Search, Memory, and Choice: An Experiment

Adam Sanjurjo[1]
Universidad de Alicante
12 December 2012[2]

Abstract: Behavior in rich search environments has recently been found to deviate systematically and substantially from optimal search. While the cause of these deviations is still unknown, they have been shown to reduce memory load for searchers. Here, an experiment tests whether such deviations are incentive compatible, by measuring whether or not reduced search memory loads result in reduced rates of choice error. In each search task, subjects are made to search in one of two *orders*; a high, or a low memory load order. Low memory load searchers are found to choose the best alternative at a substantially higher rate than high memory load searchers, when holding the information searched constant. This result suggests that observed systematic deviations from "optimal search" may actually be optimal in a slightly richer model.

KEYWORDS: Search, memory, choice, complexity, information overload.

Introduction

Search theory is one of the pillars of economic theory for good reason: all decision-making is preceded by some type of information search process- be it internal (via recall) or external (from external stimuli). Optimal search policies have been fully characterized for a wide variety of search tasks in which a single attribute of an alternative, usually its price, determines its desirability, with varying assumptions about value distributions, search costs, number of searchable alternatives, and recall options (Kohn and Shavell 1974, Lippman and McCall 1976).

Although the analysis of single attribute search models yields substantial insight, many important applications have alternatives whose values are determined by multiple attributes: we consider much more than wage when choosing a job, and much more than price when purchasing a home. Multiple attribute search environments are of particular importance because behavior in them approximates human cognition in less structured settings, and they closely resemble the kind of information displays found on popular

---

[1] Fundamentos del Análisis Económico, Universidad de Alicante, Spain. 03690 Alicante, Spain (email: sanjurjo@ua.es). I thank the students from Alicante´s 2010 graduate experimental economics course, María José Aragón, John Conlisk, Moira Decima, Eric Johnson, Vadym Lepetyuk, Asier Mariscal, Silvia Martinez, Joshua Miller, Sonia Oreffice, Carlos Oyarzun, Hal Pashler, Giovanni Ponti, Climent Quintana-Domeque, Miguel Sanchez Villalba, Carolina Silva, Luis Ubeda and especially Vincent Crawford for helpful comments.

[2] Financial support from the Spanish Ministerio de Ciencia y Tecnología and Feder Funds under project SEJ-2007-62656 is gratefully akclowledged.

internet websites, that compare anything from digital cameras to cars and homes, to health care plans and 401k´s.[3]

In contrast with the large body of theoretical work on optimal single attribute search, there has been little work on characterizing optimal policies in multiple attribute search (Bearden & Connolly 2007, Sanjurjo 2012a), presumably due to its intractability (Gabaix, Laibson, Moloche, and Weinberg 2006; henceforth GLMW).

By contrast, the experimental literature on multiple attribute search is extensive (Payne, Bettman, & Johnson 1993). Such experiments almost always track subjects´ sequences of information search along with their choices of alternative, which allows for a rich analysis.[4] GLMW conducted a particularly rich multiple attribute search experiment. In their design each task contained a matrix display of eight alternatives by ten attributes, allowed full recall, and had no order restrictions. Relative to previous multiple attribute search experiments, this was a large display, inducing high cognitive load; and GLMW remarked that subjects seemed to be suffering from memory limitations.

Sanjurjo (2012a) builds on the work of GLMW by providing a partial characterization of optimal search in their rich search environment. He finds that subjects systematically deviate from optimal search in two main ways: they search too deeply within alternatives (at the expense of searching other alternatives), and when they switch between alternatives they exhibit a strong adjacency bias, which is costly. While it is not clear what is driving these deviations, they cannot be explained by myopia, costly re-optimization, or a "western-reading bias" (Sanjurjo 2012a). By contrast, a theoretical study of memory load in search by Sanjurjo (2012b) shows how the main deviations *are* both consistent with memory load minimization. He does this by showing how the *order* that information is searched in determines the memory load experienced by the searcher.

---

[3] A consumer who wants to buy a car may compare alternatives based on a type of "internal search" in which she recalls certain attributes of each alternative in some order. Likewise, she may instead choose to walk onto a car lot, in which a salesman will describe to her various attributes of various alternatives, in some order. Alternatively, she may instead stay at home, type "compare cars" into an internet search engine, and find a long list of websites that compare alternatives across attributes in tidy matrix displays- allowing her the possibility of searching in whatever order she prefers. At the time this paper was written the first three Google hits were websites offering comparison of from 1 to 4 (one allowed a maximum of only 3) cars, and in terms of attributes, the websites displayed 9, 16, and 81, respectively.

[4] Search behavior is almost always tracked using either eye-tracking or Mouselab- a mechanical analog to eye-tracking, in which subjects "click" open closed boxes in order to observe information. Mouselab has been used by many economists, including Crawford (2008), and Camerer et. al (1993). Many, including Reutskeja, Nagel, Camerer, and Rangel (2011), have used eye-tracking.

Presumably, if subjects were indeed systematically adapting their search patterns in order to avoid high memory loads, it would be because experiencing high memory load was in some way costly for them. It seems that the most likely cost to result from increased memory load is increased choice error- a notion supported by a large literature in cognitive psychology (Miller 1956, Cowen 2001). This paper provides an experiment to test whether or not the amount of memory load endogenously incurred during the process of search affects the rate of choice error. If increased memory load is found to increase choice error then this provides a structural explanation for why searchers would want to adapt their search behavior in the ways observed: in order to reduce the rate of choice error. If increased memory load is found to decrease choice error, or have no effect, then the case for memory limitations as an explanation of subjects´ systematic deviations from optimal search will be greatly weakened.

In this paper´s experimental design the value of each searchable alternative is simply the sum of its attribute values, as assumed in GLMW and common elsewhere in the literature. The attribute values are random variables generated by a process explained clearly to subjects. In half of the tasks each alternative is represented by a row in the matrix of attributes. The experimental manipulation of primary interest is the order that subjects are made to search the given matrix of attributes in; the search order will either be left-to-right, top-to-bottom (called AL because it is within alternative intensive) or top-to-bottom, left-to-right (called AT because it is within attribute intensive).[5] In the other half of tasks each alternative is represented by a column in the matrix of attributes, and the above definitions of AL and AT search are reversed. The experimental manipulation of secondary interest is that matrices of attributes are of three different sizes- two alternatives by two attributes, three by three, and four by four. In the within subject design each subject performs 60 search tasks- 20 with each sized matrix, with half of the tasks for each matrix size AL and half AT.[6]

The model of working memory load (henceforth WML) introduced in Sanjurjo (2012b; Section I) predicts that the respective maximum WML´s for AL and AT search in a two alternative by two attribute matrix are the same: 3 and 3. In a three by three

---

[5] Here, the AL search order is how the words on this page are read. AT is the transposed version of this. These are the two archetypal search orders that collectively describe 98% of behavior in the GLMW data (Sanjurjo, 2012a).

[6] Any effect of the specific attribute values in a matrix is controlled for by using the same matrix of values for one subject as an AL searcher and for another subject as an AT searcher. Any effect of the western reading convention is tested for by transposing half of the matrices. Any order effect of the tasks is tested for by scrambling the 60 tasks in two different orders.

matrix they are 3 (AL) and 5 (AT), and in a four by four matrix they are 3 (AL) and 7 (AT). Subjects are paid only if they choose the highest valued alternative in a task. The null of the first tested hypothesis is that, because the same information is being searched for any given matrix size, for both AL and AT search, the frequency of correct choices will be the same for both search orders. The alternative of the first hypothesis is that, because the maximum WML is higher for the AL search order than the AT search order in the three by three and four by four sized matrices, the frequency of correct choices will be higher for AL search than AT search in these matrix sizes. Further, since the maximum WML is the same for AL and AT in the two by two matrix size, there will be no difference in the frequency of correct choices for the two search orders in two by two matrices of attributes.

The null of the second hypothesis is that because the information searched for any given matrix size, by either the AL or AT order, is the same, the differences in frequencies of correct choices between AL and AT search orders will be constant across different matrix sizes. The alternative to the second hypothesis is that because the differences in maximum WML between the AL and AT search orders increases monotonically in matrix size, the differences in the frequencies of correct choices between AL and AT search will increase monotonically with matrix size.

The results are that in the two by two matrices, AL search yields 95% correct choices while AT yields 93%. In the three by three matrices AL search yields 88% and AT search yields 78%. In the four by four matrices AL search yields 77% and AT search yields 49%. For both hypotheses, the null is rejected in favor of the alternative. These results provide strong evidence that the frequency of errors in choice increases markedly with the amount of endogenously induced WML, in rich multiple attribute search environments. One implication of these results is that while deviating from “optimal” search is costly in the sense that slightly less valuable information will be searched, the subsequent processing of that information will be made easier. A further implication of these results is that strictly the *order* that information is searched in is a fundamental determinant of choice behavior, independent of the *type* or *amount* of information searched.

While this paper is most closely tied to the search literature, it also clearly relates to the cognitive load literature, and the cognitive skills literature. Cognitive load experiments (Shiv and Fedorikhin 1999; Duffy and Smith 2012) have subjects perform a particular decision task while concurrently loaded with some other cognitively demanding task, such as remembering a seven digit number. The experimental design presented here differs fundamentally in that cognitive load is not induced exogenously in a separate task, but endogenously, within the natural context of the decision task of interest. Experiments in the cognitive skills literature (Burks, Carpenter, Götte, and Rustichini 2009; Dohmen, Falk, Huffman, and Sunde 2005) have subjects perform a decision task, and also perform separate tests of cognitive abilities; it is then checked if the test-based measure of a subject´s cognitive abilities is correlated with the subject´s performance in the decision task. Again, the experimental design presented here differs fundamentally in that its subjects only perform one decision task of interest, and within the context of that task, the independent variable of interest (memory load) is manipulated directly.

Section I states the tested hypotheses formally. Section II provides the experimental design. Section III presents the results. Appendix I provides the instructions used in the experiment. Appendix II gives an alternative formulation of the hypotheses which accounts for not only the maximum WML of a search sequence, but the aggregate WML as well.

I. Hypotheses

Sanjurjo (2012b) introduces a model of working memory load (henceforth WML) in search that imposes minimal structure on the memory process. The model takes any search sequence over a well-defined matrix of attributes, and calculates the endogenously incurred WML for each step of the search sequence. Thus, the maximum WML of any search sequence, along with the aggregate WML of the sequence, are easily computed. Maximum WML seems to be the more cognitively binding of the two measures, so tends to be exclusively focused on in the cognitive psychology literature (Cowen, 2001) so this same approach will be taken here.[7] Sanjurjo´s model of WML predicts that the maximum WML of two by two, three by three, and four by four matrices of attributes, according to either AL or AT search (as defined in the introduction), is as appears in Table 1, directly below.

---

[7] See Appendix II for a set of hypotheses that contain predictions as a function of both maximum and aggregate WML. Those hypotheses are almost identical to the ones presented here, and yield the same qualitative results if tested on the data.

Matrix Size	AL/AT	Maximum WML
2x2	AL	3
	AT	3
3x3	AL	3
	AT	5
4x4	AL	3
	AT	7

Table 1: Maximum WML as a function of search sequence and matrix size

Table 1 shows that there is clearly no separation in the maximum WML between AL and AT search in the two by two matrix of attributes. However, in the three by three matrix there is a difference of two units of maximum WML and in the four by four matrix there is a difference of four units of maximum WML.

The null and alternative for hypotheses one and two are presented here, more formally than in the introduction:

$$\begin{aligned} H1_0: & \ AL_2 = AT_2, \\ & \ AL_3 = AT_3, \\ & \ AL_4 = AT_4 \end{aligned}$$

$$\begin{aligned} H1_A: & \ AL_2 = AT_2, \\ & \ AL_3 > AT_3, \\ & \ AL_4 > AT_4 \end{aligned}$$

where O_m is the frequency of correct choice of the search order $O \in \{AL, AT\}$ for matrix size $m \in \{2, 3, 4\}$.

$$H2_0: AL_2 - AT_2 = AL_3 - AT_3 = AL_4 - AT_4$$

$$H2_A: AL_2 - AT_2 < AL_3 - AT_3 < AL_4 - AT_4$$

II. Experimental Design

The experiment was conducted at the LaTEx laboratory at the University of Alicante, in Spain. 58 subjects were chosen at random from a pool of 878 students who had recently registered to participate in decision making experiments. The recruitment was campus-wide and registered all university undergraduate students who responded affirmatively to the advertisement. Six experimental sessions were conducted during the week of October 29, 2010 to November 5, 2010. Sessions were held to 12 or fewer students, who were monitored closely by three research assistants in order to assure that no external memory aids of any type were being used during the experiment.[8]

The experimental design was within subject; each subject responded to 60 search tasks, using the Mouselab experimental interface.[9] Each search task contained either two, three, or four alternatives. Each alternative´s value was initially unknown to the subject, but would be fully revealed over the process of search. In the tasks in which there were two alternatives, the value of each was determined by the sum of its two attribute values. For tasks with three alternatives, the sum of three attributes determined the value of each alternative, and for tasks with four alternatives the sum of four attributes determined the value of each alternative. The list of alternatives in each search task was displayed in the form of a matrix of attribute values, where each alternative was represented clearly by a row of the matrix, or alternatively each was represented as a column. Attribute values were revealed to the subject one at a time. While the order that attributes were searched in was fixed by the experimenter, the rhythm was not; subjects could take as much (or as little) time as they wanted before moving on to observe the next attribute value.[10]

Once the subject had searched each attribute of the matrix one and only one time he/she was asked to choose an alternative. In each task subjects knew that one of the alternatives had the unique largest value. At the end of the experiment, subjects knew that four of the 60 tasks would be chosen at random and that four euros would be paid for each of those tasks in which they had chosen the highest valued alternative; otherwise no money would be paid for that task. This variable reward was added to a

---

[8] Two pilots were run previous to the experimental sessions. One pilot was run on the students from a PhD level experimental economics course at the University of Alicante. The other was run on colleagues-young assistant professors from the University of Alicante.

[9] Mouselab is a mechanical analog to eye-tracking, in which subjects "click" open closed boxes in order to observe information. Mouselab has been used by many economists, including Crawford (2008), and Camerer, Johnson, Rymon, and Sen (1993).

[10] Rushing was discouraged by subjects knowing that they could only leave after a minimum of spending 40 minutes on the search tasks (and a short questionnaire), which was plenty of time to perform all of the search tasks at a casual pace.

five euro show up fee for each subject.

While attribute values were initially unknown to subjects, the random process that determined the attribute values was fully explained to the subjects. The main objective of the random process designed by the experimenter was to make it difficult for the subject to choose the highest valued alternative at a rate better than chance without assimilating each and every searched attribute value. This was emphasized to subjects in the instructions. Because the order of search in each task was fixed by the experimenter, the search task simplifies into a memory exercise- the question being whether the subject can correctly choose the highest valued alternative as a function of the order she is made to search the attribute values in. As such, the design aims to make success in the task a function of memory capacity, and nothing else.

The random process that was used to generate the matrices of attributes is as follows. First each attribute value in a two by two, three by three, or four by four matrix is independently drawn from a uniform distribution over the integer values from -10 to 10. In order for such a matrix to then be used in the experiment, it must survive three filters. The first filter requires that the first attribute of each alternative, and the final sum value of each alternative, has a unique value. Having a non-unique first attribute value would allow the subject to possibly "chunk" (Miller, 1956) the memory of the non-unique value for multiple alternatives, thus effectively reducing experienced WML. Having a unique final sum value for each alternative provides the most precise target for correct choice. The second filter requires that differences in the running alternative sums between any two alternatives never exceed 5 after any number of attributes has been searched for all of the alternatives in the task. Having the alternatives in a "close race" makes it unappealing for the searcher to stop attending to a particular alternative because it is "too far behind," which would also reduce experienced WML. The third and final filter is that after any strict subset of attributes have been searched for all alternatives, it is equally likely that as additional attributes are searched for all alternatives, the current highest valued alternative will change positions any number of times. For example, if a task has three alternatives, and after searching the first attribute for each of the alternatives, the current highest valued alternative is the first, then as the second and third attributes are being searched for all three alternatives it is equally likely that the position of the highest valued alternative will switch one time, two times, or not at all. The purpose of this filter is to give subjects incentive to continue attending to attribute values as they search, rather than feeling comfortable ignoring later attributes, thus reducing experienced WML.[11]

There were two primary experimental manipulations. The first was that in any given task the subject would either have to search in the AL or AT order.[12] The second was that any given task could require search in a two alternative by two attribute, three by three, or four by four matrix of attributes. The purpose of these two primary experimental manipulations was to provide a clear test for the effects on choice of variable endogenously incurred WML, as a function of search order. The hypotheses presented in Section I were constructed with the Sanjurjo (2012b) model of WML in search expressly in mind.

The design was within subject, and each subject responded to 60 search tasks, which included 20 two by two matrices of attributes, 20 three by three, and 20 four by four. For each matrix size 10 tasks required AL search and 10 required AT search. 5 of the 10 tasks of each type had alternatives represented as rows in the matrix of attributes, and 5 had the alternatives represented as columns.

The order of the 60 tasks given to each subject was randomly scrambled in one way for half of the subjects, and in another way for the other half of subjects. This allowed for a test of task order effect. As half of the tasks given to each subject had alternatives represented as rows, and the other half as columns, a possible subject preference or relative facility of one search pattern (horizontal vs. vertical) over the other was controlled for. In addition, this design feature allowed for a test of the "western reading effect." Sanjurjo (2012a) discusses how one possible interpretation of subjects´ systematic deviations from optimality in GLMW´s experiment is that subjects are simply biased towards reading the matrix of attributes as they would read a book. This experimental design provides a clear test of whether subjects actually perform relatively better when the search order is western reading versus its transposed cousin. Finally, a possible effect due to the particular attribute values drawn in the tasks faced by a subject was controlled for by implementing a counter-balancing scheme- 60 matrices of attributes were generated in all. Half of the subjects would search any particular matrix of attributes in the AL order while the other half of subjects would search the same matrix of attributes in the AT order.

---

[11] Subjects were explained clearly that the purpose of the filter was to make it hard for them to do well at the task without attending to all of the attributes in the matrix. If they wanted to pay attention to the specifics of the data generating process they were able to, as the instructions were read aloud to all. For more details see the instructions in Appendix I.

[12] See the introduction for definitions of the AL and AT search orders.

III. Results

The nulls of both hypotheses stated in Section I are rejected by the data, in favor of the alternative hypotheses. All things equal, as the amount of WML endogenously induced by a search order increases the frequencies of correct choices fall. Also, as the differences in AL and AT´s endogenously induced WML´s increase with matrix size, the differences in the frequencies of correct choices increase as well. In addition, no order effects are found, but a slight reading effect is (which is relatively small compared to the effect of WML).

Subjects´ frequencies of correct choices for each matrix size and search order pairing are given in Table 2.

Matrix Size	AL/AT	Maximum WML	Correct (%)
2x2	AL	3	95
	AT	3	93
3x3	AL	3	88
	AT	5	78
4x4	AL	3	77
	AT	7	49

Table 2: Frequency of correct choices as a function of search order and matrix size

The null of the first hypothesis states that because the same information is being searched for the same matrix size, no matter what the order of search, there will be no difference in the frequencies of correct choices following either search order, for any of the matrix sizes. This null is rejected in favor of the alternative hypothesis- namely, that the frequency of correct choices will be higher for AL search than AT search in the three by three and four by four matrix sizes, due to AL´s lower maximum WML, while the frequency of correct choices will be the same in the two by two matrix size, where AL and AT yield the same maximum WML. Table 2 shows that in the two by two matrices of attributes AL searchers make the correct choice in 95% of tasks while AT searchers make the correct choice in 93% of tasks; this difference is not statistically significant.[13] Table 2 shows that in the three by three matrix AL searchers make the

---

[13] The Wilcoxon signed rank test (5% level) is used to compare frequencies and differences in

correct choice in 88% of tasks, while AT searchers make the correct choice in 78% of tasks; this difference is statistically significant. Likewise, in the four by four matrix, AL searchers make the correct choice in 77% of tasks, while AT searchers make the correct choice in 49% of tasks; this difference is statistically significant as well.[14]

The null of the second hypothesis states that because the information searched for any given matrix size, by either the AL or AT order, is the same, the differences in frequencies of correct choices between AL and AT search orders will be constant across matrix sizes. The null is rejected in favor of the alternative hypothesis- namely, that as the matrix size grows larger, and thus the difference in the maximum WML´s between AL and AT search, the differences in frequencies of correct choices between AL and AT search will increase monotonically. Table 2 shows that in the two by two matrix the difference in frequency of correct choice between AL and AT search orders is 2%, while the difference goes up to 10% in the three by three matrix, and 28% in the four by four matrix. The difference between 2% and 10% is statistically significant, as is the difference between 10% and 28%.

Thus the null of both hypotheses tested are rejected in favor of their respective alternative hypotheses. As the maximum WML of a search sequence increases, all else equal, so too does the frequency of choice error. Also, as the differences in the maximum WML´s between two search orders of the same matrix size increase, so too do the differences in the frequencies of correct choices.

The experimental design also allowed for the testing of several “second order” effects; task order effects, reading effects, and effects of the particular values of attribute matrices.

The order in which subjects received search tasks was scrambled randomly in two different ways. There is no order effect in the frequency of correct choices, pooling across all three matrix sizes (80% vs. 81%). In addition, neither of the two hypotheses from Section I can be rejected for either of the two randomly selected task orderings.

A test for a “reading effect” was performed by making alternatives represented by rows of the matrix of attributes for half of each subject´s tasks, and columns in the

---

frequencies. Paired T-tests yield the same test results.

[14] Two of the 60 tasks were removed from the data, due to a data entry error by the experimenter. Namely, the sign of an attribute value was switched in each of the two tasks. Results with or without these tasks removed are identical in terms of the results of hypotheses tests, and more generally are virtually identical in every regard. In addition, it seems very unlikely that this data entry error affected subjects´ choices in the other 58 tasks at all.

other half of tasks. Presumably, students may perform differently when they are searching the matrix of attributes in the same pattern that they read a book in, as opposed to this search order transposed. When pooling all three matrix sizes, subjects get 83% of choices correct when reading, as if in a book, and get 77% of choices correct when the reading pattern is transposed; this difference is statistically significant.[15]

While the same 60 tasks were used for all subjects, half of the subjects searched any given task in the AL order while the other half searched the task in the AT order. In this way the particular values drawn in attribute matrices could be controlled for while the search order was manipulated. A check of performance comparing the two halves of subjects shows no statistically significant differences in the frequencies of correct choices, when pooling all of the tasks.

One can directly compare the relative magnitudes of the WML effect, the reading effect, and the order effect. Because, by construction, there is only variation in maximum WML in the three by three and four by four matrix sizes, the comparison will pool data from both. The WML effect is 18%- AL searchers make the correct choice in 82% of tasks, while AT searchers do so in 64% of tasks. The reading effect is 7%- tasks in which one searches as they read a book yields 77% correct choices, while this search pattern transposed yields 70% correct choices. The order effect is 0%- 73% correct choices for both search orders.

IV. Conclusion

Recent research has shown that in rich search experiments subjects systematically deviate from optimal search in two main ways: they search too deeply within alternatives (at the expense of searching other alternatives) and they have a strong adjacency bias when switching from one alternative to another. While the cause of these costly deviations is unknown, it is clear that candidate explanations of myopia, costly re-optimization, and “western reading bias” are unable to adequately explain the deviations. On the other hand, deviations were found to be consistent with memory

---

[15] If one goes further and compares (only) AL tasks in which alternatives are represented by rows vs. columns, the difference in frequencies of correct choices, when pooling matrix sizes, remains statistically significant- 90% (rows) vs. 83% (columns). This is an intriguing finding given that displays in which alternatives are represented by columns seem very common on internet websites, and that the majority of search in rich search tasks has been observed to be AL (Sanjurjo, 2012a). Are information displays commonly found on the internet more difficult for people to process than they need to be? When typing “compare cars” into google, of the first 10 websites that compared multiple cars over multiple attributes, eight of them had alternatives represented as columns, presumably because scrolling vertically is more natural than scrolling horizontally.

load minimization- a finding which begged for further investigation. This paper reported the results of an experiment designed to test whether or not changes in search memory load affect subjects´ choice error rates. To the extent that search memory load is *not* tied to choice error rates then the fact that deviations are consistent with reduced memory load seems coincidental; to the extent that search memory load *is* tied to choice error rates it suggests that the deviations may likely be a form of adaptive rationality.

The null of the first tested hypothesis was that, because the same information is being searched for any given matrix size, for both AL and AT search, the frequency of correct choices would be the same for both search orders. The alternative of the first hypothesis was that, because the maximum WML was higher for the AL search order than the AT search order in the three by three and four by four sized matrices, the frequency of correct choices would be higher for AL search than AT search in these matrix sizes. Further, since the maximum WML was the same for AL and AT in the two by two matrix size, there would be no difference in the frequency of correct choices for the two search orders in the two by two matrix size.

The null of the second hypothesis was that because the information searched for any given matrix size, by either the AL or AT order, was the same, the differences in frequencies of correct choices between AL and AT search orders would be constant across different matrix sizes. The alternative to the second hypothesis was that because the differences in maximum WML between the AL and AT search orders increases monotonically in matrix size, the differences in the frequencies of correct choices between AL and AT search would increase monotonically with the size of the matrix.

The results show that both null hypotheses are rejected in favor of the alternative hypotheses. These results suggest that the systematic deviations from optimal search recently observed in rich search problems indicate a form of "adaptive rationality" by subjects; they intentionally search in orders that induce less WML (at the cost of searching slightly less valuable information) so that they can then be more likely to process the information they have searched correctly.

These results also show that the *order* that information is searched in can be critical to consequent decision making, aside from the *type* or *quantity* of that information. To homo-economicus the order that information is searched in is irrelevant, because working memory capacity is unlimited. By contrast, it has long been recognized in the cognitive psychology literature that human working memory capacity is finite, and small. Thus, for mortals, the order that information is searched in is inextricably linked with the working memory load induced, to the extent that the information can be grouped in natural ways, as in the case of choice from a set of alternatives evaluated over multiple attributes.

Such an insight begs further testing given that decision making in experimental environments is often based on information provided to the subject by the experimenter. Because the experimenter can monitor the order in which the subject acquires information, it is straightforward to estimate induced working memory load as a function of that order. Perhaps there are additional observed systematic decision making behaviors that have previously been considered puzzling, or misattributed to some other cause, that are actually due to a general tendency of people to try to minimize the amount of working memory load that they experience. Further, perhaps we, as decision makers, are primarily concerned with making the best choice based on our information, and not so concerned with whether the information that we have is the best.

Appendix I- Instructions[16]

Overview

Welcome!

Today you will be asked to make a series of choices, for which you will be rewarded more or less money depending on the choices you make. I expect the average participant to earn around 16 euros but it is possible to earn as much as 21 or as little as 5, depending on the choices you make. You should expect to be here for around one hour. You will be paid as you leave.

You will make 60 choices in all; for each choice you will choose one alternative from a set of either two, three, or four alternatives.

I will now give more details about the choices you will be making and how you will be paid, then you will see a demo with three examples of exactly what your choice tasks will look like, then you will start the series of 60 paid choice tasks. After you finish the paid choice tasks you will be asked to fill out a questionnaire. After the questionnaire you will be paid. As we go through the instructions please know that there is a strict culture of honesty in economics experiments, so you will not be mislead or deceived in any way. It is also very important that you do not communicate with anyone until after you have finished the experiment and left the room. Also you will not be able to use

---

[16] The instructions provided here are an English translation (from Spanish) of the instructions given to the students that participated in the experiment. The author would like to thank Maria Jose Aragón for doing the bulk of the translation.

paper and pencil, or anything other than what you are instructed to use on the computer display. If any of these conditions are violated we will unfortunately have to ask you to leave and you will not be paid.

More about the tasks

For each task you will need to choose from a set of either two, three, or four alternatives. Each alternative contains multiple characteristics, and the value of each alternative is simply the sum of its own characteristics. The way to earn the most money is to choose the alternative with the highest value in each task. When there are two alternatives to choose from each alternative consists of two characteristics, when there are three alternatives each consists of three characteristics, and when there are four alternatives each alternative consists of four characteristics. In any choice task, the alternatives can either be represented as rows or columns, for the duration of the task. An example of the two different ways alternatives can be represented is shown directly below:

alternativa a	alternativa b	Alternativa c

Alternativa a			
Alternativa b			
Alternativa c			

A very important feature of the tasks is that rather than see all characteristic values at the same time, instead **you will see only one characteristic value at a time**. Consider a two alternative choice task in which the alternatives are represented as rows and characteristic values are shown according to Order 1. The four pages you would see in sequence would look like this:

Page 1:

alternativa a	0	
alternativa b		

Page 2:

alternativa a		5

alternativa b		

Page 3:

alternativa a		
alternativa b	-2	

Page 4:

alternativa a		
alternativa b		6

You will observe each page once and only once, and you will control when you go to the next page by clicking on a button labeled “next page” on the bottom of the page.

For expositional purposes I now show all characteristic values together (though you will never see this in the experiment) along with calculated alternative values.

alternativa a	0	5
alternativa b	-2	6

Value of alternative a = 0 + 5 = 5
Value of alternative b = -2 + 6 = 4

For each of the two, three, and four alternative tasks there are two distinct orders that the characteristic values can be shown to you. You will be told throughout each task which order the characteristic values will be shown to you in (so you do not have to remember the orders). The two orders that characteristic values can be shown to you in the two alternative task, for example, are

Order 1:

alternativa a	1- first	2- second
alternativa b	3- third	4- fourth

alternativa a	alternativa b
1- first	2- second
3- third	4- fourth

Or

Order 2:

alternativa a	1- first	3- tirad
alternativa b	2- second	4 - fourth

alternativa a	alternativa b
1- first	3- third
2- second	4- fourth

The two orders in the three and four alternative tasks are similar.

Notice that there are three things that can change from one choice task to the next:

1. the number of alternatives (thus characteristics also)
2. whether the alternatives are represented as rows or columns
3. which of the two orders the characteristic values will be shown in

You will choose in 20 two alternative choice tasks, 20 three alternative choice tasks, and 20 four alternative choice tasks. For each number of alternatives 10 choice tasks will have alternatives represented as rows, and 10 as columns. Further, 5 of the 10 choice tasks in which alternatives are columns will have order 1 and five will have order 2; the same is true of choice tasks in which alternatives are represented as columns. The order of these choice tasks was chosen randomly before the beginning of the experiment.

Characteristic Values

The characteristic values for each choice task can be at lowest -10 and at highest 10. They were chosen using a random method designed to make it difficult to guess which alternative has the highest value without summing all of the observed characteristic values for each alternative.

In case you want more details on how the characteristic values were generated (optional):

> -- First, all characteristic values (integers) between -10 and 10 are drawn with equal probability, independently (so the drawing of any particular characteristic value in no way affects draws of other characteristic values).

Then, in order to be selected as the characteristic values for one of your choice tasks the values also had to comply with the following three rules:

> -- The maximum difference between any two temporary alternative value sums is five.
>
> -- It is equally likely that the temporary *highest* alternative value sum changes any number of times after the temporary alternative value sum that consists of one characteristic only. For example, in a four alternative task, if alternative A is the temporary highest alternative value sum after seeing only the first characteristic value for each alternative, then as temporary sums including two, three, and four characteristic values are compared, it is equally likely that the highest temporary alternative value sum changes zero, one, two, or three times.
>
> -- Initial and final alternative value sums are never equal across alternatives.

If this seems confusing do not worry. What is important is that you understand that it is difficult to "guess" the highest alternative value by considering only some of the characteristic values. The only way to ensure that you select the highest value alternative and report its correct value is by using all characteristic values.

Payment

You will be paid 5 euros for simply showing up and following the instructions of the experiment. In addition, after you have finished we will check the answers for 4 tasks that were chosen at random before the start of the experiment (we will show you the numbers from inside of a sealed envelope that we will open). For each of these 4 tasks you will be paid 4 euros for a correct answer. Thus, the most money you can earn is 21 euros, and the least is 5.

Please keep in mind that you will need to remain in your seat, at your computer for at least 40 minutes once your choice tasks begin, in order to be eligible for payment. Thus, there is no reason to rush through the experiment. Once at least 40 minutes have passed (we will indicate this clearly on the chalk board) you may raise your hand to indicate that you have finished the decision tasks and the questionairre. Please wait silently while we calculate your payoff; this may take a few minutes. Once we have calculated your payoff we will call you one by one and lead you outside of the room where we will pay you according to your performance in the 4 tasks chosen at random; also, we will give you a sheet of paper indicating which of these tasks you answered correctly and which you answered incorrectly. If you would like to verify incorrect answers, or the numbers selected at random to determine the 4 paid tasks, you need only ask us to see the official data when we are giving you your payment.

Summary

Remember, throughout each choice task you will know how many alternatives you are choosing between, whether the alternatives are represented by alternatives or columns, and what order the characteristic values will be shown to you in. The more choices in which you choose the highest value alternative, the higher you can expect your earnings to be. As the experiment will require a good deal of attention, please make an effort not to distract other students during the duration of the experiment.

Examples

We will now go through three examples in order to see more precisely what the choice tasks you will be seeing look like. Unlike when your choice tasks begin, for these examples we will show the calculated value of each alternative, and thus which alternative value is higher.

After finishing the demo you will have an opportunity to ask questions before starting the paid portion of the experiment. It is important that you ask any questions that you have then as you will not be able to ask questions once the paid portion of the experiment begins.

We will now start the demo.

Appendix II- Hypotheses With Aggregate WML

While the working memory load literature in cognitive psychology is sharply focused on the maximum WML incurred during the decision making process (Cowen 2001), the model of WML in search introduced by Sanjurjo (2012b) allows also for computing the estimated aggregate WML. Table 3 shows both the maximum and aggregate WML´s for the AL and AT search orders in matrices of attributes of size two by two, three by three, and four by four.

Matrix Size	AL/AT	Maximum WML	Aggregate WML
2x2	AL	3	5
	AT	3	7
3x3	AL	3	17
	AT	5	32
4x4	AL	3	35
	AT	7	87

Table 3: Maximum and aggregate WML as a function of search order and matrix size

Aggregate WML may be useful as a second order measure of memory load, given that it measures sustained induced working memory load. If it were believed that aggregate WML might have even a weak effect on predicting the frequency of correct choices, then a revised pair of hypotheses from Section I could look like the following:

$$\begin{aligned} H1_0: \mathrm{AL}_2 &= \mathrm{AT}_2, \\ \mathrm{AL}_3 &= \mathrm{AT}_3, \\ \mathrm{AL}_4 &= \mathrm{AT}_4 \end{aligned}$$

$$\begin{aligned} H1_A: \mathrm{AL}_2 &\geq \mathrm{AT}_2, \\ \mathrm{AL}_3 &> \mathrm{AT}_3, \\ \mathrm{AL}_4 &> \mathrm{AT}_4 \end{aligned}$$

where O_m is the frequency of correct choice of the search order $O \in \{\mathrm{AL}, \mathrm{AT}\}$ for matrix size $m \in \{2, 3, 4\}$.

$$H2_0: \mathrm{AL}_2 - \mathrm{AT}_2 = \mathrm{AL}_3 - \mathrm{AT}_3 = \mathrm{AL}_4 - \mathrm{AT}_4$$

$$H2_A: \mathrm{AL}_2 - \mathrm{AT}_2 < \mathrm{AL}_3 - \mathrm{AT}_3 < \mathrm{AL}_4 - \mathrm{AT}_4$$

Thus the only change relative to the hypotheses presented in Section I, which only depend on maximum WML, occurs in the alternative of H1 for two by two matrices of attributes. Because the aggregate WML of AT search is greater than AL search in this case the alternative hypothesis is that for two by two matrices of attributes the AL order will yield a weakly larger frequency of correct choices (rather than the same frequency). Under these alternative hypotheses the results are the same as those presented in Section III: both hypotheses are rejected in favor of their alternatives.

References

Bearden, Neil, and Terry Connolly (2007): “Multi-attribute sequential search,” *Organizational Behavior & Human Decision Processes, 103,* 147-158.

Burks, Stephen, Jeffrey Carpenter, Lorenz Götte, and Aldo Rustichini (2009): “Cognitive Skills Explain Economic Preferences, Strategic Behavior, and Job Attachment,” *Proceedings of the National Academy of Sciences*, 106(19), 7745-7750.

Camerer, Colin, Eric J. Johnson, Talia Rymon, and Sankar Sen (1993): “Cognition and Framing in Sequential Bargaining for Gains and Losses,” *Frontiers of Game Theory,* ed. Ken Binmore, Alan Kirman, and Piero Tani, 27-47. Cambridge, MA: MIT Press.

Cowan, N. (2001): “The Magical Number 4 in Short-term Memory: A Reconsideration of Mental Storage Capacity,” *Behavioral and Brain Sciences* 24 (1).

Crawford, Vincent P. (2008): "Look-ups as the Windows of the Strategic Soul: Studying Cognition via Information Search in Game Experiments," in Andrew Caplin and Andrew Schotter, editors, *Perspectives on the Future of Economics: Positive and Normative Foundations,* V1, *Handbooks of Economic Methodologies,* Oxford University Press.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2010): "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*, 100(3), 1238-1260.

Dow, James (1991): "Search Decisions with Limited Memory," The Review of Economic Studies, 58, 1-14.

Duffy, Sean, and John Smith (2012): "Cognitive Load in the Multi-Player Prisoner´s Dilemma Game: Are there Brains in Games?" Rutgers University-Camden manuscript.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg (2006): "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model," *American Economic Review*, 96, 1043-1068.

Kohn, Meir and Steven Shavell (1978): "The Theory of Search," *Journal of Economic Theory*, 9, 93-123.

Lippman, Steven and John McCall (1976): "The Economics of Job Search: A Survey," *Economic Inquiry*, 14, 155-189.

Payne, John, James Bettman, and Eric Johnson (1993): *The Adaptive Decision Maker*. Cambridge, U.K.: Cambridge University Press.

Piccione, Michele, and Ariel Rubenstein (2003): "Modeling the Economic Interaction of Agents with Diverse Abilities to Recognize Equilibrium Patterns," Working paper.

Reutskaja, Elena, Rosemarie Nagel, Colin Camerer, and Antonio Rangel (2011): "Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study," American Economic Review, 101(2), 900-926.

Sanjurjo, Adam (2012a): "Search with Multiple Attributes and Alternatives: Theory & Empirics," Working paper.

Sanjurjo, Adam (2012b): "Using a Model of Memory Load to Explain Deviations from Optimal Search," Working paper.

Shiv, Baba, and Alexander Fedorikhin (1999):"Heart and Mind in Conflict: The interplay of Affect and Cognition in Consumer Decision Making," *Journal of*

Consumer Research, 26(3), 278-292.