and then integrates that information with the potential coverage options in working memory. At the same time, the decision maker has to remember the coverage gap amount (long term memory), remember how much the drug costs without insurance (long term memory), figure out how many months they will not be covered (working memory), and add this amount to the out-of-pocket expense total (working memory).

Hypotheses about the long-term outcome of specific plan choices are generated and compared. Choosing a prescription drug plan requires several hypotheses for each plan; one for cost and the effect on personal budget (task number 6.0 in Appendix B), one for satisfaction (task number 5.0 in Appendix B), and another one for how nationwide coverage might affect them (task number 4.0 in Appendix B). All 3 hypotheses will need to be compared between each plan, which may be impossible with over 40 plans and limitations of working memory capacity.

Step Three: Integration of Outcomes and Action Selection

In this step, the decision maker tries to determine which option will produce an outcome that best meets the goal. Possible actions are generated by recalling experience with, or knowledge of similar situations from long term memory, and then integrating it with information from the current situation (in working memory). This process allows the decision maker to generate possible outcomes and consequences of taking a specified action (in this case, choosing one plan over another). If a plan is chosen for its low monthly premium but also has a low satisfaction rating, the decision maker has to consider the potential implications of both attributes together. Similar to the previous steps, this step is error-prone because the determination of possible outcomes may be influenced by past experiences and WM capacity limits the number of comparisons that can be made simultaneously.

Step Four: Monitoring and Evaluating Actions

Once an action is selected and carried out (a decision is made), the outcome is monitored and evaluated against new cues or information, and new hypotheses about the state of the system are formed. Working memory capacity limits the amount of new information selected and compared to the current state of the system and any subsequent actions needed.

The Use of Heuristics and Subsequent Biases

Heuristics or "rules of thumb" are shortcuts that people may use to make decisions quickly and with little effort (Tversky & Kahneman, 1974). An everyday example of such a heuristic is buying a Toyota Prius without considering other fuel-efficient options because of its high salience and availability to mind. Not considering other options reduces the task's WM demand and allows decisions to be made more quickly. A decision made using incomplete information is a biased decision. A biased decision may lead to the acceptance of an option that isn't the most optimal choice. When heuristics are based on inaccurate information individuals may make poor decisions. An example is choosing a brand name product over a non-branded product with the assumption that the quality is better when a closer examination reveals the two products are exactly the same. While there are many heuristics that may lead to biases (see Glovich, Griffin, & Kahneman, 2002 for a review), the following are examples of how a few might be utilized to simplify a decision task in the context of choosing a prescription drug plan.

Representativeness Heuristic

The representativeness heuristic is the use of *prototypes* to make judgments. Decision makers judge a set of cues based on how similar they match a prototype or category from previous experience (stored in long term memory). The decision maker relies on the probability that a certain group of information or cues generally describes a situation or system. If for example the decision maker wants to purchase an environmentally friendly hybrid car, he or she might immediately consider a Toyota Prius as the best choice because of its popularity in the media as the prototypical environmentally friendly car. There could be other cars available that are better for the environment, but the decision maker ignores these and purchases the Prius. Using this heuristic reduces working memory demands because instead of generating multiple hypotheses about how different vehicles may affect the environment (step2); the decision maker selects the prototypical hybrid car. Although the decision led the user to choose a hybrid car that is better for the environment, the decision was biased toward the prototype. In other words the decision maker made the inaccurate judgment by evaluating an unimportant factor: how representative the prototype is to the current situation.

One consequence of using the representative heuristic is a tendency to ignore base rates of phenomena (Tversky & Kahneman, 1974). For example, each prescription drug plan differs by some combination of the attributes of a plan (monthly premium, annual deductible, or gap coverage). Ideally, these attributes will be weighted and compared one-by-one in order to choose the optimal plan. This is a working memory intensive task given the large number of plans (as much as 40) and the number of comparisons between plans and attributes that need to be evaluated. Instead, the decision might be biased because of a prior experience where a plan with the higher cost had the most coverage. A biased choice is one that, on the surface, appears to be best plan because its total yearly cost is more expensive and is thus expected to have more coverage (assuming all plans that are expensive have more coverage), when in reality the gap coverage may be much lower (so there is lower coverage).

Availability heuristic

The availability heuristic is the use of information that comes readily to mind (Tversky & Kahneman, 1973). The fluency by which the information comes to mind is misjudged as accuracy or reliability and is used in subsequent decision-making. The availability heuristic is manifested as judging more familiar and salient information as more probable or truthful (Tversky & Kahneman, 1974), or overestimating the frequency of two events occurring at the same time because the experience of them occurring together came to mind first (Chapman & Chapman, 1969). An example of the availability heuristic would be if an older adult purchasing a drug plan bases their decision on one attribute (e.g., gap coverage) because they recall a neighbor who neglected to purchase any gap coverage and wound up paying a lot of money out-of-pocket. They might not consider other factors that would incur out-of-pocket expenses and narrow down the choices to only plans with the highest coverage instead of calculating the overall cost of the plan (using the other attributes). The decision task's working memory demands are reduced because eliminating the plans with less coverage reduces the number of comparisons the consumer will have to make.

Biases and Decision Making Strategy

Two decision making strategies used to reduce resource demands are satisficing (Simon, 1955) and elimination by aspects (Tversky, 1972). For both strategies, the decision maker must determine evaluation criteria. Biases introduced in the criteria development process can lead to a poor decision. For example, in the satisficing strategy, options are evaluated along criteria until an acceptable option is found without necessarily considering all options. Satisficing can be an efficient strategy because additional effort is not expended considering all possible options.

Elimination-by-aspects is used to eliminate all choices that do not meet threshold for a particular aspect or criteria (e.g., in the drug plan decision a plan would be eliminated if it doesn't meet the threshold - less than $40 per month). In the example of the availability bias, Mary might set her threshold for the cost of the criteria to be only plans that offer coverage of costs in the gap (see Appendix A for an explanation). If she uses a satisficing strategy, she would look at the options one by one until she found one plan that minimally matched her criteria. When she found a plan that offered coverage in the gap, she would choose that plan and not look any further. If she were to use an elimination-by-aspects strategy, she would go through all of the plans and eliminate all plans that do not meet her criteria of having coverage in the gap. Thus, if the criteria that determine whether an option is either "good enough" or eliminated are biased, then the decision will also reflect that bias. Again, this can be an efficient strategy because it allows the decision-maker to focus on a few criteria at a time, rather than consider all criteria.

Decision Making and Aging

Older adults' reduced working memory capacity (Salthouse, 1991) limits the number of integration and comparison tasks that can be made at a given time and thus may affect their ability to make optimal decisions (Mata, Schooler, & Rieskamp, 2007). Age-related limitations may lead older adults to exhibit more frequent heuristic-based decision-making. For example, one study examining age differences in decision making strategy found that older adults were more likely than younger adults to use a satisficing heuristic in a financial decision making task (Chen & Sun, 2007). In their study, younger adults chose the relatively more involved strategy of remembering and comparing up to six monetary offers (higher working memory load), while older adults chose the less effortful strategy of memorizing one offer (lower working memory load). Surprisingly, despite the varying strategies between younger and older adults (and older adults lower working memory capacity), there were no age differences in performance. Of course, this represents a case where the use of heuristics leads to effective decision making. As mentioned previously, heuristics will not always lead to an optimal decision.

Although older adults are sometimes successful in adapting their strategy to meet the task demands, they tend to perform worse on tasks that require integrating information (comparing more than two pieces of information), rather than extracting information (finding one piece of information; Finucane, et.al, 2002). Comparing information that is presented in different units (e.g., monetary units and satisfaction ratings) may make the task more difficult for older adults (Finucane, et. al, 2005, Tanius, et. al, 2009). Choosing a prescription drug plan exemplifies this task; one must compare multiple cost values and multiple satisfaction ratings among many possible plans. Older adults tend to commit more errors and have more difficulty comprehending information than younger adults when the task requires integrating information (Finucane, et. al, 2005) among many choices (Tanius, et. al, 2009).

Decision Aids and Environmental Support

Aids that specifically reduce working memory demands are called environmental supports (Craik, 1986). Environmental supports (ES) can improve task performance for older adults (Morrow & Rogers, 2008) by reducing task demands or supporting the use of existing resources (Morrow & Rogers, 2008). An example of a successful ES is in a study that examined navigation efficiency in an automated voice menu system (Sharit, Czaja, Nair, & Lee, 2003). In study 1, Sharit et al. found that older adults' had lower performance than younger adults during a complex auditory navigation task, with measures of working memory contributing the most to the variance. A follow up study was conducted to examine performance with the use of a graphical aid (a form of ES). The ES was designed to reduce task demands by allowing the user to rely on the external environment (the graphic aid) for information instead of working memory (internal components). The graphical aid displayed a hierarchy of the automated voice menu system allowing users to see direct relationships between menu items rather than having to remember the steps they took (a working memory and spatially demanding task). No age differences in performance were found in the graphical aid condition, suggesting that providing an environmental support, designed to reduce working memory demands, enhanced performance for older adults (Sharit, et al., 2003).

Current Literature

Several studies with younger adults have shown that providing an ES reduces WM demand by facilitating visual search and automatic perceptual processing of information (Lohse, 1997; Ratwani, Trafton, & Boehm-Davis, 2008) for example, when color is used to facilitate automatic visual integration of related information into meaningful "chunks" (Lohse). In that study, participants in the color condition no longer had to shift attention between the legend and the graph, nor did they have to remember the items in the legend or their locations within the graph. Instead, participants were able to allocate memory and attention resources on making meaningful comparisons between these chunks, rather than on their formation. Ratwani, Trafton, and Boehm-Davis further examined the cognitive process used to successfully integrate and extract information from graphs, and theorized that when information within the graph is already organized into visual “clusters”, 1) less effort is needed to group similar information together, 2) reducing the working memory demand to the task. When similar information is grouped together (e.g., in this study it was counties with similar attributes were grouped into visual clusters using color), the user can focus attention on the differences between the groups, rather than first actively integrating information into clusters.

Reducing the need for the effortful comparison of information may allow the user to allocate more resources to other steps in the decision making process (Ratwani, Trafton, Boehm-Davis, 2008). Older adults may benefit from a decision aid designed to shift information from working memory to an external memory aid where it can be perceived by the relatively age-insensitive pre-attentive visual perceptual system (Plude and Doussard, 1989). Although some perceptual abilities decline with age (i.e., visual acuity, hearing loss), the ability to detect and process meaning of a single target feature (e.g., color, shape) does not decline with age (Plude & Doussard, 1989). For example, a multi-ordered brightness scale allows people to make comparisons between choices without having to process a number and assign it meaning before serially moving onto the next choice (Breslow, Ratwani, and Trafton, 2009). Instead, meaning is automatically processed using perceptual features (e.g., darker green may represent a higher number than a lighter green - the scale is based on the color density). In addition, it is much faster to search for a color singleton than to find a number target (Treisman, 1982). This suggests one avenue of providing an environmental support-based decision-making aid: shifting the working memory burden to the perceptual processing system by eliminating the need to comprehend and compare each option semantically and instead comparing the information perceptually.

The display design principles found in Appendix C provide some suggestions for altering tasks/displays to reduce overall cognitive processing demand. For example, the proximity compatibility principle suggests that information that needs to be processed or integrated should be placed close together to facilitate more efficient processing (Wickens & Carswell, 1995). Close proximity of information facilitates processing because it reduces the need to switch attention between two pieces of information. Switching attention requires the user to remember the first piece of information, consciously direct attention to another area of the display, extract another piece of information, integrate and then finally interpret the information. Thus, keeping information close in proximity can also help reduce the need for executive attention thereby reducing the working memory load of the task. The purpose of the current study is to extend Lohse's (1997) and Ratwani, Trafton, & Boehm-Davis's (2008) findings to design of information visualizations that reduce working memory demands. Reducing working memory demands is expected to reduce the likelihood of using heuristics which may lead to better decisions.

Overview of the Current Study

The goal of the current study is to examine whether older adult decision making performance can be enhanced by the use of graphical decision aids designed to reduce WM demands. Reducing WM demands is expected to lessen reliance on heuristic strategies, and improve decision quality. Decision quality is measured as to how well the choice met the criterion in the question. The assumption is that when the decision making task is reduced from cognitively complex to relatively easy, decision makers would not need to rely on heuristics and would consider all information. The first experiment was designed to assess information visualizations that reduce the working memory demands of the task. The second experiment was conducted to validate the efficacy of the information visualization in an older adult group.

EXPERIMENT 1:
DESIGNING INFORMATION VISUALIZATIONS THAT REDUCE WORKING MEMORY DEMAND

The goal of Experiment 1 was to test alternative information visualizations on their ability to work effectively (i.e., reduce the chance of bias decision-making) under conditions of high working memory load. A concurrent memory load is primarily meant to induce people into heuristic decision making (and thus is a rudimentary simulation of older adult decision making). The actual design of the alternative information visualizations was based on existing human factors display design principles (e.g., proximity compatibility principle). How well a particular information visualization reduced working memory demands was examined in a dual-task paradigm. Younger adults performed a primary decision making task while also performing a secondary working memory task. A concurrent task paradigm was used to constrain younger participants' working memory capacity to simulate the conditions an older adult with a lower working memory capacity may experience and to "force" them to utilize heuristics.

The information visualizations are expected to facilitate visual integration and perceptual comparisons in place of effortful cognitive integration and comparisons, thus reducing the WM demands of the task (Lohse, 1997; Ratwani, Trafton, & Boehm-Davis, 2008). Reducing the tasks WM demand was predicted to improve decision quality and accuracy because the decision maker could then consider more of the options and rely less on heuristics. In addition, decision-making speed is predicted to be faster in the info-vis conditions than the table conditions because perceptual comparisons (e.g., size and color) don't require higher level cognitive processing (Lohse, 1997; Treisman, 1987).

For the two levels of task difficulty, it was predicted that quality, accuracy, and task time would be better in the low difficulty task compared to the high difficulty task because the high difficulty task requires more comparisons (either visual or cognitive). Finally, task performance on all dependent measures was expected to be worse for both levels of difficulty with the addition of the WM task; however the addition of this task would negatively affect the table condition more so than the info-vis conditions because the info-vis conditions have a lower WM demand than the table condition.

METHODS

Participants

Thirty-four younger adults were recruited from psychology courses and all subjects received course credit for participating. Groups of 1 to 4 participants were tested simultaneously, however participants worked independently at separate workstations. The only exclusion criteria for participation were the presence of color-blindness and the inability to read a computer screen.

Design

Experiment 1 was a 3 (decision aid: table, color information visualization, size information visualization) x 2 (task difficulty: low, high) x 2 (WM task condition: single, dual) mixed design (see Figure 2), with decision aid as the between subjects variable and task difficulty and WM demand as the within subjects variables.

Participants made decisions over 40 trials. The trials were organized around 8 blocks of 5 questions per block. A randomized blocked design was utilized for questions of varying task difficulty and WM demand. The questions within each block were also randomly presented.

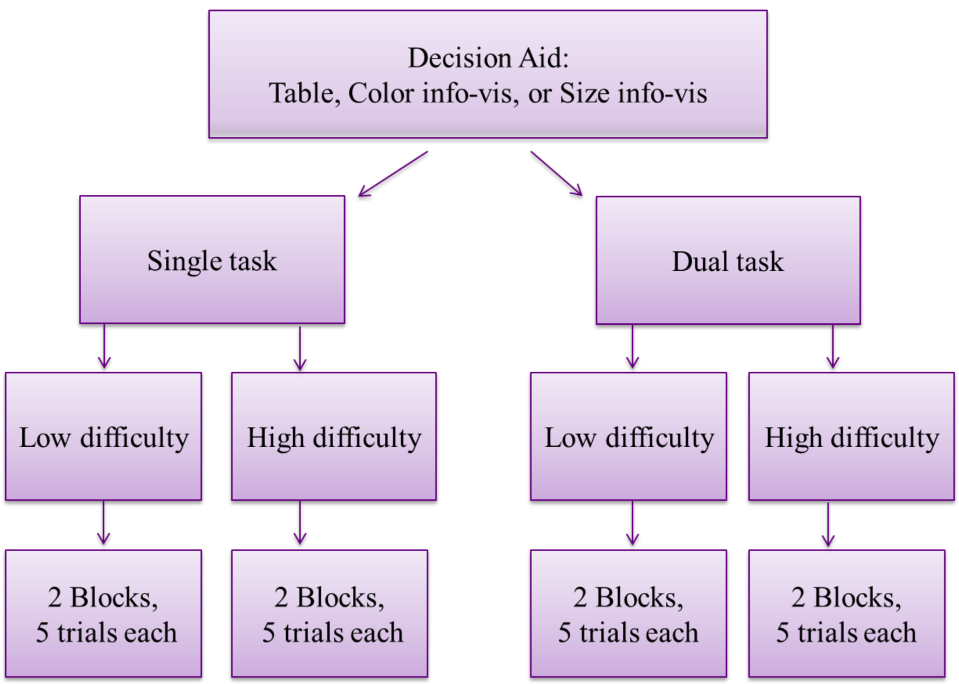


Figure 2. Experimental blocks.

Dependent measures were decision accuracy (sum score of number correct), decision quality (sum score of scaled decision ratings), decision task time (in seconds), and n-back accuracy score (sum score of number correct).

Independent Variables

Decision aids.

The table condition was a replica of the table found on the Medicare website (at the time of the study proposal). The table (shown in Figure 3) included a row for each of the fifteen prescription drug plans and columns for four of the plan's attributes.

Which plan has the highest satisfaction rating?

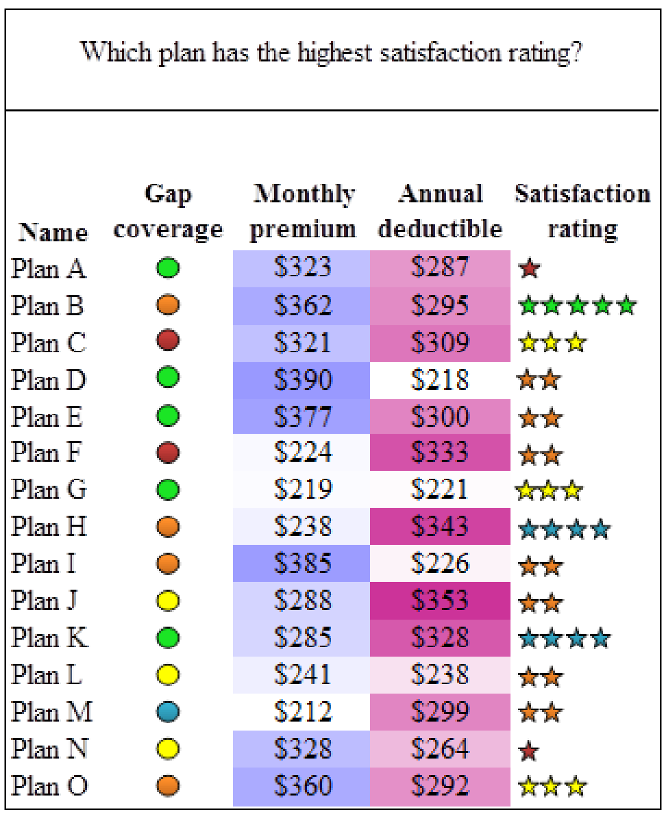
Name	Gap coverage	Monthly Premium	Annual Deductible	Satisfaction Rating
Plan A	All generics	$323	$287	1.0 out of 5 stars
Plan B	Some generics	$362	$295	5.0 out of 5 stars
Plan C	No gap coverage	$321	$309	3.0 out of 5 stars
Plan D	All generics	$390	$218	2.0 out of 5 stars
Plan E	All generics	$377	$300	2.0 out of 5 stars
Plan F	No gap coverage	$224	$333	2.0 out of 5 stars
Plan G	All generics	$219	$221	3.0 out of 5 stars
Plan H	Some generics	$238	$343	1.0 out of 5 stars
Plan I	Some generics	$385	$226	2.0 out of 5 stars
Plan J	Many generics	$288	$353	2.0 out of 5 stars
Plan K	All generics	$285	$328	4.0 out of 5 stars
Plan L	Many generics	$241	$238	2.0 out of 5 stars
Plan M	Most generics	$212	$299	2.0 out of 5 stars
Plan N	Many generics	$328	$264	1.0 out of 5 stars
Plan O	Some generics	$360	$292	3.0 out of 5 stars

Figure 3. Example layout of a low difficulty decision task in the table condition. Fifteen plan options are shown with four plan attributes (gap coverage, monthly premium, annual deductible, and satisfaction rating).

The information visualization conditions were created by adding graphics instead of (or in addition to) text to represent specific attributes. Two information visualizations (shown in Figures 4 and 5) were created utilizing well-accepted display design principles (see Appendix C.) (e.g., proximity compatibility principle, color gradients, pictorial representations, and redundancy). Our task analysis (Appendix B) illustrated the working memory-intensive nature of this task (steps 6-10). The information visualizations used in this study were designed to alleviate the working memory intensive parts of the task specifically by converting them into easier perceptual tasks using color and size manipulations.

Figure 4 shows the color information visualization (color info-vis) in which multi-colored scales (heat map color scale) replace the categorical gap coverage text. The same multi-colored scale was used in the stars that replace the number scales for satisfaction ratings. Multi-colored scales have been shown to be facilitate identification tasks – where one has to select a target value represented by a color (e.g., identify the plans that have gap coverage level of all generics – represented by the color green), and in cases where a particular absolute value (i.e., all generics) is more important than a relative value (i.e., the plan with the lowest amount of coverage) (Breslow, Ratwani, & Trafton, 2009). In the current study, the multi-colored scale was used to represent the five specific categories of both gap coverage and satisfaction ratings and these categories were absolute, not relative to one another (e.g., "all generics" was always the highest level of gap coverage, but "some" or "many" generics are not proportionate to each other).

Brightness ordered scales (same color is used but lightest color gradient is the lowest value and the darkest color is the highest value) were added to dollar amounts in both the monthly premium and annual deductible columns. Brightness ordered scales have been shown to be superior for comparisons of relative value (Breslow, Ratwani, & Trafton, 2009) – where all values are compared to one another (e.g., which plan has the lowest or highest monthly premium). These color manipulations were added to facilitate more perceptual comparisons rather than effortful cognitive comparisons, thus reducing WM demand.



Which plan has the highest satisfaction rating?

Name	Gap coverage	Monthly premium	Annual deductible	Satisfaction rating
Plan A	●	$323	$287	★
Plan B	●	$362	$295	★★★★★
Plan C	●	$321	$309	★★★
Plan D	●	$390	$218	★★
Plan E	●	$377	$300	★★
Plan F	●	$224	$333	★★
Plan G	●	$219	$221	★★★
Plan H	●	$238	$343	★★★★
Plan I	●	$385	$226	★★
Plan J	●	$288	$353	★★
Plan K	●	$285	$328	★★★★
Plan L	●	$241	$238	★★
Plan M	●	$212	$299	★★
Plan N	●	$328	$264	★
Plan O	●	$360	$292	★★★

*Figure 4.*Color information visualization (color info-vis).

Unlike the first info-vis, the second information visualization (size info-vis) illustrated in Figure 5 used area and size characteristics to help shift WM demand to the perceptual system. Bar graphs and pie charts are two commonly used graphing methods that use size comparisons to communicate relative differences of data points visually. Bar graphs are particularly useful in displaying differences in a dependent variable over levels of an independent variable (Gillan et. al, 1998). For this task, the decision maker needs to make comparisons along the monthly premium amount (a dependent variable) across multiple plans (levels of the independent variable). Bar charts were used in addition to the dollar amounts, providing redundancy and not forfeiting the tables' superiority in comparing exact values (Meyer, Shinar, & Leiser, 1997). Individual stars were used to create the bars that represent the satisfaction ratings. For the monthly premium, annual deductible, and satisfaction rating, the length of the bar represents the amount such that a smaller bar indicates a lesser dollar amount or lesser satisfaction rating and a longer bar indicates a higher dollar amount or high satisfaction rating.

Performance with pie charts is best when the size of each slice or piece represents a proportionate value (e.g., percentage) of the whole pie (Gillan, et. al, 1998). For example, one slice may indicate 25% of the whole – and thus would take up one-fourth of the total pie area. For the gap coverage attribute, there were five distinct categories that can be considered absolute, rather than proportional. However, because each category represents one level of gap coverage out of five possible levels, a pie chart may be ideal because each slice represents more or less gap coverage – making it proportional to the maximum (all generics) and minimum (no gap coverage) category. Decision makers simply need to understand that each level of gap coverage (or slice) will always be one-fifth the size of the pie and it is not necessary to know an exact difference (because this information is unknown) to make this decision (e.g., how much coverage in dollars). Thus, a pie chart was chosen to display this attribute. As in the color info-vis, the additional perceptual information is expected to reduce cognitive comparisons (WM demand) and instead rely on perceptual comparisons.

Which plan has the highest satisfaction rating?

Name	Gap coverage	Monthly premium	Annual deductible	Satisfaction rating
Plan A	●	$323	$287	★
Plan B	◔	$362	$295	★★★★★
Plan C	○	$321	$309	★★★
Plan D	●	$390	$218	★★
Plan E	●	$377	$300	★★
Plan F	○	$224	$333	★★
Plan G	●	$219	$221	★★★
Plan H	◔	$238	$343	★★★★
Plan I	◔	$385	$226	★★
Plan J	◑	$288	$353	★★
Plan K	●	$285	$328	★★★★
Plan L	◑	$241	$238	★★
Plan M	◕	$212	$299	★★
Plan N	◑	$328	$264	★
Plan O	◔	$360	$292	★★★

Figure 5. Size information visualization (size info-vis).

Task difficulty

Task difficulty was directly manipulated by varying the number of plan attributes that must be considered in order to accurately complete the task. In the low difficulty condition, participants selected a plan based on one attribute (e.g., which plan has the lowest monthly premium?). The high difficulty condition required the participant select a plan by integrating and comparing three attributes of each plan (e.g., which plan has the lowest monthly premium, highest gap coverage, and highest satisfaction rating?). For both conditions, the data was structured so that only one plan best met all of the criteria in the question. This manipulation required participants to make a compensatory decision and use an analytical decision strategy in order to select the best answer.

Secondary Task Workload Inducement

WM demand was induced by adding a secondary concurrent task to the primary decision making task. The n-back task requires participants to remember a series of letters and later recall the letters in reverse order and identify a letter some number (n) back from the end of the sequence. A high working memory demand was induced to encourage participants to opt for heuristic-based decision making (less optimal decision making) and to test the efficacy of the aid in reducing working memory demand (evidenced by more optimal decision making).

Materials

Equipment

Participants used PC-compatible computers and donned headphones during the experiment. The experiment was programmed using E-prime (version 1.1).

Surveys & Abilities

Demographic information, health information, insurance experience, technology experience, and an exit survey were collected from each subject. A blocked design allowed us to administer the NASA-TLX at the end of each block for each level of task difficulty and WM demand.

Tasks

Decision task

All participants were assigned to one of the three decision aid conditions and performed tasks at both levels of difficulty and WM demand. A standardized format was used so that the question, plan data, and choice set always appeared in the same location for each trial. The question was located at the top of the screen, with the decision aid below it. Decision performance was assessed by measures of accuracy, decision quality, and task time.

Working memory task

The purpose of this task was to place an additional memory burden on participants in order to examine performance with a decision aid when the task demands constrained the user's working memory capacity. An auditory n-back WM lag task was used for this purpose.

Pilot testing revealed that participants were unable to perform the task when the list length was greater than 6 letters. In addition, participants noted that they realized the first letter presented does not need to be remembered because it was never part of the recall portion (e.g., 8 letter series, and only asked for up to 4 letters back). The letter set varied from 4 to 6 letters to prevent participants from anticipating the recall task. The recall task asked for the letter that was between 1-back through 6-back from the end of the sequence, so that the user must remember all letters presented to successfully complete the task. For example, the participant heard a letter sequence A-B-C-D, and was told to remember all letters in sequence. After the decision task, they would be asked to recall a letter some number (n) back from the end of the sequence (in this example 1-back is D, 3-back is B). Dependent measures of accuracy and task time were used to assess WM task performance. WM task accuracy scores under two standard deviations from the mean were established as exclusion criteria because a score that low indicates that these participants were not performing both tasks.

Procedure

Experimental sessions were administered in groups of 1 to 4 participants; however each participant worked independently. After signing consent forms, the experimenter administered a paper and pencil working memory ability test, the Reverse Digit Span (Wechsler, 1997), before moving on to the computerized portion of the task. Participants were instructed to listen to the experimenter for instructions and to follow along on the computer screen.

The terms used in the decision task were defined by the experimenter and also presented visually on the screen. Next, the experimenter guided participants through a series of practice sessions. The first session introduced the low difficulty decision making task. Once participants were oriented to the screen, the experimenter walked participants through an example question step by step. Participants chose an answer by pressing the letter on the keyboard associated with the selected plan (e.g., participants pressed the “A” key to select Plan A). The practice questions did not have a time limit to ensure time for questions, but did include feedback to make sure participants understood the task. Next, participants completed another example on their own. The experimenter then followed the same procedure for a high difficulty task, such that there was a worked example and then an individual practice example.

The second practice session introduced the n-back task. Again, participants were oriented to the display, and the experimenter walked participants through a worked example. Participants used headphones to listen to a pre-recorded series of letters (at a rate of 1 letter every 3 seconds) and were asked to remember those letters in serial order. The recall portion of memory task was displayed visually for 30 seconds on the computer screen, instructing the participant to key in the letter that was n-back from the last letter. For example, if the subject heard the letter series: ABCDEFG, and the recall portion asks for the letter that was 4-back from the end the correct answer would be D. The recall portion in this practice session was not timed. Instead, the program waited for the user's response before moving on to the next screen. Next, participants completed a timed example of the n-back task.

The third and final practice task was included to help participants understand the dual-task paradigm. The experimenter explained a complete example trial, which included first the auditory presentation of the letter set, then the decision task, and lastly the recall question for the n-back. The last example had the same time limit as the actual experiment; 3 minutes for the decision task and 30 seconds for the recall portion of the n-back task. At the end of this practice task, a screen prompted users to fill out the NASA TLX survey. The experimenter explained part 1 and part 2 of the paper and pencil survey, and participants then filled out both parts on a practice survey.

An overview of the experimental procedure was given to participants both verbally by the experimenter and visually on the screen. During pilot testing, participants expressed a tendency to ignore the n-back task because of its relative difficulty to the decision task, which was reflected in their low n-back performance scores. The instructions were changed by telling participants that their most important task was the n-back memory task, rather than treating both tasks equally. In addition, feedback on the n-back task was given during both the practice and during the study. Pilot data reflected an increase in performance scores on the n-back task and so the instructions and feedback were added to the actual study.

At the end of the practice sessions, participants were instructed to move on to the actual study. Instructions were provided on the screen before and after each block of questions. Each participant completed a total of 40 trials (8 blocks of 5 questions each). The computer notified participants when they had completed all trials and then participants completed a computerized exit survey, demographics and health survey, technology experience survey, an insurance purchasing experience questionnaire, and an exit survey.

RESULTS

Data from 5 subjects were removed from the analysis. One was removed because of technical difficulties (the program was not responding) that prevented that subject from completing the experiment. Two subjects were removed because they did not follow directions; one wrote down the letter series during the n-back task and the other did not fill out any of the NASA-TLX surveys after the practice block. Two subjects were removed because their n-back scores were lower than 2 standard deviations from the mean, indicating that they were not performing both tasks in the dual-task portions of the experiment. The remaining 29 subjects, ages 18-26 (M=18.62, SD=1.63) were used in the analysis of all dependent variables. Remaining participant characteristics can be found in Table 1. Chi-square analysis revealed no significant ($p > 0.05$) differences between decision aid groups in WM ability (assessed using the Reverse Digit Span test), computer experience, health, education, or insurance purchasing experience.

Table 1.
Experiment 1: Participant Characteristics (N=29)

Category		Frequency	Percentage
Gender			
	Female	18	62%
	Male	11	38%
Race/Ethnicity			
	Black/African American	5	17%
	White	23	80%
	Multiracial	1	3%
Health			
	Fair	1	3%
	Good	5	17%
	Very Good	15	52%
	Excellent	8	28%
Marital status			
	Single	28	97%
	No answer	1	3%
Highest Education			
	High School diploma	24	83%
	Some college	5	17%
Experience with computers?			
	Yes	29	100%
Computer experience (years)			
	1 year but less than 3 years	1	3%
	At least 5 years	28	97%
Insurance types of which participant are named on the policy			
	Health insurance	27	93%

Table 1. (continued)
Experiment 1: Participant Characteristics (N=29)

Category		Frequency	Percentage
	Prescription drug insurance	14	48%
	Health savings account	5	17%
	Medicare plans	5	17%
	Dental insurance	22	76%
	Vision insurance	10	35%
	Motor vehicle insurance	22	75%
	Homeowner's insurance	3	10%
	Renter's insurance	2	7%
	Life insurance	11	38%
Insurance purchased types			
	Health insurance	3	10%
	Prescription drug insurance	1	3%
	Health savings account	2	7%
	Medicare plans	3	10%
	Dental insurance	3	10%
	Vision insurance	1	3%
	Motor vehicle insurance	2	7%
	Homeowner's insurance	2	7%
	Renter's insurance	2	7%
	Life insurance	3	10%
	None of the above	23	80%
Number of times purchased insurance			
	Never	24	83%
	1 time but less than 5 times	2	14%
	6 times but less than 10 times	3	7%

The remainder of the analyses is grouped by task; the decision making task and the WM task. The decision task had several dependent variables including accuracy (number correct), quality (how well the answer matched the criterion in the question), and mean decision time (in seconds). The WM task had 2 dependent variables; accuracy (number correct) and mean reaction time (in seconds). All analyses were conducted at an alpha level of .05 and all post-hoc tests and pairwise comparisons used the Bonferonni degrees of freedom adjustment. The main effects are reported but not explained if an interaction was present for that variable.

Decision Task

Decision Accuracy

A decision accuracy score was calculated for each level of difficulty (low and high) and level of WM demand (no WM task and WM task). A score of 1 (correct) was when a participant chose the best answer (the answer that met all the criteria in the question). All other choices were scored with a zero (incorrect). The number correct was summed to create a total accuracy score for each of the four conditions (low and high with the WM task, and low and high without the WM task). There were 10 questions for each condition, therefore the maximum score was 10 points and the minimum score was zero points.

A 3 (decision aid condition: table, color info-vis, size info-vis) x 2 (task difficulty: low, high) x 2 (WM demand task: with, without) mixed measures ANOVA was conducted to analyze decision accuracy with decision aid as the between subjects variable, task difficulty and WM demand as the within subjects variables. The results are graphed in Figure 6.

For decision accuracy, there were significant main effects of decision aid (F (2, 26) = 8.42, p=.002, η_p^2 =.39), task difficulty (F (1, 26) = 70.81, p < .000, η_p^2 =.73) and WM demand (F (1, 26) =5.674, p=.025, η^2 =.18). The type of decision aid significantly interacted with the WM demand task on decision accuracy (F (2, 26) = 6.956, p=.004, η_p^2

=.34). Post-hoc analysis revealed the source of the interaction to be in the table condition; participants' accuracy scores diminished significantly with the addition of the WM task (*M*=6.61, *SD*=1.49) as compared to without the WM task (*M*=8.06, *SD*=1.01). For the color and size visualization conditions, there were no significant differences in accuracy with the addition of the WM task.

There was also a significant interaction between task difficulty and WM demand on decision accuracy (F (1, 26) =4.449, p=.045, η_p^2 =.15). Only when task difficulty is high does performance accuracy significantly decrease with the addition of the WM task (*M*=6.41, *SD*=2.46) compared to performance without the WM task (*M*=7.20, *SD*=1.83). The three-way interaction between decision aid, task difficulty, and WM demand was not significant, however the observed power was low (.228) and may have decreased the ability to detect an effect (an increase in probability of committing a Type-2 error).

In sum, increasing WM demand only had significantly detrimental effects on decision making accuracy in the table condition. This finding suggests that WM demand is an important factor when the decision maker uses a table but not when using information visualization. In line with the predicted effects, information visualizations were able to mitigate WM demand enough to prevent accuracy decrements. As expected, accuracy was lower in the high difficulty questions than in the low difficulty questions. The tasks were designed such that the low difficulty tasks required fewer comparisons than the high difficulty tasks, and thus the low difficulty tasks were expected to have a lower WM demand than the high difficulty tasks.

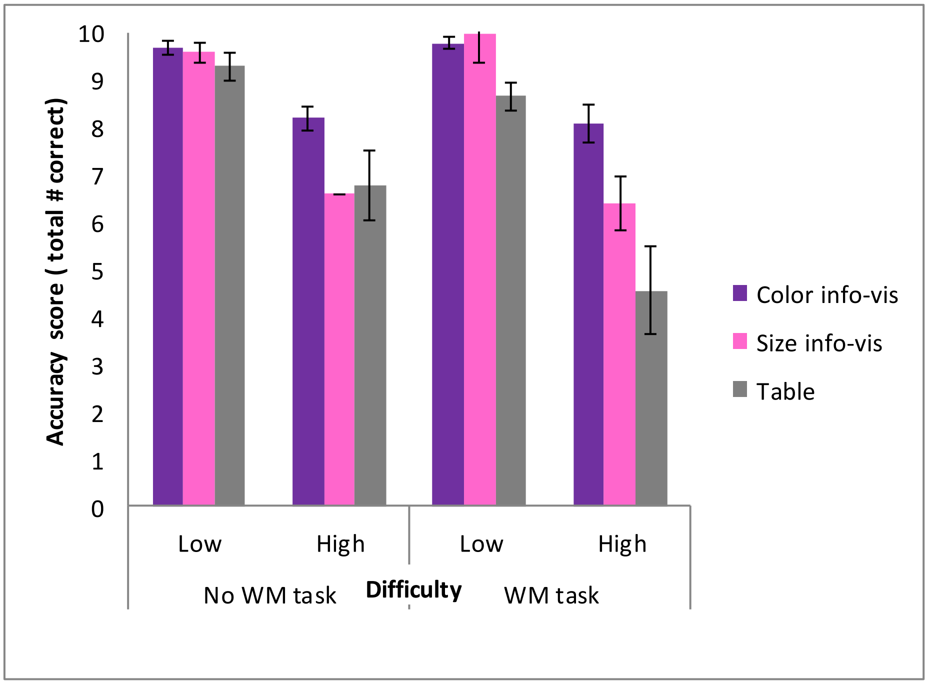


Figure 6. Decision task accuracy by decision aid for both low and high difficulty tasks and with and without WM task. Error bars represent standard error of the mean.

Decision Quality

For each high difficulty question, the plan data was created so that only one option met all of the criteria presented in the question during each trial. The other plan options met 0, 1, 2 or of the 3 possible criterion. Choosing the correct plan assumes that each criterion was used in the assessment. Thus, a maximum score of 3 is possible for each question and represents the best answer. A minimum score of 0 indicates that the plan chosen met none of the criteria in the question. These points were added together to compute a total decision quality score for both the high difficulty with WM task and high difficulty without the WM task. For the computed score, the maximum score was 30 points (3 x 10 questions) and the minimum score was 0 points.

A 3 (decision aid condition) x 2 (WM demand) mixed measures ANOVA revealed a significant main effect of decision aid (F (2, 28) =3.47, p=.045, η_p^2 =.20) and a significant interaction between decision aid and WM demand (F (2, 28) = 4.10, p=.027, η_p^2 =.28, see Figure 7) on decision quality. Only in the table condition did the addition of the WM task significantly diminish decision quality (without the WM task, M=25.11, SD=3.62; with the WM task, M=20.78, SD=5.95). Decision quality was adversely affected when WM demand was increased in the table condition only. Similar to the findings for decision accuracy, decision quality remained stable in the info-vis conditions as WM demand increased.

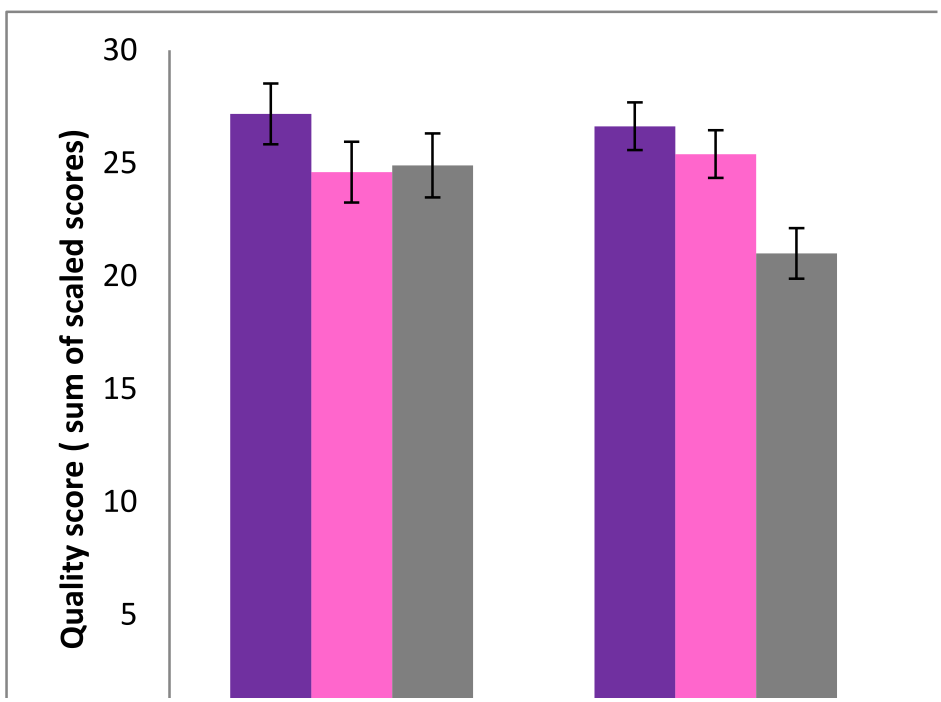


Figure 7. Decision quality by decision aid for high difficulty tasks and WM demand. Error bars represent standard error of the mean.

Mean Decision Task Time

Task time was recorded (in ms) and began when the decision task appeared on the screen and ended when the participant selected an answer. Milliseconds were converted to seconds before analysis for simplicity. A 3 (decision aid condition: table, visualization A, visualization B) x 2 (task difficulty: low, high) x 2 (WM demand task: with, without) mixed measures ANOVA was conducted to analyze decision reaction time in seconds, with decision aid as the between subjects variable, task difficulty and WM demand as the within subjects variables. The results are graphed in Figure 8.

There was a significant main effect of task difficulty on decision task time (F (1, 26) = 177.49, p<.000, η_p^2 =.87). There were no significant main effects of decision aid and WM demand on reaction time, however there was a significant three way interaction between task difficulty, WM demand, and decision aid (F (2,26)=4.00, p=.031, η_p^2 =.24). The addition of the WM task led to an increased decision task time in the size info-vis for high difficulty tasks only (without WM task M=31.15, SD=11.47; with WM task M=40.66, SD=23.89) and for the table condition for low difficulty tasks only (without WM task, M=7.97, SD=1.32; with WM task, M=10.65, SD=3.03). The addition of the WM task did not negatively affect decision making time significantly in the color info-vis condition for either level of task difficulty.

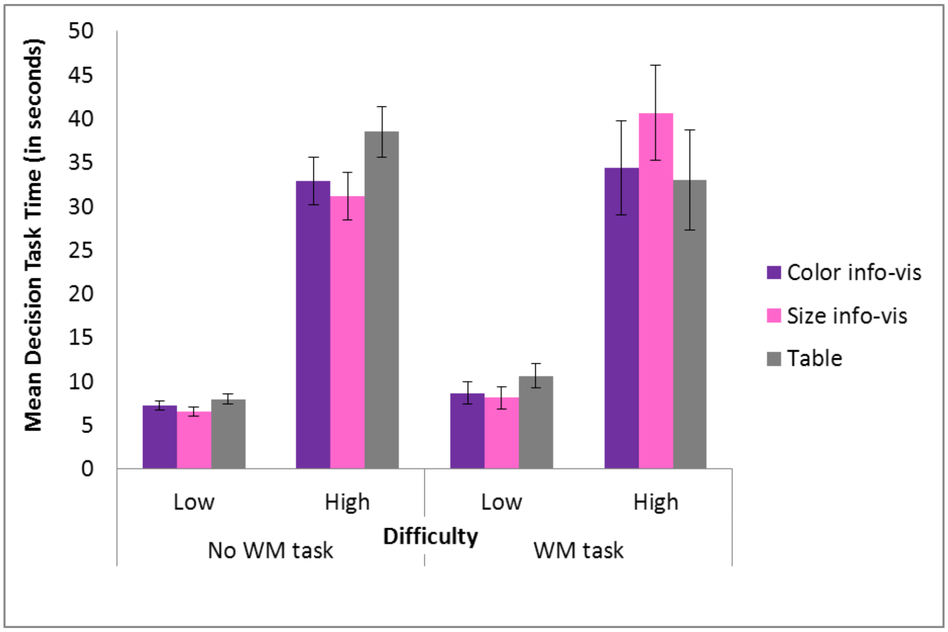


Figure 8. Mean Decision Task Reaction Time (in seconds) by decision aid for both low and high difficulty tasks and with and without WM task. Error bars represent standard error of the mean.

Working Memory Task

WM Task Accuracy and Mean Reaction Time

Performance accuracy on the WM task (n-back) was calculated by summing the number of correct answers (1 point for each correct answer, 0 for each incorrect answer). Only 4 blocks contained this task: 2 blocks of low difficulty questions without the WM task and 2 blocks of high difficulty questions with the WM task. The maximum score for each type of question was equal to the number of trials, so 10 was the maximum score and 0 was the minimum score. A cut-off of 3 points was used to eliminate subjects from the analysis because a score this low (over 2 standard deviations from the mean) indicated that the subject was not completing both tasks, and thus performance scores were not comparable to cases with the additional WM task.

In addition to the outlier analysis, a 3 (decision aid: table, color info-vis, size info-vis) x 2 (task difficulty: low, high) mixed measures ANOVA was conducted for both n-back accuracy and n-back reaction time to be sure that performance on this task did not confound decision task performance. No main effects or interactions of decision aid or task difficulty were found for either dependent variable (p>.05).

Subjective Workload – NASA TLX survey

A subjective workload survey, the NASA TLX, was administered as a manipulation check of task difficulty and WM demand. Only overall scores were analyzed and are graphed in Figure 9. A main effect of task difficulty was significant (F (1, 26) = 55.08, p<.000, η_p^2 =.68) and WM demand was significant (F (1, 26) = 49.03, p<.000, η_p^2 =.65). These main effects were qualified by a significant interaction between task difficulty and WM demand (F (1, 26) =33.94, p<.000, η_p^2 =.57). Participants rated low difficulty tasks without the additional WM task as having a lower subjective workload (*M*=38.22, *SD*=13.27) than with the WM task (*M*=60.91, *SD*=17.35) and high difficulty tasks without the additional WM task lower (*M*=57.34, *SD*=14.83) than with the WM task (*M*=65, *SD*=16.03). There were no significant main effects or interactions for the decision aid variable. Power was low (.219) for the three way interaction (decision aid, task difficulty, and WM demand), the two way interactions (decision aid by task difficulty (.365); decision aid by WM demand (.413)), and the main effect of decision aid (.455) which may help explain the lack of significance.

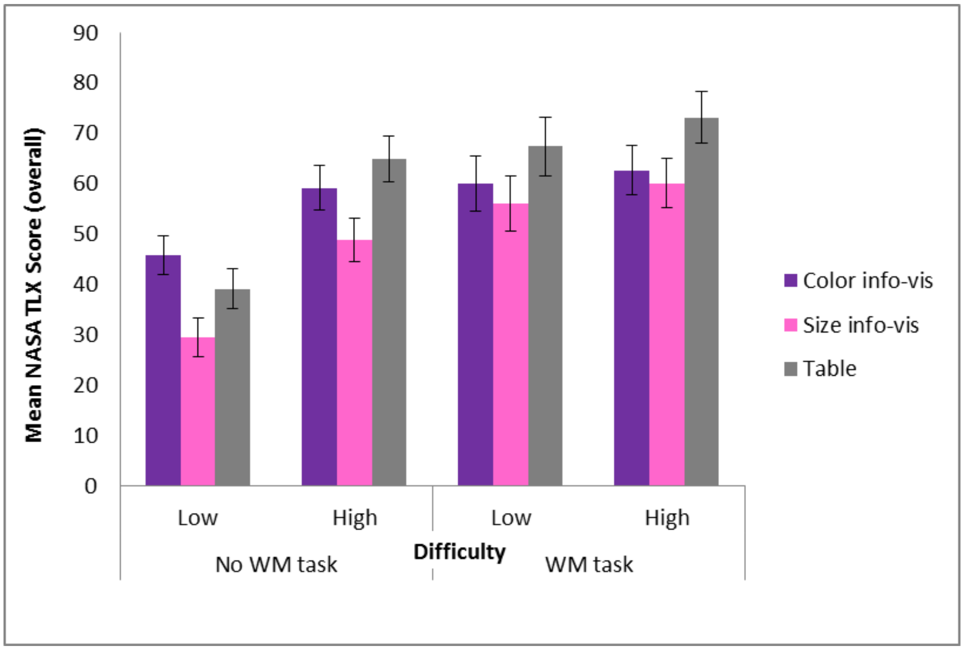


Figure 9. NASA TLX subjective workload scores by decision aid for both low and high difficulty tasks and with and without WM task. Error bars represent standard error of the mean.

Exit Survey

Participants were asked to rate a series of questions including the clarity of the instructions, difficulty of specific tasks, and dividing attention on a 1-5 point Likert scale. The results are listed in Table 2 below. Overall, participants indicated that they understood the directions but had a hard time dividing their attention between the decision task and the WM task. Consistent with the TLX results, participants rated the low difficulty decision task as less difficult than the high difficulty decision task.

Table 2.
Experiment 1: Exit Survey Results

Question	Mean	SD	Category
How clear were the directions in telling you what you were supposed to do?	4.28	0.96	Moderately to Extremely Clear
How difficult did you find the n-back memory task?	3.69	1.14	Somewhat to Moderately Difficult
How difficult did you find the decision task only 1 criterion?	1.21	0.41	Not at all to Slightly Difficult
How difficult did you find the decision task with more than 1 criterion?	3.21	1.01	Somewhat to Moderately Difficult
How difficult was it for you to divide your attention in the decision task and memory task part of the study?	3.97	0.19	Somewhat to Moderately Difficult

DISCUSSION

The goal of Experiment 1 was to design an info-vis that would reduce the WM demand of the decision making task. Identifying WM demanding comparison subtasks (via a task analysis) provided the opportunity to employ an environmental support to shift the WM demand to the more automatic, visual perception system using color and size manipulations. Making comparisons using visual cues (size or color) was predicted to reduce the WM demand of the task. Reducing the WM task demands would allow the decision maker to use an analytical strategy and compare more options, rather than rely on heuristic strategies to reduce the WM demand.

Following this logic, main effects of decision aid were expected for accuracy and quality. These hypotheses were directional in that the info-vis conditions were expected to improve both accuracy and quality compared to the table condition. Furthermore, when WM demand was increased by adding the n-back task, it was predicted that the info-vis conditions would not see a performance decrement but that the table condition would; indicating that WM is the mechanism responsible for performance differences. Decision task time was also predicted to be faster in the info-vis conditions compared to the table.

A main effect of task difficulty was expected such that performance across all decision aids and variables would be higher in the low difficulty conditions than in the high difficulty conditions. A main effect of decision aid was also predicted for subjective workload, such that the participants would rate the table condition as having a higher workload than the info-vis conditions. In addition, low difficulty tasks were expected to be rated as having a lower subjective workload than high difficulty tasks.

Although the size info-vis followed the same trends as the color info-vis, only the color info-vis was statistically better than the table condition on both accuracy and quality. Previous research indicates that size comparisons are more difficult and more cognitively demanding than color comparisons because size comparisons may place higher demands on visuo-spatial WM (Tricket & Trafton, 2007). Therefore, this may be one plausible explanation for why the size info-vis was unable to reduce WM demand as significantly as the color info-vis. The same logic may also explain why subjects spent more time making a decision on high difficulty tasks with the size info-vis when WM demand was high, than in low difficulty tasks without the WM demand was lower.

Unexpectedly, participants did not perform the decision task significantly faster with the info-vis conditions than with the table. One plausible explanation is that because participants were able to make more perceptual comparisons with the info-vis conditions, they were instead able to make more comparisons between plans in the same amount of time as those in the table condition. Similar to the accuracy and quality results, the color info-vis was the only condition that showed no task time decrements with the addition of the WM task.

Subjective workload was assessed using the TLX and the exit survey questions as a manipulation check for task difficulty. The predicted direction of the main effects of task difficulty and WM demand were confirmed. Low difficulty tasks were rated as having a lower workload than high difficulty tasks; and tasks with the WM demand task were rated as having a higher workload than without the WM task. It was expected however, that participants would rate the info-vis conditions as having a lower workload if they reduced the WM demand of the task and improved performance. Only two scores (samples) were assessed for each block of the within subjects variables (task difficulty and WM demand) which may have reduced the power needed to detect a significant effect of decision aid (the between subjects variable).

Although participants did not make faster decisions with either info-vis, for the task of choosing a prescription drug plan accuracy and quality are arguably the most important dependent measures. In addition, decision time wasn't any slower than the table condition. For the high difficulty tasks, only the color-info-vis was resistant to performance decrements with the additional WM task on all dependent measures. There were no performance differences between the two info-vis conditions, however accuracy and quality were significantly better than the table condition under high difficulty, high working memory tasks with the color info-vis and not with the size info-vis. Based on the results of Experiment 1, Experiment 2 compared decision performance of older adults in the color info-vis condition to their performance in the table condition.

EXPERIMENT 2:

VALIDATING THE INFORMATION VISUALIZATION IN AN OLDER ADULT SAMPLE

The goal of Experiment 2 was to test the color information visualization from Experiment 1 as a viable decision support system for older adults who, because of reduced working memory capacity, may be more susceptible to less-optimal decision making. Older adults performed the same decision task from Experiment 1 using either a table or the color info-vis from Experiment 1. A concurrent task paradigm was not used because the decision task alone should have constrained older adults' reduced working memory capacity.

Based on the results of Experiment 1, older adults were hypothesized to make more accurate and better quality decisions using the color info-vis than the table for both levels of task difficulty (low, high). It was also hypothesized that older adults would be more accurate and faster in the low difficulty tasks compared to the high difficulty tasks. Decision task time in the color info-vis condition was hypothesized to be faster or not significantly different than the table condition (based on results in Experiment 1).

METHODS

Participants

Twenty-three older participants ages 65-80 were recruited through an existing database of volunteers in the surrounding community. Older adults received $14 in compensation for participating. Similar to Experiment 1, color-blindness and the inability to read a computer screen were the only exclusion criteria.

Materials

Decision aids

The same table condition and color info-vis from Experiment 1 was used in Experiment 2. All other surveys and tasks were identical to Experiment 1.

Design & Procedure

Experiment 2 is a 2 (decision aid: table, color info-vis) x 2 (task difficulty: low, high) mixed design, with decision aid as the between subjects variable and task difficulty as the within subjects variable. Each participant was randomly assigned to one of the decision aid conditions, and completed trials at both levels of task difficulty. The procedure for Experiment 2 was identical to Experiment 1, excluding the secondary WM task (n-back task). Excluding the WM task reduced the number of trials by half (20 trials instead of 40).

RESULTS

Participants

Twenty-three older adults (12 female) between the ages of 66 and 80 (M=72.4, SD=3.73) participated in this study. No significant differences ($p > .05$) were found between decision aid groups on computer experience, health, insurance purchasing experience, working memory, or age. More detailed participant characteristics can be found in Table 3. All subjects were included in the following analyses.

Table 3.
Experiment 2: Participant Characteristics (N=23)

Category		Frequency	Percentage
Gender			
	Female	12	52%
	Male	11	48%
Race/Ethnicity			
	White	22	96%
	Other	1	4%
Health			
	Fair	5	22%
	Good	5	22%
	Very Good	7	30%
	Excellent	6	26%
Marital status			
	Single	1	4%
	Married	21	92%
	Widowed	1	4%
Highest Education			
	High School diploma	7	30%
	Vocational training	2	9%
	Some college/Associate's degree	6	26%
	College graduate	5	22%
	Master's degree (or other post-graduate training)	3	13%
Experience with computers?			
	Yes	23	100%
Computer experience (years)			
	Less than 6 months	1	4%
	6 months but less than 1 year	1	4%
	1 year but less than 3 years	0	0%
	3 years but less than 5 years	3	13%
	At least 5 years	18	79%
Insurance types of which participant are named on the policy			
	Health insurance	23	100%

Table 3. (continued)
Experiment 2: Participant Characteristics (N=23)

Category	Frequency	Percentage
Prescription drug insurance	18	78%
Health savings account	2	9%
Medicare plans	22	96%
Dental insurance	6	26%
Vision insurance	4	17%
Motor vehicle insurance	23	100%
Homeowner's insurance	23	100%
Renter's insurance	0	0%
Life insurance	20	87%
Insurance purchased types		
Health insurance	20	87%
Prescription drug insurance	14	61%
Health savings account	2	8%
Medicare plans	19	83%
Dental insurance	7	30%
Vision insurance	5	22%
Motor vehicle insurance	23	100%
Homeowner's insurance	23	100%
Renter's insurance	2	8%
Life insurance	15	65%
Number of times purchased insurance		
Never	1	4%
1 time but less than 5 times	2	8%
6 times but less than 10 times	10	44%
At least 10 times	10	44%

Decision Accuracy

A 2 (decision aid) x 2 (difficulty) mixed measures ANOVA revealed a significant main effect of task difficulty on decision accuracy (F (1, 21) = 39.88, p<.000, η^2 =.65, see Figure 10). Participants performed the decision task more accurately in the low difficulty condition (*M*=8.87, *SD*=1.39) than in the high difficulty condition (*M*=6.30, *SD*=2.05). There was no significant main effect of decision aid (F (1, 21) = 3.81, p=.064,

η_p^2 =.15) nor an interaction between task difficulty and decision aid (F (1, 21) =.829, p=.373, η_p^2 =.04). However, because the hypothesis being tested was directional (the color info-vis would *improve* performance), a one-tailed significance test is appropriate. The result is a significant main effect of decision aid (F (1, 21) = 3.81, p=.032, η_p^2 =.15), and confirms the hypothesis that older adults would perform significantly better in the color info-vis condition (M=8.13, SD=1.21) than the table condition (M=7, SD=1.55). The interaction remained insignificant.

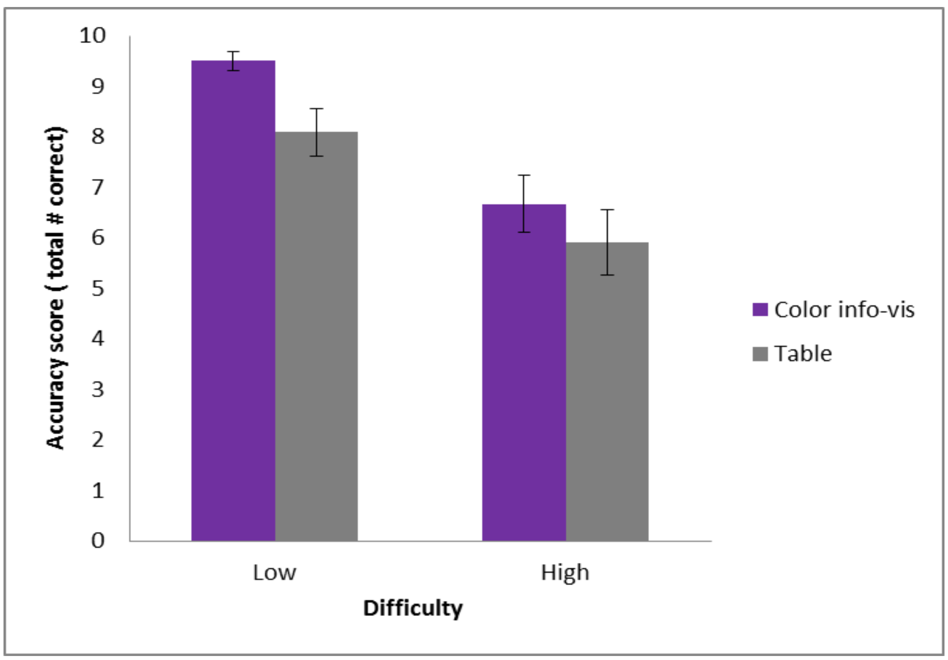


Figure 10. Decision task accuracy by decision aid for both low and high difficulty task. Error bars represent standard error of the mean.

Decision Accuracy by Attribute in the Low Difficulty Condition

For the low difficulty decision tasks, participants were asked to find a plan that best meets the single criterion (one attribute, i.e., satisfaction rating). Thus, we can analyze performance for each attribute (gap coverage, monthly premium, annual deductible, and satisfaction rating) individually to examine why participants were more accurate in the info-vis condition than in the table condition.

The data was analyzed using a 2 (decision aid) x 4 (plan attribute) mixed measures ANOVA. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi 2$ (5) = 36.65, p < .001), therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity (ε = 0.573) (Huynh & Feldt, 1976). Main effects of attribute type (F (1.72, 36.11) = 15.61, p<.000, η_p^2 =.43) and decision aid (F (1, 21) =7.1, p=.015, η_p^2 =.25), were qualified by a significant interaction between plan attribute and decision aid (F (1.72, 36.11) =8.81, p=.001, η_p^2 =.30). See Figure 11. Participants were better able to accurately answer questions about the gap coverage attribute in the color info-vis condition (M=91.7%, SD=20.77%) than in the table condition (M=51.73%, SD =27.51%). This difference is the source of the main effect of decision aid on accuracy.

Mean Decision Time by Attribute

2 (decision aid) x 4(plan attribute) mixed measures ANOVA on decision time (in seconds) was run to look for evidence of a speed-accuracy trade-off that might explain the effect of decision aid on accuracy with gap coverage questions. Mauchly's test was significant, indicating a violation of the sphericity assumption ($\chi 2$ (5) = 28.25, p < .001), thus degrees of freedom were corrected using the Huynh-Feldt estimates of sphericity (ε = 0.598) (Huynh & Feldt, 1976). The analysis revealed a significant main effect of decision aid (F (1, 21)=4.5, p=.046, η_p^2 =.18) and a significant main effect of plan attribute (F (6.8, 37.68) =6.82, p=.004, η_p^2 =.25), but not a significant interaction between decision aid and attribute (p=.079). Power was low for the interaction (.491), most likely

again due small sample sizes. Since this was not a planned analysis, the number of questions analyzed per attribute may not have been enough to detect an effect. Participants did spent more time answering the gap coverage questions than the other attributes and more time answering questions about this attribute in the table condition than in the color info-vis condition (see Figure12).

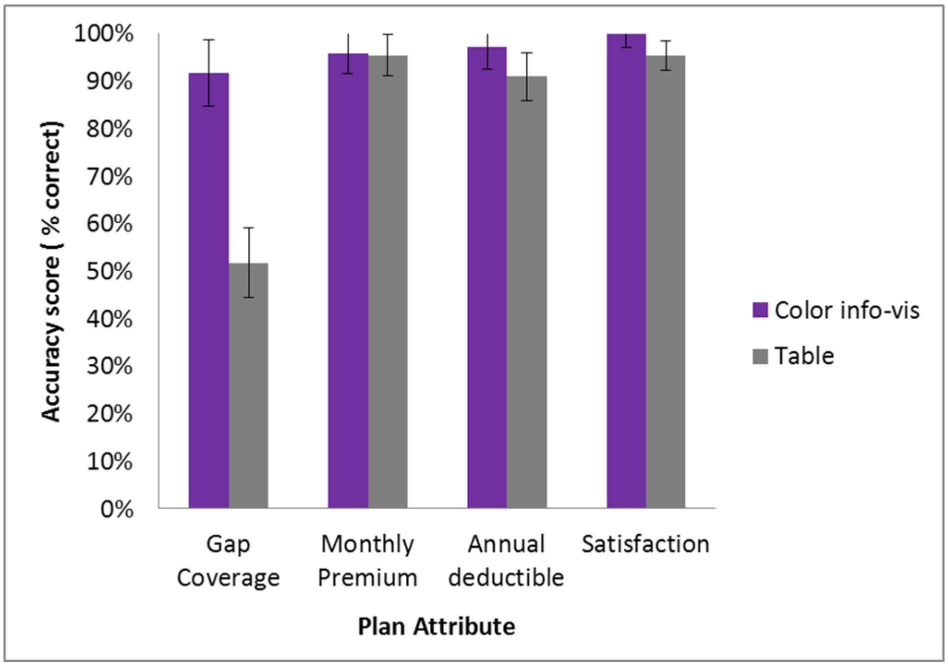


Figure 11. Percent accuracy on low difficulty tasks by plan attribute and decision aid. Error bars represent standard error of the mean.

Participants answered the decision ask significantly faster in the color info-vis condition (*M*=16.93, *SD*=5.95) than in the table condition (*M*=23.5, *SD*=8.35). Questions about the satisfaction rating attribute (*M*=13.69, *SD*=8.81) took significantly less time than the annual deductible (*M*=19.64, *SD*=5.22), gap coverage (*M*=25.41, *SD*=17.66), and monthly premium (*M*=19.56, *SD*=6.86). This indicates that there was not a speed-accuracy tradeoff that would explain significantly lower accuracy for gap coverage questions in the table condition versus the color info-vis condition.

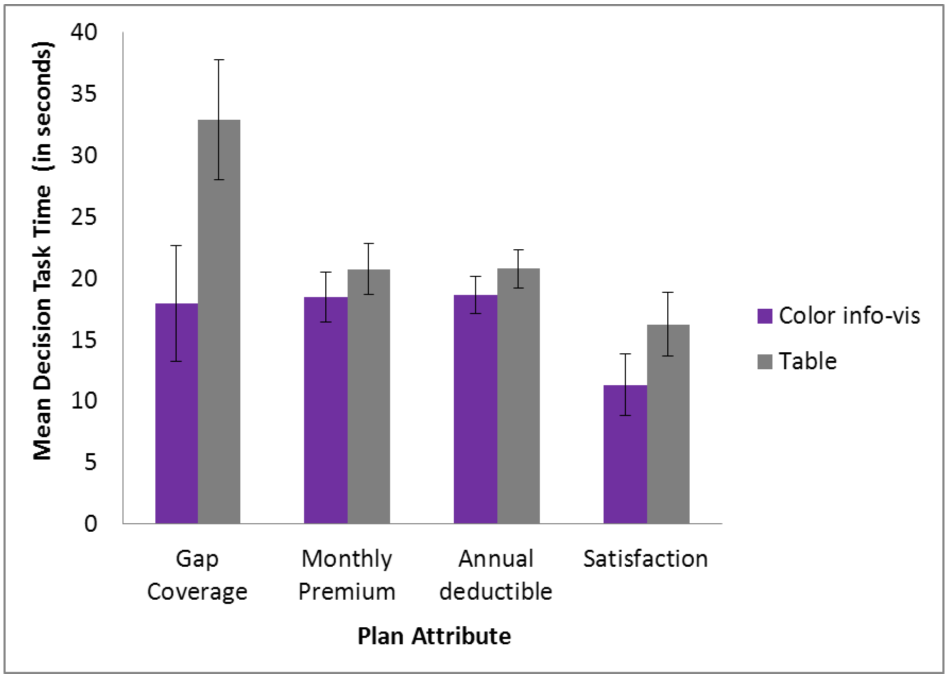


Figure 12. Mean decision time (in seconds) by plan attribute and decision aid for the low difficulty condition. Error bars represent standard error of the mean.

Decision Quality

An independent samples T-test was conducted between decision aid conditions on decision quality score and revealed that quality did not differ significantly by decision aid (t=.7, p=.492). A one-tailed significance test did not change the effect of the decision aid variable on decision quality.

Mean Decision Task Time

A 2(decision aid) x 2 (difficulty) mixed measures ANOVA was run to assess decision task time and revealed a significant main effect of difficulty (F (1, 21) = 155.73, p<.000, η_p^2 =.88), such that participants were much faster in the low difficulty condition

(M=20.07 sec, SD=7.78) than in the high difficulty condition (M=70.69, SD=20.92). See Figure 13. There was no significant main effect of decision aid (F (1, 21) = 1.07, p=.314, η_p^2 =.05) on task time, nor was there an interaction between decision aid and difficulty (F (1, 21) = .081, p=.779, η_p^2 =.01). This finding was consistent with Experiment 1.

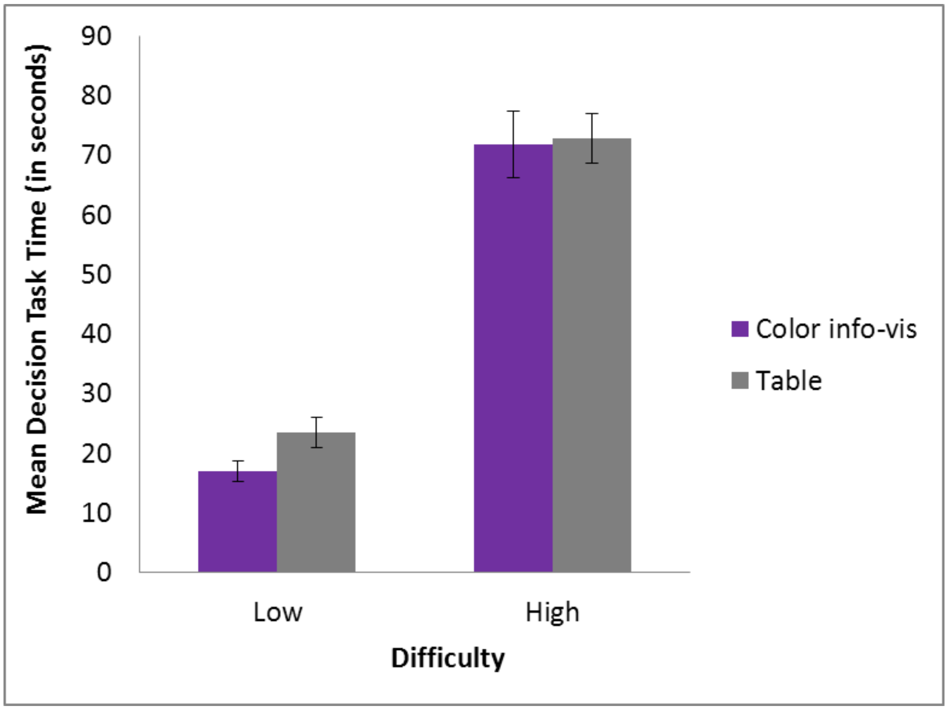


Figure 13. Decision task time by decision aid for low and high difficulty tasks. Error bars represent standard error of the mean.

Subjective Workload – NASA TLX Survey

Subjective workload ratings were assessed by conducting a 2 (decision aid) x 2(difficulty) mixed measures ANOVA. A significant main effect of difficulty (F (1, 21) =74.2, p <.000, η_p^2 =.78) was revealed and in the direction expected. See Figure 14. As in study 1, this was a manipulation check for difficulty and indicates a successful manipulation because participants rated the high difficulty tasks significantly more difficult (*M*=58.63, *SE*=3.57) than the low difficulty tasks (*M*=35.35, *SE*=2.99). Similar to Experiment 1, there was no main effect of decision aid (*F* (1, 21) = 1.5, *p*=.234, η_p^2=.07), nor an interaction effect of decision aid and difficulty (*F* (1, 21) =.06, *p*=.815, η_p^2=.003). Again, power was low for both the main effect of decision aid (.215) and its interaction with task difficulty (.056).

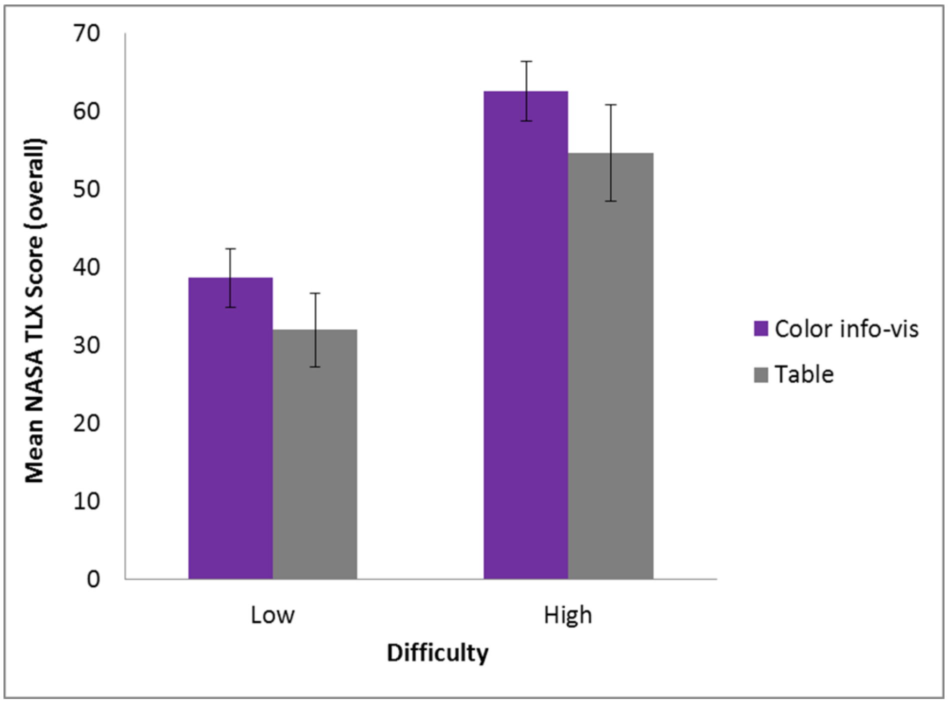


Figure 14. NASA TLX subjective workload scores by decision aid for both low and high difficulty tasks. Error bars represent standard error of the mean.

Exit Survey

Results from the exit survey questions are shown in Table 4 below. Responses to the exit survey confirmed the subjective workload results from the TLX survey; participants rated the high difficulty task as more difficult than the low difficulty task. There were no significant differences in ratings between decision aids.

Table 4**.**
Experiment 2: Exit Survey Results

Question	Mean	SD	Category
How clear were the directions in telling you what you were supposed to do?	4.09	0.95	Moderately to Extremely Clear
How difficult did you find the decision task only 1 criterion?	1.35	.71	Somewhat to Moderately Difficult
How difficult did you find the decision task with more than 1 criterion?	3.26	1.01	Somewhat to Moderately Difficult

DISCUSSION

Experiment 2 examined whether color information visualizations can be used as a decision support for older adults making complex decisions. Previous research has shown that older adults exhibit difficulty in choosing a prescription drug plan on the Medicare.gov website, possibly because of a combination of usability issues and normative changes in cognitive abilities such as reduced working memory capacity (Czaja, Sharit, & Nair, 2008).

It was hypothesized for Experiment 2 that older adults would perform better (higher accuracy and quality) in the color info-vis condition than in the table condition for both high and low difficulty tasks. Our results show that accuracy was significantly higher in the color info-vis condition (shifting processing burden from cognitive resources to perceptual resources) than in the table condition, indicating that older adults did not use heuristics but instead an analytical decision making strategy similar to younger adults in Experiment 1.

If older adults did not choose the best plan option, they were able to select a plan that was "good enough" in quality regardless of the decision aid. This finding is consistent with the current literature in that older adults' are likely to use heuristic strategies at a lower level of WM demand than younger adults and that they can be successful heuristic users (Chen & Sun, 2003).

Although the color info-vis may have been successful in reducing the WM demand for comparing plans on a single attribute (low difficulty task), the info-vis did little to support integration of more than one attribute (i.e., the three attributes required in the high difficulty tasks). The lack of an effect of condition on accuracy in the high difficulty tasks indicates that relying on perceptual capacities cannot fully accommodate age-related declines in cognitive capacities (color info-vis condition). Future research should evaluate ways to support, via information visualizations, more complex decision-making tasks where multiple attributes must be compared.

In the graph reading literature, a low difficulty condition is generally termed an extraction task because the user is asked to find a specific bit of information (e.g., what is plan B's monthly premium amount), rather than perform a comparison of one attribute among many options (e.g., which plan has the lowest monthly premium) as in this study. This may be why there was an effect in the low difficulty condition that isn't consistently found in other studies within the graph reading literature (e.g., Ratwani, Trafton, & Boehm-Davis, 2008).

In the low difficulty condition, older adults were much more successful choosing the correct answer when the question was about the gap coverage attribute. This finding is interesting for a number of reasons. First, the user had to remember what each of the colors meant or refer to the legend, which on the surface appears to increase WM demand. However, in the table condition gap coverage had to be evaluated based on textual values (e.g., all generics vs. some generics). This requires reading and comprehension of the text, rather than a less WM demanding visual search for a target color (Treisman, 1982). Second, previous literature has suggested that numeracy (ability to process numerical information) and processing speed (or how fast one can process information and perform tasks without focused attention) is responsible for performance differences with a large data set (24 plan options; Tanius, et. al, 2009). Using color comparisons rather than numerical comparisons may be a good option for those who do not have high numeracy abilities, WM abilities, and those with slower processing speed. In addition, our study suggests that using colors rather than the recommendation to assign categorical values to numerical data to help those with low numeracy (Tanius, et. al, 2009) may be more effective at increasing decision accuracy.

Whether or not the use of color is in fact allowing the user to make faster, less demanding comparisons might be a question that can be answered using eye-tracking data. For example, recording fixation durations and plotting saccadic amplitude could help answer the question of whether color is facilitating a less cognitively demanding search (Velichovsky, 2005). Long fixation durations might indicate focal vision which is indicative of selective attention while short saccades indicate a scanning behavior akin to ambient vision or more automatic (pre-attentive) processing.

Future research should examine how perceptual manipulations (e.g., color and size) interact together and whether high difficulty comparisons and integration tasks can be simplified. This study examined color and size separately and did not examine the effects of size and color together, or how these manipulations can improve specific types of data (e.g., categorical vs. interval).

CONCLUSION

The results of this study indicate that information visualization may be successful as an environmental support for both older and younger decision makers for comparison tasks. Reducing the WM demand of the task through the use of an environmental support can improve decision accuracy in some cases. Further research is needed to examine how information can be visualized to help with more difficult cognitive integration tasks.

REFERENCES

Baddeley, A. (1986). Working memory. New York: Clarendon Press/Oxford University Press.

Breslow, L.A., Ratwani, R.M., &Trafton, J.G. (2009). Cognitive models of the influence of color scale on data visualization tasks. *Human Factors,* 51 (3), 321-338.

Centers for Medicare & Medicaid Services, U.S. Department of Helath and Human Services. (2010). Medicare and You: National Medicare Handbook. Baltimore, MD.

Chen, Y., & Sun, Y. (2003). Age differences in financial decision-making: Using simple heuristics. *Educational Gerontology, 29*(7), 627-635.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.). *Human memory and cognitive capabilities, mechanisms, and performances* (pp. 409–422). Amsterdam: Elsevier Science.

Czaja, S.J., Sharit, J., & Nair, S. N. (2008). Usability of the Medicare Health Web Site. *Journal of the American Medical Association*, 300 (7), 790-792.

Finucane, M.L., Mertz, C.K., Slovic, P., & Scholze Schmidt, E. (2005). Task complexity and older adults' decision-making competence. *Psychology and Aging*, 20 (1), 71-84.

Finucane, M.L., Slovic, P., Hibbard, J.H., Peters, E., Mertz, C.K., & MacGregor, D.G. (2002). Aging and decision-making competence: An analysis of comprehension and consistency skills in older versus younger adults considering health-plan options. *Journal of Behavioral Decision Making*, 15, 141-164.

Glovich, T., Griffin, D.W., & Kahneman, D. (2002). Heuristics and biases: the psychology of intuitive judgment. Cambridge, UK: Cambridge University Press.

Hsu, J., Fung, V., Price, M., Huang, J., Brand, R., Hui, R., et al. (2008). Medicare beneficiaries' knowledge of part D prescription drug program benefits and responses to drug costs. *JAMA: The Journal of the American Medical Association*, *299*(16), 1929-1936.

Huynh, H. & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. Journal of Educational Statistics, 1, 69-82.

Lohse, G. L. (1997). The role of working memory on graphical information processing. *Behaviour & Information Technology, 16*(6), 297-308.

Mata, R., Schooler, L.J., Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology and Aging*, 22(4), 796-810.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity to process information. *Psychological Review*, 63, 81-97.

Meyer, J., Shinar, D., & Leiser, D. (1997). Multiple factors that determine performance with tables and graphs. *Human Factors*, 39(2), 268-286.

Morrow, D.G., & Rogers, W.A. (2008). Environmental Support: An integrative framework. *Human Factors*, 50(4), 589-613.

Plude, D.J. & Doussard-Roosevelt, J.A. (1989). Aging, selective attention, and feature integration. *Psychology and Aging*, 4(1), 98-105.

Ratwani, R.M., Trafton, J.G., & Boehm-Davis, D.A. (2008). Thinking graphically: connecting vision and cognition during graph comprehension. *Journal of Experimental Psychology:Applied*, 14(1), 36-49.

Salthouse, T.A. (1990). Working memory as a processing resource in cognitive aging. *Developmental Review, 10*, 101–124.

Sharit, J., Czaja, S. J., Nair, S., & Lee, C. C. (2003). Effects of age, speech rate, and environmental support in using telephone voice menu systems. *Human Factors, 45*, 234–251.

Simon, H.A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99-118.

Tanius, B.E., Wood, S., Hanoch, Y., Rice, T. (2009). Aging and choice: Applications to Medicare Part D. *Judgment and Decision Making*, 4(1), 92-101.

Triesman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 194-214.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281-299.

Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.

Tversky, A., Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185, 1124-1131.

Velichkovsky, B.M., Joos, M., Helmert, J.R., & Pannasch, S. (2005). Two visual systems and their eye movements: evidence from static and dynamic scene perception. *CogSci 2005: Proceedings of the XXVII Conference of the Cognitive Science Society*. July 21-23 Stresa, Italy, pp. 2283-2288.

Wechsler, D. (1997). Wechsler Adult Intelligence Scale (3rd ed.). San Antonio, TX: Psychological Corp.

Wickens, C.D. & Carswell, C.M.(1995). The proximity compatibility principle: its psychological foundation and relevance to display design. *Human Factors*, 37 (3), 473-494.

Wickens, C.D., Lee, J.D., Liu, Y., & Gordon Becker, S.E. (2004). An introduction to human factors engineering (2nd ed.). New Jersey: Pearson-Prentice Hall.

APPENDIX A

Medicare Prescription Drug Plan Terms and Definitions

Term	Definition
Annual Deductible	The amount you must pay for your prescriptions or other medical care, before your Medicare drug plan or Medicare Health Plan begins to pay. These amounts can change every year.
Cost Sharing	The amount you pay for health care and/or prescriptions. This amount can include copayments, coinsurance, and/or deductibles.
Coverage Gap	Medicare drug plans may have a "coverage gap," which is sometimes called the "donut hole." A coverage gap means that after you and your plan have spent a certain amount of money for covered drugs (no more than $2830), you have to pay out-of-pocket all costs for your drugs while you are in the "gap:" The most you have to pay out-of-pocket in the coverage gap is $3610. This amount doesn't include your plan's monthly premium that you must continue to pay even while you are in the limit, you will have "catastrophic coverage." This means that you will only pay a coinsurance amount (like 5% of the drug cost) or a copayment (like $2.50 or $6.30 for each prescription) for the rest of the calendar year.
Monthly Premium	The periodic payment to Medicare, an insurance company, or a health care plan for health care or prescription drug coverage. In a few cases, a note will say "Under Review" instead of a premium amount. This means Medicare and the company are still discussing the amount.
Generic Drug	A prescription drug that has the same active-ingredient formula as a brand-name drug. Generic drugs usually cost less than brand-name drugs. The Food and Drug Administration (FDA) rates these drugs to be as safe and effective as brand-name drugs.
Out-of-Pocket Costs	Health care costs that you must pay on your own because they are not covered by Medicare or other insurance.
Tiers	Drugs on a formulary are often organized into different drug "tiers," or groups of different drug types. Your cost depends on which drug tier your drug is in. For example, a plan may form tiers this way: • Tier 1 - Generic drugs. • Tier 2 - Preferred brand-name drugs. • Tier 3- Non-preferred brand name drugs

APPENDIX B

Task Analysis for Choosing a Prescription Drug Plan from the Medicare.gov Website.

Task #	Task	Task/Knowledge Requirements*	Feedback	Potential Problems
1.0	**Observe the table**			
1.1	Read each of the headings	Reading comprehension[7]	None	Does not understand jargon
1.1.1	Click on "What is this?" if unsure what the heading means	Visual search[8]	Small window appears	Fails to see link
1.1.2	Read the definition	Reading comprehension[7]	None	Does not understand definition
1.1.3	Press the x button to close the window	Declarative knowledge[4]	Small window closes	Unable to figure out how to close window and return to previous window
2.0	**Find total yearly fixed cost**			
2.1	Locate plan with cheapest monthly premium	Visual search[8]	None	Forget which plan had the lowest amount
2.1.1	Extract the monthly premium amount	Visual search[8]	None	
2.2	Multiply premium by 12 months on paper	Numerical computation[6]; Working memory[9]	None	May miscalculate amount
2.3	Locate the annual deductible	Visual search[8]; Working memory[9]	None	Forget which plan to use
2.4	Add deductible to the premium total	Numerical computation[6]; Working memory[9]	None	Forget which value was the premium total
3.0	**Calculate out-of-pocket expenses**			

3.1	List current drug costs	Declarative knowledge[4]	None	May not remember all costs; may miscalculate amount
3.1.2	Add up the monthly cost of your drugs	Numerical computation[6]; Working memory[9]	None	May miscalculate cost
3.1.3	Multiply monthly drug cost by 12 months	Numerical computation[6]; Working memory[9]	None	May miscalculate cost; drug costs may differ over the year
3.2	Calculate expenses outside of the gap	Numerical computation[6]; Working memory[9]	None	
3.2.1	Find the donut hole (gap in coverage) amount	Abstract reasoning[1]; Reading comprehension[7]	None	May not understand jargon or meaning
3.2.2	Click on the "what is this?" link in the "coverage in the gap" column	Reading comprehension[7]	Small window appears	Fails to see link; may not understand jargon
3.2.3	Read explanation	Reading comprehension[7]	None	May not understand explanation
3.2.4	Extract the donut hole amount	Visual search[8]	None	May not identify correct #
3.2.5	Subtract total yearly drug cost from the donut hole amount	Numerical computation[6]; Working memory[9]	None	Forget which amount was the total yearly drug cost
3.2.6	Divide that number by monthly cost of drugs to determine months of no coverage	Numerical computation[6]; Working memory[9]	None	Forget the monthly cost of drugs; May not understand how to do this task
3.2.7	Round that number up	Abstract reasoning[1]	None	
3.2.8	Multiply number of months without	Numerical	None	Forget the monthly cost

	coverage by monthly cost of drugs (cost w/o using insurance)	computation[6]; Working memory[9]		or calculated number of months without coverage
3.2.9	Add this to number 3.1.3	Numerical computation[6]; Working memory[9]	None	May forget the monthly drug cost before the donut hole
3.3	Calculate drug sharing costs		None	May not understand jargon
3.3.1	Click on the plan name to get to the detail page	Good visual acuity[5]; Attentional control[2]; Abstract reasoning[1]	Web page displays plan details	Fails to see or understand link
3.3.2	Scroll down to find the drug cost sharing box	Visual search[8]	Page moves with scrolling action	Fails to scroll to see more details
3.3.3	Read tier cost information	Reading comprehension[7]	None	May not understand jargon
3.3.4	Contact the drug plan to find out what tier your drugs are in (recommended)	Declarative knowledge[4]	Speak to representative	Unable to do this task independently
3.3.5	Calculate cost of drugs by multiplying the tier cost by number of drugs	Numerical computation[6]; Working memory[9]	None	Forget which drugs are in what tier or how much each tier was worth; Forget which plan is being evaluated
3.3.6	Add up totals	Numerical computation[6]; Working memory[9]	None	May miscalculate totals; Forget to add in a total
3.3.7	Add this number from 3.2.9	Numerical computation[6]; Working memory[9]	None	Forget amount from 3.2.9
4.0	**Find out if the plan is nationally recognized**			

4.1	Click "go back to plans in your state" to get back to the table	Visual search[8]; Abstract reasoning	Web page displays plan table	Fails to see link
4.1.1	Find the column labeled "Plan Name and ID numbers"	Visual search[8]; Abstract reasoning[1]	None	Fail to recognize that this information is under this header
4.1.2	Read information under plan name	Reading comprehension[7]	None	May not understand jargon
5.0	**Consider customer satisfaction ratings**			
5.1	Find the column for summary ratings	Visual search[8]	None	
5.1.1	Extract summary rating	Visual search[8]	None	
5.1.2	Click on the plan name to get to the detail page	Abstract reasoning[1]; Good visual acuity[5]	Web page displays plan details	Fail to realize detailed information exists
5.1.3	Find the box that contains all of the rating categories	Visual search[8]	None	Fail to scroll down
5.1.4	Read the label and definition of each rating category	Reading comprehension[7]	None	May not understand jargon
5.1.5	Compare the number of stars for each category	Comparison[3], Working memory[9]	None	May not understand stars
6.0	**Compare yearly cost and monthly cost to current budget**	Comparison[3], Working memory[9]	None	May forget monthly or yearly totals
7.0	**Repeat steps 2.0-5.1.2 for another plan**		None	May forget comparable values from prior plan
8.0	**Compare plan to another plan**			May forget comparable values from prior plan
8.1	Compare yearly fixed cost	Comparison[3], Working memory	None	May forget comparable values from prior plan
8.2	Compare out-of-pocket expenses	Comparison, Working memory[9]	None	May forget comparable values from prior plan

8.3	Compare nationwide coverage	Comparison[3], Working memory[9]	None	May forget comparable values from prior plan
8.4	Compare satisfaction ratings	Comparison[3], Working memory[9]	None	May forget comparable values from prior plan
9.0	**Repeat steps 2.0-8.4 for another plan**		None	May forget comparable values from prior plan
10.0	**Compare against all plans**		None	May forget comparable values from prior plan

NOTE: *Definitions for task /knowledge requirements:
[1]*Abstract reasoning*: Process of perceiving issues and reaching conclusions through the use of symbols or generalizations rather than concrete, factual information
[2] *Attentional control:* Controlled processing on difficult tasks or tasks that use unfamiliar items
[3] *Comparison:* Examination of 2 or more items to establish similarities and differences
[4] *Declarative knowledge:* Knowledge about facts or things
[5] *Good visual acuity:* Clarity or sharpness of vision, the ability to see fine detail (e.g., reading test, recognizing symbols)
[6] *Numerical computation:* Ability to solve mathematical equations
[7] *Reading comprehension:* Ability to understand what is read
[8] *Visual search:* Ability to actively scan the visual environment for a particular object or feature (target) among other objects or features.
[9] *Working memory:* Brief, immediate memory for material that is currently being processed; a portion of working memory also coordinates ongoing mental activities

APPENDIX C

Display Design Principles (adapted from Wickens, 2004)

Principles	Explanation
Perceptual Principles	
1. Make displays legible or audible	Information should be clearly presented
2. Avoid absolute judgment limits	Operator should not be required to judge a the level of a variable on the basis of a single sensory variable
3. Top down processing	Variables should be presented in accordance with expectations
4. Redundancy gain	Under degraded conditions information should be presented more than once
5. Discriminability	Similar elements cause confusion, highlight dissimilar information
Mental Model Principles	
6. Principle of pictorial realism	If possible, the display should look like the variable it represents
8. Principle of the moving part	Dynamic information should be compatible with user's expectations
Principles Based on Attention	
8. Minimizing information access cost	Minimize the effort and time it takes to direct selective attention
9. Proximity compatibility principle	Information that needs to be mentally integrated should be close in proximity
10. Principle of multiple resources	Use multiple resources (visual system, auditory) to present large amounts of information concurrently.
Memory Principles	
11. Replace memory with visual information knowledge in the world	Display necessary information rather than requiring the user to retain information.
12. Principle of predictive aiding	Display predictive information visually to reduce memory load
13. Principle of consistency	Display designs should be consistent with previous or conceptually similar displays.