Title
Essays in decision making under cognitive load

Author
Sanjurjo, Adam Angel

Publication Date
2009

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Essays in Decision Making Under Cognitive Load

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Economics

by

Adam Angel Sanjurjo

Committee in charge:

Professor Vincent P. Crawford, Chair
Professor Richard Carson
Professor John Conlisk
Professor Patrick Fitzsimmons
Professor David Schkade
Professor Joel Sobel

2009

The Dissertation of Adam Angel Sanjurjo is approved, and is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2008

EPIGRAPH

Tis Boolean-
Zero or 1.
There are no other states,
No in between states-
Just confusion.

To be there is to know.
To be elsewhere is to wonder.

It is simple,
Yet mysteriously unknown.

If there is some mystical energy in words
It is because there is structure behind them-
One level beyond our current awareness…

And thus there are words and experiences
Mystical to me
That will no longer be
Once their mechanism is revealed.

To automate and to rise.
To master the technical
And reach the level of emergence
Where one creates using heuristic pieces-
This is greatness
For artist, for scientist,
That magic flash that moves around known
pieces in new ways-
Into a final product that seems too beautiful
for the hand of man.

There is only one state,
And then,
There is confusion.

Balance.
Imbalance.
No in between.

1-18-7

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge all of the advisors on my committee for their assistance and great generosity- Richard Carson for supporting me in running experiments for my second year project, John Conlisk for meeting with me in the early years to discuss information overload, Patrick Fitzsimmons for giving me deeper insight into proof methods for my second chapter, David Schkade for encouraging me in my research of working memory limits, Joel Sobel for challenging me to better understand proof methods in the second chapter, and Vincent P. Crawford for the vast amount of time and energy he devoted towards helping me progress on all fronts, as well as the great wisdom he showed in drawing, not forcing, progress out of me. I would also like to acknowledge Robert Johnson for his great friendship and wisdom, and my family and friends for their love and support.

Chapters 1, 2, and 3 are currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this material. Chapter 1 was co-authored with Peter Katuscak.

VITA

2000	Bachelor of Arts, University of California, Santa Barbara
2000	Student Intern, Securities Exchange Commission
2000-2002	Research Assistant, Reserve Board of Governors
2002-2009	Teaching Assistant, University of California, San Diego
2009	Instructor, University of California, San Diego
2009	Doctor of Philosophy, University of California, San Diego

FIELDS OF STUDY

Psychology and Economics
Professors Richard Carson, John Conlisk, Vincent P. Crawford, David Schkade, Joel Sobel

Mathematics
Professor Patrick Fitzsimmons

ABSTRACT OF THE DISSERTATION

Essays in Decision Making Under Cognitive Load

by

Adam Angel Sanjurjo

Doctor of Philosophy in Economics

University of California, San Diego, 2009

Professor Vincent P. Crawford, Chair

My first chapter tests several hypotheses of information overload in an experiment where subjects estimate security prices under variable signal loads. Peter Katuscak and I find that as information load increases subjects eventually stop assimilating further information, and they shift increasing weight towards the most salient information. This combination of results leaves information receivers vulnerable to strategic manipulation by senders.

My second chapter builds on the multiple attribute search experiment, and analysis, of Gabaix and Laibson (2006) in four ways; I provide a basic description of subjects' search behavior, study behavior on the individual subject level, provide a partial characterization of optimality, and compare subjects' behavior to my partial

characterization of optimality. I find that subjects' search behavior violates optimality at high rates, but is also highly systematic, that 98% of all search behavior can be explained by four simple, exclusive, types, and that subjects often search conditionally too deeply within alternatives and exhibit strong adjacency biases in switching between alternatives. I also observe unambiguous evidence of memory failure by subjects.

My third chapter tests the hypothesis that working memory limits can explain subjects' main systematic deviations from optimality, as well as other fact patterns in search, in the Gabaix and Laibson (2006) dataset. First I show that the most popular type of search pattern by subjects also requires the unique minimum amount of working memory load. Second, I show that more systematic search sequences require less working memory load than more "random-looking" sequences. These theoretical results strongly suggest that a simple model of search in which working memory is limited, but subjects otherwise search optimally, can explain Gabaix and Laibson's subjects' main systematic deviations from optimality as well as other fact patterns from the dataset.

An Experimental Study of Information Overload in Financial Bayesian Weighting Tasks of Varying Signal Loads

Adam Sanjurjo and Peter Katuscak[1]
University of California, San Diego
3 June 2005

Abstract: Information overload, here, refers to the act of assimilating less information in a task that contains strictly more relevant information, but is otherwise identical. We test for information overload experimentally, with subjects estimating the expected prices of securities based on variable probabilistic signal loads. A standard Bayesian updating model is used both to generate criterion expected prices, and to econometrically estimate the extent of subjects' assimilation of signals. Three hypotheses related to information overload, and one regarding the heuristic strategy of placing an excessively large amount of weight on stronger signals, and an excessively small amount of weight on weaker signals (Polarization), are tested. As the Bayesian regression model is highly sensitive to extreme expected price estimates, two sets of results are reported. The first set includes all data gathered in the experiment, including extreme expected price estimates, and shows that while decision error increases with information load there is no evidence of overload or polarization. The second set of results truncates the 4% most extreme expected price estimates, allowing for all subjects' estimates to be weighted virtually equally in our analysis, as they should be, unlike in the non-truncated dataset. In this case we show that as information load increases decision

[1] Department of Economics, University of California, San Diego. 9500 Gilman Drive, La Jolla, CA 92093-0508 (Adam's email: asanjurjo@ucsd.edu; Peter's email: Katuscak@cerge-ei.cz). Adam would like to thank Richard Carson, Vincent Crawford, John Conlisk, Armin Falk, Dan Houser, Andrew Schotter, Reinhard Selten, and Yixao Sun for helpful comments.

error increases, the amount of information assimilated plateaus, and subjects shift a disproportionately large amount of weight to the strongest signals.

"A wealth of information creates a poverty of attention." -H. Simon

Introduction

Traditional economic models tend to assume that decision makers have a scarcity of information and therefore choose to engage in costly search (Stigler, 1961; Kohn and Shavell, 1974) in order to acquire more, and that disposal of unwanted information is free. Hand in hand with these assumptions goes the prevalent belief within economics that "more information is better."

However, with the creation and vast mobilization of the internet over the last 40 years, in today's world information is not so much scarce as it is abundant, immediately available, and virtually costless. At the same time, human information processing limits are an ever-present reality (Miller, 1956, Cowen, 2001). Thus it is a non-trivial question to ask how human beings adapt their decision processes as information load increases.

According to Jacoby (1977) human beings have a finite ability to assimilate and process information during any given unit of time. Once these limits are surpassed, the human system is said to be "overloaded" and decision making "becomes confused, less accurate, and less effective."

In this paper we define information overload (henceforth IOL) as the act of processing relatively less information when relatively more is provided, on an identical task. This definition is concrete, measurable, easy to interpret, and seems to capture the main essence of Jacoby's definition, in a practical way. We test this and other related hypotheses in an experimental setting.

In our experiment subjects estimate the expected value of a financial security based on a prior along with probabilistic signals of varying quantities and strengths. We use a within subject design in which each subject performs 24 rounds of estimates, each containing three, six, or nine signals of information. Each signal can be of weak, moderate, or strong strength, in a way I will define clearly in Section I. Subjects are compensated according to the precision of their estimates relative to computed (Bayesian) criterion expected prices.

Stated informally, our hypotheses are that subjects' estimates will stray further from the criterion expected price as the number of signals increases (weak IOL), that as the number of signals increases from six to nine subjects will assimilate the same (moderate IOL) to strictly less (strong IOL) information, and that as the number of signals increases subjects will disproportionately over-weigh high strength signals and disproportionately under-weigh low strength signals (polarization).

Our results are reported in two sets. The first set includes all subject data, without accounting for the Bayesian model's extreme sensitivity to extreme expected price estimates (explanation in Section IV). In this case we show that as the number of information signals increase subjects stray further from the criterion price, but assimilate more information and show no signs of polarization.

In the second set of results we account for the Bayesian model's extreme sensitivity to extreme expected price estimates by truncating the 4% most extreme estimates. This truncation allows for all subjects' estimates to be weighted virtually equally in our analysis, as they should be, unlike in the non-truncated dataset. In this case we show that as the number of information signals increase subjects stray further from the criterion price, and as they go from six to nine signals they assimilate no more information, and disproportionately over-weigh the strongest signals.

The combination of subjects' plateauing level of assimilated information and a shifting of disproportionate weight toward the strongest signals suggests that increases in information load allow information senders a clear opportunity for the strategic manipulation of receivers.

While it is difficult to study IOL directly in field studies, laboratory experiments offer an effective alternative environment in which extraneous variables can be controlled, decision accuracy can be measured objectively, and signals can be made both relevant and independent from one another so that the effect of information load on subjects can be properly isolated. This paper accomplishes these ends more fully than previous work in the IOL literature. While our design borrows heavily from those of Tuttle & Burton (1999, 2004), it is the first in the IOL literature to use a Bayesian updating (supported by decision theory) rather than linear framework to model subjects' expected price estimates conditional on market signals. It is also simpler than previous IOL experimental designs in ways that allow for a clearer isolation of the pure variable effects of increasing information load.

This paper is composed of five sections. Section I contains the experimental design, while experimental hypotheses are found in Section II. Section III describes the Bayesian econometric model used to measure subjects' assimilation of informative signals. Section IV describes the experimental analysis, and Section V concludes.

Following the conclusion are two appendices; the first contains the instructions from our experiment and the second contains a brief literature review of IOL that serves to motivate our hypotheses and experimental design.

I. Experimental Design

First we describe the subjects, then the general structure including payoffs, and finally the procedure used in our experiment.

Subjects

The subject pool consisted of 176 UCSD undergraduate students from a Financial Investments course.

General Structure

In this experiment subjects are told they will be doing the job of a securities analyst. A new stock will be worth either $0 or $100 in one year from now- $0 if the company does poorly, $100 if it does well. Subjects start with a given prior of a .5 probability the company will do well and a .5 probability the company will do poorly. Subjects are then shown a set of probabilistic signals with which they update their prior in order to make an expected price estimate between $0 and $100 (any value) for the hypothetical security. The signals that subjects receive must take one of the two values- "Strong" or "Weak." Strong signals are more likely than Weak signals to appear when the security will do well, and Weak signals are more likely to appear when it will do poorly. Specifically,

$$p(signal = Strong \mid Well) = p(signal = Weak \mid Poorly) = \alpha,$$
$$p(signal = Weak \mid Well) = p(signal = Strong \mid Poorly) = \gamma$$
$$where\ \alpha > \gamma.$$

In addition to taking one of two values, each signal must be one of three types- Company, Economy, or Industry (following Tuttle & Burton, 1999), where

$$\alpha^c > \alpha^e > \alpha^i \ \& \ \gamma^c < \gamma^e < \gamma^i$$

with superscripts indicating signal type; c = Company, e = Economy, i =Industry. [1]

For each security subjects receive either three, six, or nine signals. If they receive three signals there will be one of each type, meaning one Company signal, one Economy signal, and one Industry signal. If they receive six signals there will be two of each type, and with nine signals they will receive three of each type. Table 1.1 shows the possible values each signal can take- "Strong" or "Weak." Table 1.2 shows the varying degrees of information content (signal strength) for signals of each type: Company, Economy, and Industry.

---

[1] Company indicators are meant to provide the most information about a company, then Economy indicators, and Industry indicators are meant to provide the least.

Table 1.1: Each signal can take either a Strong or Weak value

	Strong Value	Weak Value
Company Signals		
Company_1	Strong	Weak
Company_2	Strong	Weak
Company_3	Strong	Weak
Economy Signals		
Economy_1	Strong	Weak
Economy_2	Strong	Weak
Economy_3	Strong	Weak
Industry Signals		
Industry_1	Strong	Weak
Industry_2	Strong	Weak
Industry_3	Strong	Weak

Table 1.2: The information content of each signal type (conditional probabilities)

	Company Indicators		*Economy Indicators*		*Industry Indicators*	
	Strong	**Weak**	**Strong**	**Weak**	**Strong**	**Weak**
Do Well	0.65	0.35	0.60	0.40	0.55	0.45
Do Poorly	0.35	0.65	0.40	0.60	0.45	0.55

More informative signals should influence expected price estimates more than less informative signals. For example, if a subject receives one "Strong" Company signal, she knows that the likelihood ratio is $\left(\frac{.65}{.35}\right)$ that the signal comes from a company that is doing well vs. one that is doing poorly. That is, if the company does well the probability a Strong company signal will be drawn is *p(Strong| Well)* = .65 while the probability of a Weak signal being drawn is only *p(Weak| Well)* = .35. Similarly, if the company does well the likelihood ratio of seeing a Strong vs. a Weak Industry signal would only be $\left(\frac{.55}{.45}\right)$, as the Industry signal is the weakest.

The experiment consists of 33 rounds in all- nine of which are practice, and 24 of which are for credit. In each practice round subjects are given 50 seconds to make an estimate of the expected security price, before being shown the criterion (true) expected price. The criterion expected price is computed using Bayes' rule. In the example round provided in Figure 1.1, for instance, the criterion expected price is calculated by first computing the correct odds ratio given the informative signals, then multiplying it by $100:

Signal:	**Value**
Industry_2	Strong
Economy_1	Weak
Company_2	Strong
Company_1	Weak
Economy_2	Strong
Industry_1	Weak

Figure1.1: Example of a task

$$\frac{p(do\ well)}{p(do\ poorly)} = \left(\frac{.55}{.45}\right)\left(\frac{.40}{.60}\right)\left(\frac{.65}{.35}\right)\left(\frac{.35}{.65}\right)\left(\frac{.60}{.40}\right)\left(\frac{.45}{.55}\right) = \left(\frac{.65}{.35}\right)^{(1-1)}\left(\frac{.60}{.40}\right)^{(1-1)}\left(\frac{.55}{.45}\right)^{(1-1)} = 1:1$$

(First calculate the odds-ratio)

From this result, *p(do well | signals' information)=* $\frac{1}{2}$.

Therefore the expected stock price is equal to $\left(\frac{1}{2}\right) * \$100 = \$50$.

Note that in this example, for each signal type, 1 Strong signal is matched by 1 Weak signal, so all signals offset and leave the Bayesian posterior equal to the prior of *p(do well) = p(do poorly)=* $\frac{1}{2}$.

Once the 24 credit rounds begin subjects no longer receive feedback. The motivation behind giving subjects feedback in the 9 no-credit practice rounds is to allow them to test their individual pricing models against the Bayesian criterion expected price and adjust for the type of "conservatism," or under-reaction to signals that has been systematically observed in the Bayesian experimental literature (Camerer, 1993). In order to mitigate learning effects subjects are not given feedback during credit rounds. Nor are subjects explicitly given the Bayesian odds formula. With the formula at hand this experiment would quickly reduce to an exercise of merely sorting signals and plugging them into a formula. There would be time costs, little increase in difficulty with increasing signal loads, and predetermined model selection. Without being given a clearly specified model, on the other hand, subjects are free to use whatever means they prefer to estimate the expected security price. This feature allows the task to more closely resemble a real world decision environment than a highly stylized laboratory experiment.

Generating Signals

The data generating process used to create signal values involves four steps.

1) Randomly assign signal loads of either three, six, or nine to each of the 33 rounds of the experiment[2]

2) Randomly and independently draw whether the company will "do Well" or "do Poorly" from a binary distribution where both outcomes have probability $\frac{1}{2}$.

3) Conditioning on whether or not the company will do well, randomly draw mutually independent “Strong” or “Weak” values for each signal, using the appropriate probability distributions from Table 1.2.

[2] The data generating process just described is run for two sets of 33 rounds rather than one in order to provide more variation in the signals for our sample size.

4) Finally, for each round signal ordering is randomly scrambled.

An example of the final product generated by steps 1-4 is depicted in Figure 1.1.

Payoffs

The 176 Financial Analysis students were able to earn up to 2% extra credit toward their overall course grade by participating in the experiment. For each round, payoffs were determined by the "dartboard principle." If the subject's estimate was within \$1 of the perfect Bayesian expected price the subject would earn the full 20 points for that round. If the subject's estimated expected price was off by more than \$1 but weakly less than \$2 she would earn 19 points, off by more than \$2 but weakly less than \$3 18 points, and so on. If the subject was off by more than \$19 but weakly less than \$20 she would earn 1 point for that round. If she was off by more than \$20 she would earn no points for that round. Each subject's total payoffs for the 24 credit rounds was determined by the following formula:

$$Extra\ Credit = 2\% * \left(\frac{Aggregate\ Individual\ Score\ for\ 24\ rounds}{480} \right)$$

Subjects were required to fill in expected price estimates for each and every round in order to qualify for any extra credit for the experiment. Average subjects' payoffs were around 1% extra credit for overall course grade.

Procedure

The experiment was run in the regular time slot and location of a financial investments class at UCSD. Students were told about the experiment the previous week, both in class and via e-mail. Students were told they would be able to use calculators. Once students were seated instructions were projected onto a screen for all subjects to see, and were read word for word by one of the experimenters while the other experimenter handed out answer packets. Answer packets consisted of 5 pages. The first page contained answer spaces for all credited 24 rounds, as well as a reminder table of the relative probability weights for the different signal types (Table 1.2). The second and third sheets were scratch pages where subjects were asked to document all written work. Page 4 contained a short survey asking subjects to provide subjective responses to questions regarding the experiment. Page 5 asked subjects to report background information such as major, cumulative GPA, year at UCSD (i.e.- "Junior"), gender, and age; and sign a non-obligatory request of authorization for the experimenters to use their data. Additionally, subjects were provided a scantron sheet to which they were asked to transfer their expected price estimates at the end of the experiment.

There were two versions of the answer packets that only differed from one another in that one had a large "A" on the front page, while the other had a large "B." Versions A and B were passed out so as to alternate ABAB... within each row of subjects. During the instructions subjects were told that the overhead screen would be split into two, with signals for Version A students always on the left and signals for Version B students always on the right. After the instructions were read students had an opportunity to ask questions. They were also told they were free to excuse themselves from the experiment at any time if they did not wish to participate.

Next, students participated in 9 non-credit practice rounds. For each practice round one version A column of signals, and one version B column of signals were simultaneously projected overhead for 50 seconds. After each round the true Bayesian expected price estimate for that round was projected above for 5 seconds. Subjects were asked to write down their estimates before receiving the "true expected price" feedback. After each 5 second interval of feedback, the next round would begin by posting two fresh sets of signals over a 50 second span. At the 40 second mark of each round one of the experimenters would announce "10 seconds left."

After the nine practice rounds were completed subjects were told that the 24 credit rounds would then commence. Subjects were reminded that they would not be receiving feedback for any further rounds, and that they should write down an expected price estimate for each round before moving on to the next.

After the 24 credit rounds were completed subjects had time to transfer their expected price estimates to the provided scantron sheet (if they had not already done so) and fill out the short survey and data use authorization form.

II. Model

The model used in this experiment to determine criterion expected security prices as well as regression coefficients assumes strict Bayesian updating on multiple signals. There are either three, six, or nine signals grouped into three types- each with a different signal strength. For any given round the criterion expected price is calculated by the following formula:

$$E(P) = \$100 * \left(\frac{X}{1+X}\right), \text{ where } X = \left(\frac{.65}{.35}\right)^{\left(\sum 1_c^+ - 1_c^-\right)} * \left(\frac{.60}{.40}\right)^{\left(\sum 1_e^+ - 1_e^-\right)} * \left(\frac{.55}{.45}\right)^{\left(\sum 1_i^+ - 1_i^-\right)}$$

Here, likelihood ratios for Company, Economy, and Industry, respectfully, are found in parentheses. Exponential terms are summations of strong and weak signals for each signal type. 1_c^+ is an indicator variable for a "Strong" Company signal, while 1_c^- is an indicator variable for a "Weak" Company Signal. For example, if round 1 had the three signals as in Figure 1.2:

Signal	**Value**
Company1	Strong
Economy1	Weak
Industry1	Strong

Figure 1.2: three signal example

the criterion expected stock price would be:

$$E(P) = \$100 * \left(\frac{\left(\frac{.65}{.35}\right)^{(1)} * \left(\frac{.60}{.40}\right)^{(-1)} * \left(\frac{.55}{.45}\right)^{(1)}}{1 + \left[\left(\frac{.65}{.35}\right)^{(1)} * \left(\frac{.60}{.40}\right)^{(-1)} * \left(\frac{.55}{.45}\right)^{(1)}\right]} \right) = \$60$$

The regression analysis is run using a block-diagonal construction so as to conduct hypothesis testing across signal load treatments for our hypotheses. The model

$$\begin{aligned}\ln\left(\frac{\frac{y^{sr}}{100}}{1-\frac{y^{sr}}{100}}\right) = {} & T_3\left(c_1^{sr}\beta_{1,3}^{c} + e_1^{sr}\beta_{1,3}^{e} + i_1^{sr}\beta_{1,3}^{i}\right) + \\ & T_6\left(c_1^{sr}\beta_{1,6}^{c} + e_1^{sr}\beta_{1,6}^{e} + i_1^{sr}\beta_{1,6}^{i} + c_2^{sr}\beta_{2,6}^{c} + e_2^{sr}\beta_{2,6}^{e} + i_2^{sr}\beta_{2,6}^{i}\right) + \\ & T_9\left(c_1^{sr}\beta_{1,9}^{c} + e_1^{sr}\beta_{1,9}^{e} + i_1^{sr}\beta_{1,9}^{i} + c_2^{sr}\beta_{2,9}^{c} + e_2^{sr}\beta_{2,9}^{e} + i_2^{sr}\beta_{2,9}^{i} + c_3^{sr}\beta_{3,9}^{c} + e_3^{sr}\beta_{3,9}^{e} + i_3^{sr}\beta_{3,9}^{i}\right) + \varepsilon_{sr}\end{aligned}$$

where $T_3 = 1_{\{treatment=3\}}$, $T_6 = 1_{\{treatment=6\}}$, $T_9 = 1_{\{treatment=9\}}$

y is subject *s*'s estimate for the expected price of the security in round *r*. Subscripts on *T*'s correspond to the signal treatment they are indicator variables for. Here *c*, *e*, and *i* represent conditionally independent signals for Company, Economy, and Industry, respectfully. Subscripts on signals indicate whether they are signal number 1, 2, or 3 of type *c*, *e*, or *i*. Coefficient superscripts represent signal type, while subscripts indicate *(signal number, treatment).*

In the following hypotheses and resulting analysis, coefficient values from the above model are interpreted as subjects' sensitivity to, or assimilation of informational signals. One point that will play a very important role in the subsequent analysis is that extreme expected price estimates (near \$0 or \$100) have a relatively huge effect on coefficient estimation (Figure 1.3) because of the Bayesian log-odds regression structure. In the analysis section we will discuss how, due to this hypersensitivity, including or not including extreme values makes a huge difference in our results.

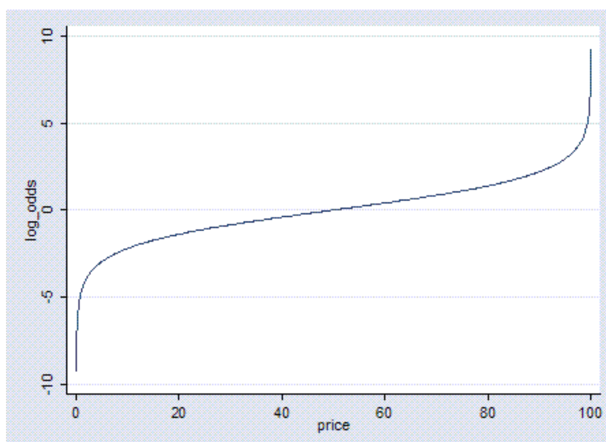

Figure 1.3: extreme sensitivity of log-odds ratio to subjects' expected price estimates

III. Hypotheses

In this section we present the four hypotheses tested in our analysis. In addition, in Appendix II we provide a brief history of the IOL literature, which motivates our choice of these particular hypotheses.

H1 posits that the average response error will increase with the number of signals shown to subjects.

$$\text{H1}: IOL\ (Weak): \frac{\sum |y_t - price_t|}{m_t} < \frac{\sum |y_s - price_s|}{m_s} < \frac{\sum |y_n - price_n|}{m_n}$$

Here t is the observation index for the three signal treatment, s for the six, and n for the nine signal treatment. y is a subject's estimated expected price. price is the computed Bayesian criterion expected price. m_i is the number of observations in treatment *i*.

IOL(weak) predicts first that subjects are not perfect Bayesians, and second that they deviate further from the Bayesian benchmark, on average, the more signals they are given. If subjects were perfectly reasoning all average absolute errors, of course, would be zero. H1 is a necessary but insufficient condition for IOL.[3]

$$\text{H2}: IOL\ (Moderate): \sum \beta_3^c + \sum \beta_3^e + \sum \beta_3^i < \sum \beta_6^c + \sum \beta_6^e + \sum \beta_6^i = \sum \beta_9^c + \sum \beta_9^e + \sum \beta_9^i$$

where subscripts denote signal treatment and superscripts denote signal type:

c = company, e = economy, and i = industry.

H2 is a moderately strong IOL hypothesis in that it predicts that information assimilation will increase between the three and six signal treatment, but then fail to increase from the six to nine signal treatment. This is a fairly strong statement given that subjects will actually need to assimilate relatively more signals in a treatment with more signals in order to do equally well as in a treatment with less signals.[4] That IOL is likely to occur between six and nine signals is an educated guess supported by (Miller, 1956).[5]

---

[3] H1 is the only one of the four hypotheses that is "model-free."

[4] For example, if a subject assimilates six signals in a six signal round her response should have zero decision error. On the other hand, if she assimilates signals one through six in a nine signal round, on average her estimate will be off by $9.5.

[5] In retrospect, we probably should have been wimpier and included a treatment with more signals- perhaps 12 or 15.

$$\text{H3}: IOL\ (Strong): \sum \beta_3^c + \sum \beta_3^e + \sum \beta_3^i < \sum \beta_6^c + \sum \beta_6^e + \sum \beta_6^i > \sum \beta_9^c + \sum \beta_9^e + \sum \beta_9^i$$

H3 only differs from H2 by one inequality. Whereas H2 predicts a plateauing of assimilated information between the six and nine signal treatments, H3 predicts a reversal effect, which is stronger.

$$\text{H4}: Polarization:$$
$$case\ 1: \left(\frac{\beta_3^c}{\sum \beta_3} < \right) \frac{\beta_6^c}{\sum \beta_6} < \frac{\beta_9^c}{\sum \beta_9}$$
$$case\ 2: \left(\frac{\beta_3^i}{\sum \beta_3} > \right) \frac{\beta_6^i}{\sum \beta_6} > \frac{\beta_9^i}{\sum \beta_9}$$

H4 predicts that as information load increases from 6 to 9 signals, subjects will concurrently put relatively more weight on the strongest signals and relatively less weight on weakest signals. For completeness, a similar effect is predicted for increasing information loads from three to six signals (in parentheses), though the primary interest in this study is what occurs in the posited IOL range between 6 and 9 signals.[6]

A subject that is prone to both overload and polarization is vulnerable to potentially severe forms of strategic manipulation, where an increased information load can result in myopic over-weighting of a relatively small number of particular signals.

IV. Results

Our results are reported in two sets. The first set includes all subject data, without accounting for the Bayesian model's great sensitivity to extreme expected price estimates. In this case we show that as the number of information signals increase

[6] H4 is an adaptation of a hypothesis taken from T & B (2004).

subjects stray further from the criterion price, but assimilate more information and show no signs of polarization.[7]

In the second set of results we account for the Bayesian model's extreme sensitivity to extreme expected price estimates by truncating the 4% most extreme estimates. This truncation allows for all subjects' estimates to be weighted virtually equally in our analysis, as they should be, unlike in the non-truncated dataset. In this case we show that as the number of information signals increase subjects stray further from the criterion price, and as they go from six to nine signals they assimilate no more information, and disproportionately over-weigh the strongest signals.

Results for the non-truncated and truncated datasets are reported in Tables 1.3 and 1.4, respectively.

Table 1.3: results of experimental data with extreme responses included in the sample. While H1 cannot be rejected, H2-H4 are.

Results: **Including Extreme Values**	"Accept"/ Reject	Values	Test
H1 (Weak IOL):	(Accept: $3 \rightarrow 6$) Accept: $6 \rightarrow 9$	3:MeanAbs.Error=11.148 6:MeanAbs.Error=12.617 9:MeanAbs.Error=14.699	Regress abs. error On treatment indicators $\beta_3 < \beta_6$: $t = -3.15$ Prob $> \lvert t \rvert = .002$ $\beta_6 < \beta_9$: $t = 4.87$ Prob $> \lvert t \rvert = .000$

[7] When including all values there is one minor adjustment that needs to be made in order for all subjects' responses to have well defined log-odds in the Bayesian model. We alter responses of $0 and $100 to $1 and $99 respectively. These changes occur for roughly 1% of all observations.

Table 1.3:(continued)

H2 (Moderate IOL):	(Accept: $3 \rightarrow 6$) Reject: $6 \rightarrow 9$	$3: \beta_3^c + \beta_3^e + \beta_3^i = 1.438$ $6: \sum \beta_6^c + \sum \beta_6^e + \sum \beta_6^i = 1.740$ $9: \sum \beta_9^c + \sum \beta_9^e + \sum \beta_9^i = 2.410$	$\beta_3^c + \beta_3^e + \beta_3^i <$ $\sum \beta_6^c + \sum \beta_6^e + \sum \beta_6^i$ $F(1,4203) = 19.77$ $\text{Prob} > F = .000$ $\sum \beta_6^c + \sum \beta_6^e + \sum \beta_6^i$ $= \sum \beta_9^c + \sum \beta_9^e + \sum \beta_9^i$ $F(1,4203) = 0.65$ $\text{Prob} > F = .4215$
H3 (Strong IOL):	(Accept: $3 \rightarrow 6$) Reject: $6 \rightarrow 9$	$3: \beta_3^c + \beta_3^e + \beta_3^i = 1.438$ $6: \sum \beta_6^c + \sum \beta_6^e + \sum \beta_6^i = 1.740$ $9: \sum \beta_9^c + \sum \beta_9^e + \sum \beta_9^i = 2.410$	Follows from H2 test results
H4 Polarization:	(Reject(C): $\rightarrow 6$) Reject(C): $6 \rightarrow 9$ (Reject(I): $3 \rightarrow 6$) Reject(I): $6 \rightarrow 9$	Company: $\frac{\beta_3^c}{\sum \beta_3} = .386$; $\frac{\beta_6^c}{\sum \beta_6} = .397$; $\frac{\beta_9^c}{\sum \beta_9} = .382$ Industry: $\frac{\beta_3^i}{\sum \beta_3} = .306$; $\frac{\beta_6^i}{\sum \beta_6} = .320$; $\frac{\beta_9^i}{\sum \beta_9} = .298$	**Company :** $\frac{\beta_3^c}{\sum \beta_3} < \frac{\beta_6^c}{\sum \beta_6}$ $F(1,4203) = 0.16$ $\text{Prob} > F = .689$ $\frac{\beta_6^c}{\sum \beta_6} < \frac{\beta_9^i}{\sum \beta_9}$ $F(1,4203) = 0.25$ $\text{Prob} > F = .619$ **Industry :** $\frac{\beta_3^i}{\sum \beta_3} > \frac{\beta_6^i}{\sum \beta_6}$ $F(1,4203) = 0.41$ $\text{Prob} > F = .524$ $\frac{\beta_6^c}{\sum \beta_6} > \frac{\beta_9^i}{\sum \beta_9}$ $F(1,4203) = 0.64$ $\text{Prob} > F = .424$

In Table 1.4 we report results for the dataset after subjects' 4% most extreme expected price estimates are excluded. Because the Bayesian regression model is highly sensitive to extreme responses, coefficient estimates will likely be affected much more by their inclusion as observations, than by the selection effect that results as a result of their exclusion. By excluding extreme observations the remaining observations are weighed roughly equally, rather than being dominated by extreme estimates. In the histogram of subjects' estimates presented in Figure 1.4, as well as the plot in Figure 1.5, all estimates contained between the two red lines are retained for analysis.

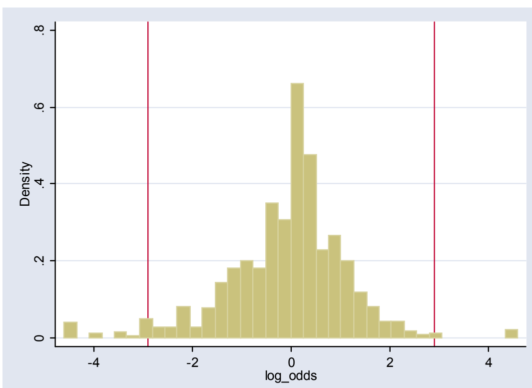

Figure 1.4: histogram with truncated 4% extreme observations

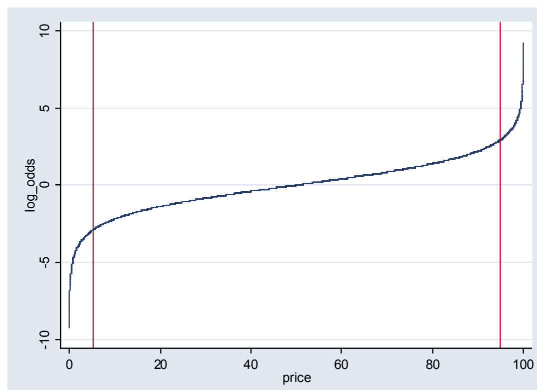

Figure 1.5: price cut-offs when 4% of extreme estimates are truncated

Table 1.4: results of experimental data with the 4% most extreme responses truncated from the sample. While H3 is rejected, H1, H2, and an important case of H4 cannot be rejected.

Results: **Excluding Extreme Values**	"Accept"/ Reject	Values	Test
H1 (Weak IOL):	(Accept: $3 \rightarrow 6$) Accept: $6 \rightarrow 9$	3: Mean Abs.Error = 10.634 6: Mean Abs.Error = 12.575 9: Mean Abs.Error = 15.028	Regress abs. error On treatment indicators $\beta_3 < \beta_6$: $t = -4.20$ Prob $> \lvert t \rvert = .000$ $\beta_6 < \beta_9$: $t = -5.42$ Prob $> \lvert t \rvert = .000$
H2 (Moderate IOL):	(Accept: $3 \rightarrow 6$) Accept: $6 \rightarrow 9$	$3: \beta_3^c + \beta_3^e + \beta_3^i = 1.159$ $6: \sum \beta_6^c + \sum \beta_6^e + \sum \beta_6^i = 1.459$ $9: \sum \beta_9^c + \sum \beta_9^e + \sum \beta_9^i = 1.536$	$\beta_3^c + \beta_3^e + \beta_3^i <$ $\sum \beta_6^c + \sum \beta_6^e + \sum \beta_6^i$ F(1,4203) = 26.07 Prob $> F$ =.000 $\sum \beta_6^c + \sum \beta_6^e + \sum \beta_6^i$ $= \sum \beta_9^c + \sum \beta_9^e + \sum \beta_9^i$ F(1,4203) = 43.59 Prob $> F$ =.000
H3 (Strong IOL):	Accept: $3 \rightarrow 6$) Reject: $6 \rightarrow 9$	$3: \beta_3^c + \beta_3^e + \beta_3^i = 1.159$ $6: \sum \beta_6^c + \sum \beta_6^e + \sum \beta_6^i = 1.459$ $9: \sum \beta_9^c + \sum \beta_9^e + \sum \beta_9^i = 1.536$	Follows from H2 test results

Table 1.4: (Continued)

H4 Polarization:	Reject(C): 3 → 6) Accept(C): 6 → 9 (Reject(I): 3 → 6) Reject(I): 6 → 9	Company $\frac{\beta_3^c}{\sum \beta_3} = .416$ $\frac{\beta_6^c}{\sum \beta_6} = .404$ $\frac{\beta_9^c}{\sum \beta_9} = .481$	Industry $\frac{\beta_3^i}{\sum \beta_3} = .274$ $\frac{\beta_6^i}{\sum \beta_6} = .292$ $\frac{\beta_9^i}{\sum \beta_9} = .243$	**Company :** $\frac{\beta_3^c}{\sum \beta_3} < \frac{\beta_6^c}{\sum \beta_6}$ F(1,4203) = 0.17 Prob > F =.6785 $\frac{\beta_6^c}{\sum \beta_6} < \frac{\beta_9^i}{\sum \beta_9}$ F(1,4203) = 3.93 Prob > F =.0475 **Industry :** $\frac{\beta_3^i}{\sum \beta_3} > \frac{\beta_6^i}{\sum \beta_6}$ F(1,4203) = 0.59 Prob > F =.4422 $\frac{\beta_6^c}{\sum \beta_6} > \frac{\beta_9^i}{\sum \beta_9}$ F(1,4203) = 2.41 Prob > F =.1206

The combination of subjects' plateauing level of assimilated information and a shifting of disproportionate weight toward the strongest signals suggests that increases in information load allow information senders a clear opportunity for the strategic manipulation of receivers. [8]

V. Conclusion

[8] One can imagine a car company that dominates the competition in a critical characteristic, such as resale value, "overloading" people so that they place a disproportionately large amount of weight on this characteristic.

In an age of rapidly increasing information availability the non-trivial tension between abundant information and finite human cognitive capacity (Miller, 1956) becomes ever more pronounced. We define information overload as the act of assimilating relatively less information in a task that contains strictly more relevant information, but is otherwise identical.

While it is difficult to test for information overload (IOL) in field studies, laboratory experiments offer an effective alternative environment where extraneous variables can be controlled, decision accuracy can be measured objectively, and signals can be made both relevant and independent from one another so that subjects' assimilation of individual information signals can be measured.

In our experimental design subjects are asked to estimate expected security prices after being given priors along with variable information loads in the form of (binary) probabilistic signals. In each round of a within subjects design subjects make an estimate after being provided with either three, six, or nine probabilistic signals- each of varying signal strength weak, moderate, or strong.

We test for four overload hypotheses. They posit that as the number of information signals increases subjects' estimates will stray further from the criterion expected price (weak IOL), that as the number of signals increases from six to nine subjects will assimilate the same (moderate IOL) to strictly less (strong IOL) information, and that as the number of signals increases subjects will disproportionately over-weigh high strength signals and disproportionately under-weigh low strength signals (polarization).

Our results are reported in two sets. The first set includes all subject data, without accounting for the Bayesian model's extreme sensitivity to extreme expected price estimates. In this case we show that as the number of information signals increase subjects stray further from the criterion price, but assimilate more information and show no signs of polarization.

In the second set of results we account for the Bayesian model's extreme sensitivity to extreme expected price estimates by truncating the 4% most extreme estimates. This truncation allows for all subjects' estimates to be weighted virtually equally in our analysis, as they should be, unlike in the non-truncated dataset. In this case we show that as the number of information signals increase subjects stray further from the criterion price, and as they go from six to nine signals they assimilate no more information, and disproportionately over-weigh the strongest signals.

The combination of subjects' plateauing level of assimilated information and a shifting of disproportionate weight toward the strongest signals suggests that increases in information load allow information senders a clear opportunity for the strategic manipulation of receivers.

In retrospect, given the observed plateau effect in information assimilation between the six and nine signal treatments, it would have been interesting to have included a twelve-plus signal treatment in the design; perhaps we would have observed a reversal.

By choosing a Bayesian updating task we provided ourselves with a decision theory backed decision model which allowed us the ability to interpret regression coefficient estimates as subjects' assimilation of information signals. In our analysis we were forced to make a difficult decision and omit a small fraction of our data, due to the Bayesian model's great sensitivity to extreme responses. In future experiments such difficult decisions can be avoided by the creation of designs with exclusively model-free hypothesis testing, like our H1, for example. Another option is to use a revealed-preference approach where subjects choose between different bundles of information with varying information loads. [9] On the other hand, neither one of these types of alternative designs allows for the type of rich analysis of information assimilation that we were able to conduct here.

Appendix I

Instructions

Motivation

Suppose you are a stock analyst and you are researching a new stock. Your task is to provide a price estimate for this stock before it becomes publicly traded. You know that in one year the stock will be worth either \$100 or \$0: \$100 if the company does well and \$0 if it does poorly. You do not know, however, whether the company will do well or poorly, and without knowing anything specific about the company you consider both outcomes equally likely.

Then you acquire information about the Company. This information consists of a set of indicators (pieces of information). Each of these indicators is more likely to be strong when the company will do well, and more likely to be weak when the company will do poorly. You use these indicators to update your probability belief that the company will do well. Your price estimate is equal to the expected value of the stock based on the updated probability that the stock will be worth \$100. For example, if after analyzing the given indicators you believe the company will do well with probability

[9] As suggested by Vince Crawford.

0.75, your price estimate for the stock will be $75. If you believe that the company will do well with probability 0.25, your price estimate for the stock will be $25.

Procedure

We will now go over detailed instructions after which you will have an opportunity to ask questions. Then you will have 9 practice rounds. Practice rounds will not count towards your payoff. The main purpose of the practice rounds is for you to familiarize yourself with the decision task and see how your stock price estimates differ from the true prices. The true price will change from round to round. The 9 practice rounds will be followed by 24 scored rounds in which you will be able to earn credit. The scored rounds also differ from the practice rounds in that you will not be given true prices at the end of the scored rounds. That is, true prices are only reported after each practice round. Each of you will be assigned to either Group A or Group B for the entire exercise. Please note your group and use information for that group to calculate your price estimates.

Each of the 9 practice rounds will consist of the following stages:

· Information related to the company whose stock you are going to price will appear on the screen. For those of you in Group A, the information will appear on the left side of the screen. For those of you in Group B, the information will appear on the right side of the screen. The number of the round will appear on top of the screen. The information will stay on the screen for 50 seconds. After that, the true price for that round will be displayed on the screen for 5 seconds. Then the experiment moves to the next round.
· During the 50 seconds, you need to come up with your estimate for the stock price which should be as close as possible to the true price based on the information available to you. Please record your estimate on your answer sheet. Then the true price for that round will be displayed for 5 seconds. Please record your estimate before you see the true price. Remember, you are not earning credit for the practice rounds, and therefore you do

not gain anything from waiting to see the true price and recording it instead of your estimate. Once a practice round ends, please stop working on that round and move on to the next one.

Each of the 24 credit rounds will be similar to the practice rounds, except that you will not receive feedback on what the true price is.

We provide you with two pages of scratch paper in the answer sheet packet. Please use these pages for your calculations and do not use any other scratch paper. You are not allowed to use any notes, but you are allowed to use a calculator.

After the experiment, please record your price estimates from the 24 credit rounds on the scantron sheet provided with your answer sheet packet. Record your answer for the first round, rounded to the nearest integer, in lines 1 and 2 of the scantron sheet. Please mark the first of the two digits in line 1, and the second of the two digits in line 2. Use the "10-bubble" as zero. If you want to enter a single-digit number, for example 7, enter it as 07. If you want to enter 0, enter it as 00. If you want to enter 100, leave the first line blank, and enter 0 in the second line. For the second round, mark your estimate in the same way in lines 3 and 4, and so on for the following rounds. In order for you to be eligible for extra credit, you must provide a price estimate for each round on the scantron sheet.

Note: no scantron answers = no extra credit.
At the end of the experiment, you will be asked to fill out a short survey.

Indicators

In each round, you will receive information consisting of the combination of values of three types of indicators: Company Indicators, Economy Indicators, and Industry Indicators. Each indicator can take only one of two values: "Strong" or "Weak."

Here is a table that lists the three types of indicators, together with possible values they may take:

	Strong Value	Weak Value
Company Signals		
Company_1	Strong	Weak
Company_2	Strong	Weak
Company_3	Strong	Weak
Economy Signals		
Economy_1	Strong	Weak
Economy_2	Strong	Weak
Economy_3	Strong	Weak
Industry Signals		
Industry_1	Strong	Weak
Industry_2	Strong	Weak
Industry_3	Strong	Weak

We generated the indicator values in each round using the following steps:

Step 1: In each round, it is first randomly determined whether the company will do well or do poorly. The probability of each outcome is 0.5. You, however, do not know which of the two is the outcome in any given round.

Step 2: Conditional on the outcome of the first step, the value of each indicator is generated randomly to be either "Strong" or "Weak" with certain probabilities. This is done independently for each individual indicator. These probabilities are given in the following table:

	Company Indicators		*Economy Indicators*		*Industry Indicators*	
	Strong	**Weak**	**Strong**	**Weak**	**Strong**	**Weak**
Do Well	0.65	0.35	0.60	0.40	0.55	0.45
Do Poorly	0.35	0.65	0.40	0.60	0.45	0.55

For example, conditional on the Company doing well, each individual Company indicator will be "Strong" with probability 0.65 and "Weak" with probability 0.35.

Conditional on the Company doing poorly, each individual Company indicator will be "Strong" with probability 0.35 and "Weak" with probability 0.65.

That is, in each round, regardless of whether the company will do well or do poorly, each indicator may me Strong or Weak. Therefore none of the indicators is a perfect signal of whether the company will do well or do poorly. But each indicator is statistically related to how well the company will do. That is, any given indicator is more likely to be "Strong" when the Company will do well, and is more likely to be "Weak" when the company will do poorly.

Your Payoff

In this in-class exercise you can earn up to 2% extra credit toward the overall course score. You will be awarded points following the dartboard principle for your estimates in each round. If your estimate is within 1 unit of the stock price, you will earn 20 points, if it is within 2 units of the stock price, you will earn 19 points, and so on, up until if your answer is within 20 units of the stock price, you will earn 1 point. That is, the closer your estimate is to the true price, the higher is your reward. If your estimate is more than 20 units away from the stock price, you will not earn any points in that round.

Your extra credit in terms of the percentage gain for your overall class score is then equal to the fraction of 2% given by the ratio of the points you earn to total possible point earnings of 480. That is, your extra credit is

$$Extra\ Credit = 2\% * \left(\frac{Aggregate\ Individual\ Score\ for\ 24\ rounds}{480} \right) .$$

You are free to form your estimates in any way you desire so as to maximize your credit. But please note that in order to earn your credit, you need to provide a price estimate on the scantron sheet in every credit round.

This is an independent exercise, so please do not communicate or share information with other participants during the experiment. Because you will be rewarded based on your individual participation and performance, communicating or working with others is strictly prohibited and will result in you being excused from the exercise, and thereby exempt from the extra credit opportunity.

If you do not wish to participate in this experiment, you are free to leave the room now. In case you decided to stay, we would a

Practice Rounds

We will now perform 9 practice rounds in order for you to become better accustomed to making price estimates under limited time of 50 seconds and varying amounts of information. Remember, these 9 practice rounds do not count for your payoffs, and, unlike for the credit rounds, you will be given the true price feedback after each practice round. Therefore we encourage you to use these 9 practice rounds to see how well your estimated prices approximate corresponding true prices.

Credit Rounds

We are now going to start the 24 credit rounds. In each round, the indicator information will be displayed for 50 seconds. Remember, you will not receive any feedback on the true price. Once a round is over, please stop working on it and move to the next round. Good luck and make sure to enter your estimates on the scantron sheet.

Appendix II

In 1956 Miller ran an exhaustive list of experiments tracing out human channel capacities at 7 plus or minus two. Channel capacity is the asymptotic "upper limit on the extent to which the observer can match his responses to the stimuli (he is given)." In these experiments subjects were given variable amounts of "input information" and the amount of "transmitted information" was recorded. IOL builds on Miller's work by recognizing channel capacities and asking the question: What happens when people face sets of information that exceed their channel capacities? Three possibilities are:

1) no change due to perfect reasoning and/or free disposal
2) deterioration in decision accuracy due to assimilating more information than is optimal (naive and inexperienced)
3) use of simplifying rules of thumb to reduce information assimilated (naive and experienced)[10]

In 1967 Schroder, Driver, and Streufert introduced the theory of an inverted U-shape relationship between information load and information assimilated. The first IOL studies were then conducted by Jacoby et al. in 1974. These preliminary experiments found results of decreasing "decision accuracy" with greater information loads. Malhotra (1984), among others, pointed out several concerns with the Jacoby et al. design. Most notably Jacoby et al. used a subjective measure of decision accuracy based on comparing subjects' responses to their own reported "ideal responses." Also, Jacoby et al. did not draw a clear distinction between the amount of information contained in the number of choice alternatives available and the amount of information contained in the number of attributes defining each of those alternatives. Later studies showed that number of alternatives and number of alternatives' attributes should be treated independently (Malhotra, 1980; Leckenby, 2001).

Malhotra (1980) followed Jacoby et al. (1974, 1977) by using the benchmark of subjects' individually selected ideal choices. He computed decision error by summing euclidean distances between ideal and selected attributes across all attributes. This measure was highly problematic, as in order for it to be an accurate measure all attributes needed to be independent, which they clearly were not in Malhotra's experiments.

Building on Malhotra's work, Tuttle & Burton (henceforth T&B; 1999, 2004) used statistical tests to check for absence of cross-correlations between signals in their designs. Both papers belonged to a class of literature studying hypothetical financial decisions made by experimental subjects (Chewning & Harrell, 1990; Stocks & Harrell, 1995; Stocks & Tuttle, 1998). These studies were especially important in that they

[10] Clearly, two and three are not mutually exclusive.

provided a means of measuring the amount of information subjects assimilate. This was done by manipulating an orthogonal informational signal design, then running statistical regressions with subjects' responses as the dependent variable and observed informational signals as the independent variables.

Although these studies have made significant contributions to the IOL literature, collectively they beg for a more simplified design offering an objective measure of decision accuracy. Chewning & Harrell (1990) classified distance between individual response and group average response as error, and assessed consistency of responses. They provided experimental materials to students in a non-timed, take-home format, which enabled an unmanageable list of factors to affect subject performance. T&B (1999) assumed a linear design to information updating in a stock pricing task though it is not clear why subjects would use this type of model (check). There were 2 different signal loads for subjects, again, with no limit to decision making time. T&B (2004) added a signal load treatment to the (1999) design and ran the experiment in a double auction market. In both T&B experiments memory was an important determinant of performance as names of signals, their types, and their value designations needed to be recalled by subjects. Signal usage was inferred by counting the number of significant signal coefficients when regressing subjects responses on informational signals.

Our experimental design draws from economic decision theory via the standard model of Bayesian belief updating on informative signals. Previous experimental literature shows that subjects tend systematically towards conservatism when updating across multiple informative signals (Camerer, 1993). Conservatism is a general under-adjustment in posteriors generated from a prior and informative signals. For this reason informational signal coefficients are not expected to be individually significant and will not be focussed on. Instead, we study relative changes in coefficient weights as indicative of subjects' relative assimilation of (or sensitivity towards) signals in different signal load conditions. This paper offers the first experimental study of a Bayesian updating task with varying informative signal loads.

The current design also makes strides in isolating the pure variable effect of information load as the independent variable affecting subjects' responses. Relative to

T&B, this design's naming of signals, signal types, and signal values is made uniform and as intuitive as possible in order to minimize subjects' confusion due to extraneous variables such as requirements on memory.

References

Agnew, J., Szykman, L. (2004). Asset Allocation and information overload: The Influence of Information Display, Asset Choice and Investor Experience. Center for Retirement Research at Boston College.

Busemeyer, J. R. (1993) Violations of the Speed-Accuracy Tradeoff Relation: Decreases in Decision Accuracy with Increases in Decision Time. Taken from Time Pressure and Stress in Human Judgment and Decision Making. (Eds. Svenson,O., Maule, A. J.) Plenum Press, New York, 1993.

Chewning, E., Harrell, A. (1990). The Effect of Information Load on Decision Makers' Cue Utilization Levels and decision Quality in a Financial Distress Decision Task. Accounting, Organizations and Society, Vol. 15, No. 6.

Conlisk, J. (1996) Why Bounded Rationality? Journal of Economic Literature, Vol. 34, No. 2 (Jun., 1996)

Cowan, N. (2001): "The Magical Number in Short-term Memory: A Reconstruction of Mental Storage Capacity," Behavioral and Brain Sciences 24 (1).

De Bondt, W. F. M., & Thaler, R. H. (2002). Do analysts overreact? In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 678-685) Cambridge: Cambridge University Press.

Epley, N. & Gilovich, T. (2002). Putting adjustment back in the anchoring and adjustment heuristic. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 139-149) Cambridge: Cambridge University Press.

Eppler, M., Mengis, J. The Concept of information overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines. The Information Society, 20: 325-344, 2004.

Gabaix, X., & Laibson, D. (2000). Bounded Rationality and Directed Cognition. Forthcoming.

Gabaix, X., & Laibson, D. (2004) Competition and Consumer Confusion. Draft version

Gigerenzer, G. (2001). The Adaptive Toolbox. In Bounded Rationality. The MIT Press.

Greene, W.H., (1997) Econometric Analysis. Prentice-Hall Inc.

Mattsson, Weibull (2002). Probabilistic choice and procedurally bounded rationality. Games and Economic Behavior. pgs. 61-78.

Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. IEEE Transactions on Systems, Man, and Cybernetics, 17, 753-770.

Jacoby, J. (1977). Information Load and Decision Quality: Some Contested Issues. Journal of Marketing Research. Vol. 10.

Jacoby, J. (1984'). Perspectives on information overload. Journal of Consumer Research. Vol 10.

Kahneman, D. & Fredrick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 49-81) Cambridge: Cambridge University Press.

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

Leckenby, J., (2001). ADV 392 Doctoral Pro-Seminar http://www. ciadvertising.org/student_account/fall_01/adv392/bklee/paper2/intro.htm

Malhotra, N. (1982). Information Load and Consumer Decision Making. Journal of Consumer Research. Vol. 3.

Malhotra, N. (1984). Reflections on information overload Paradigm in Consumer Decision Making. Journal of Consumer Research. Vol. 10.

Manski, C. Measuring Expectations. Econometrica. Evanston: Sep 2004. Vol. 72, Iss. 5; p. 1329 (48 pages)

Miller, J. (1956) The Magical Number Seven Plus or Minus Two: Some Limits on our Capacity for Processing Information. Psychological Review 63: 81-97.

Payne, J. W., Bettman, J. R., Johnson, E. J. (1993). The adaptive decision maker. Cambridge University Press 1993.

Payne, J., Bettman, J. (2001). Preferential Choice and Adaptive Strategy Use. In

Bounded Rationality. The MIT Press.

Plott, C., Sunder, S., (1988). Rational Expectations and the Aggregation of Diverse Information in Laboratory Security Markets. Econometrica, Vol. 56, No.5.

Gilovich, D. Griffin, and D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 534-547) Cambridge: Cambridge University Press.

Selten, R. (2001) What is Bounded Rationality? In Bounded Rationality. The MIT Press.

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), Heuristics and biases: The psychology of intuitive judgment (pp. 379-396) Cambridge: Cambridge University Press.

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

Stigler, J. (1961) The Economics of Information. In the "Journal of Political Economy" 69, no. 3, june.

Todd, P. Fast and Frugal Heuristics for Environmentally Bounded Minds. In Bounded Rationality. The MIT Press.

Tuttle, B., Burton, G., (1999). The Effects of a Modest Incentive on information overload in an Investment Analysis Task. Accounting, Organizations, and Society. Vol. 24.

This chapter is currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this material. This chapter was co-authored with Peter Katuscak.

Search with Multiple Attributes: Theory and Empirics

Adam Sanjurjo[1]
University of California, San Diego
20 November 2008

Abstract: Optimal search policies have been fully characterized, and tested in the laboratory, for a wide variety of single-attribute search problems. Few authors, however, have addressed optimality in multiple-attribute search environments, though a number of experiments have been conducted. Gabaix and Laibson (2006) conduct a multiple-attribute experiment, performing an analysis on the data that is largely aggregative, and sharply focused on comparing several behavioral models. I build on their analysis in three ways. First, I provide a partial characterization of optimality in the form of several necessary conditions for optimal search. To my knowledge, these are the first systematic theoretical results for multiple-alternative, multiple-attribute search with full recall. Second, I show that experimental subjects violate these conditions frequently; 97% of all search problems, and 62% of all search actions within problems, violate at least one of the conditions. Third, I analyze Gabaix and Laibson's dataset on an individual level, which yields a classification of four exclusive search "types" that together describe 98% of all search behavior. I find that behavior varies systematically across these types, both in terms of violation rates of necessary conditions, as well as several other descriptors of search and performance. I close by exploring possible explanations for observed

[1] Department of Economics, University of California, San Diego. 9500 Gilman Drive, La Jolla, CA 92093-0508 (email: asanjurjo@ucsd.edu). Many thanks to Nageeb Ali, Karim Chalak, John Conlisk, Richard Carson, David Eil, Patrick Fitzsimmons, Mark Machina, David Miller, Joshua Miller, Justin Rao, David Schkade, Ricardo Serrano-Padial, Joel Sobel, and Yixiao Sun for insightful comments, to Xavier Gabaix and David Laibson for providing their data, and to Vince Crawford for invaluable guidance throughout the entire process of creation.

behaviors, including deviations from optimality, based on a simple model of working memory load that I develop in an accompanying paper.

Introduction

Optimal search policies have been fully characterized for a wide variety of search problems 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 these analyses yield substantial insights, many important applications have alternatives whose values are determined by multiple attributes: we consider more than wage when choosing a job, and more than price when purchasing a home.

Gabaix and Laibson (2006; henceforth "GL") study search with multiple attributes (ten) and alternatives (eight) experimentally, with full recall and no order restrictions. Each of their subjects faces a series of search problems, either (in separate treatments, described below) with an explicit time limit per problem or with a fixed budget of time in which to do as many problems as desired. GL's analysis is of particular interest because the richness of their search environment approximates human cognition in less structured settings more closely than most other models, and because their experimental interface bears a close family resemblance to the kind of information displays commonly used in internet commerce.[1] [2]

---

[1] Typing "compare homes" or "compare cars," etc. into Google will yield several hits for websites that compile information so that several goods of the requested type can be compared across multiple attributes.

[2] There are other multiple attribute search experiments I could have chosen to study here, such as those in Payne, Bettman, and Johnson (1988), but they would not have allowed as rich an analysis of optimality.

Although optimal search would be a natural benchmark with which to compare subjects' behavior, GL find that the high dimensionality of their problem make characterizing optimal search analytically and numerically intractable (p. 1066). Instead they focus on comparing their "Directed Cognition" model of search, which is myopic in that it ignores option value, but otherwise fully rational, with "naïve" heuristics taken from the psychology literature (Tversky, 1972). Further, although GL's data include detailed observations of behavior at the individual level, search problem by search problem, most of their data analysis is conducted at the aggregate level, pooling both subjects and problems.

This paper builds on GL's analysis in three ways, with the goal of advancing our understanding of observed multiple-attribute search and how it relates to optimal search with perfect recall. First, I conduct a detailed analysis of GL's search data on the individual level. Second, although a full characterization of optimal search in GL's environment does indeed appear to be intractable, I give a partial characterization in the form of several necessary (but not sufficient) conditions for optimality.[3] Third, I compare subjects' behavior to this partial rational benchmark.

My first necessary condition allows me to identify instances of conditional over-search and under-search within alternatives, which include violations that occur both when the searcher switches alternatives, and when the searcher continues to search in the same alternative. Behaviorally, these two types of violations are quite distinct, so further insight is gained by studying each of them separately. In GL's design the attributes in

[3] Collectively, my necessary conditions assume risk neutral payoff maximization, perfect reasoning (including memory), and costly search.

each alternative have information values that vary in a simple way, which is announced to the subjects. My second necessary condition for optimality identifies cases where subjects search an attribute that is strictly dominated in the value of information sense. The remaining three conditions are theoretically trivial under my assumptions. Two of them identify search decisions that are strictly uninformative, but costly. The last simply identifies which alternative must be chosen at the completion of search, which determines subjects' payoffs. All proofs are contained in the appendix.

To my knowledge, these necessary conditions are the first systematic theoretical results on multiple-alternative, multiple-attribute search with full recall.[4] They facilitate solving numerically for fully optimal search policies in versions of GL's search problem with fewer alternatives and attributes, which will make it possible to conduct experiments that compare observed behavior with a complete optimal-search benchmark. When used on GL's individual-level data, they allow useful inferences about subjects' main systematic deviations from optimal search.

For example, a robust finding in single attribute experiments, as well as Bearden et. al.'s (2007) two-attribute experiments, is that subjects systematically search too few alternatives (Camerer 1995). The multiple-attribute model poses two related questions, which are closely linked to the extent that search costs limit the total number of look-ups: whether subjects tend to search too many or two few alternatives, and too few or too many attributes within a given alternative. My first necessary condition for optimality allows me to partially address these questions, by identifying instances where the search

[4] The only other theoretical work on multiple attribute search that I am aware of is by Bearden et al. (2005), who numerically solve for the optimal policy with two attributes and ten alternatives, ruling out recall and imposing strong restrictions on the order of search.

of a particular attribute simultaneously indicates conditional over-search of the alternative that attribute is in, and conditional under-search in another alternative. 50% of subjects' search actions violate this condition.

The remainder of the paper is organized as follows. In Section I I explain GL'S experimental design. In Section II I describe the basic properties of GL's subjects' search behavior in a completely model-free way, finding that even an analysis at this basic level eliminates entire classes of candidate theories of search behavior. Section III begins my analysis of GL's data on the individual level, treating each subject and search problem as a separate observation, and in as model-free a way as possible, using the results from Section II to classify 98% of behavior into one of four mutually exclusive search "types."[5] Section IV compares subjects' behavior to my five necessary conditions for optimality, first in the pooled sample of all search problems, then conditional on the four search types classified in Section III, along with corresponding summary statistics reflecting other basic features of search behavior and performance for each type. Even my partial characterization of optimality allows me to identify a large number of violations of optimal search; 97% of all problems, and 62% of all search actions within problems violate at least one of my five necessary conditions. Also, violation rates of necessary conditions, and search behavior in general, vary systematically across search types. In Section V I explore possible explanations for subjects' deviations from optimality, based on the model of working memory load in Sanjurjo 2008b (Newell and Simon 1972, Johnson, Bettman, and Payne 1993, Crawford 2008), which successfully

[5] I study search on the individual problem level because search patterns vary not only across subjects but within subjects as well. Sanjurjo 2008c studies whether time pressure may be an important cause of this within subject variation in search patterns.

predicts multiple seemingly unrelated search behaviors, including violations of optimality, observed in the GL data. Section VI concludes the paper, and is followed by an appendix containing proofs for those of section IV's necessary conditions whose proofs are not immediate.

I. GL's Experimental Design

GL's subjects face a series of problems in which they choose one of eight alternatives, under time pressure. The value of each alternative is equal to the sum of its ten individual attribute values, nine of which are initially unknown to the subject (imagine comparing eight different potential employers using information collected on attributes such as wage, benefits, commuting distance, etc.). Attributes are mean zero, normally distributed, independent, and linearly decaying in variance from attribute one (σ^2) to ten ($.1\sigma^2$).[6] By imposing time pressure GL create a shadow cost of time in both of their experimental treatments. In the "Endogenous" treatment subjects are given 25 minutes to complete as many different problems as they choose to, given a 20 second buffer screen between problems. In the "Exogenous" treatment subjects are allocated between 10 and 49 seconds, drawn from a uniform distribution, to complete each of 12 different problems.[7] The design is within subject, so each subject completes both of the treatments, with half of the subjects starting in the Exogenous, and the other half in the Endogenous, in order to control for order effects.

[6] Sigma varies across problems. The effects of this variation on search behavior is explored in Sanjurjo 2008c.
[7] In all but the econometric analysis, I pool data from the first Exogenous and first Endogenous treatments, excluding the second treatments in order to mitigate learning effects, and pooling the first treatments because their results are relatively similar (except for one special case mentioned in Section V).

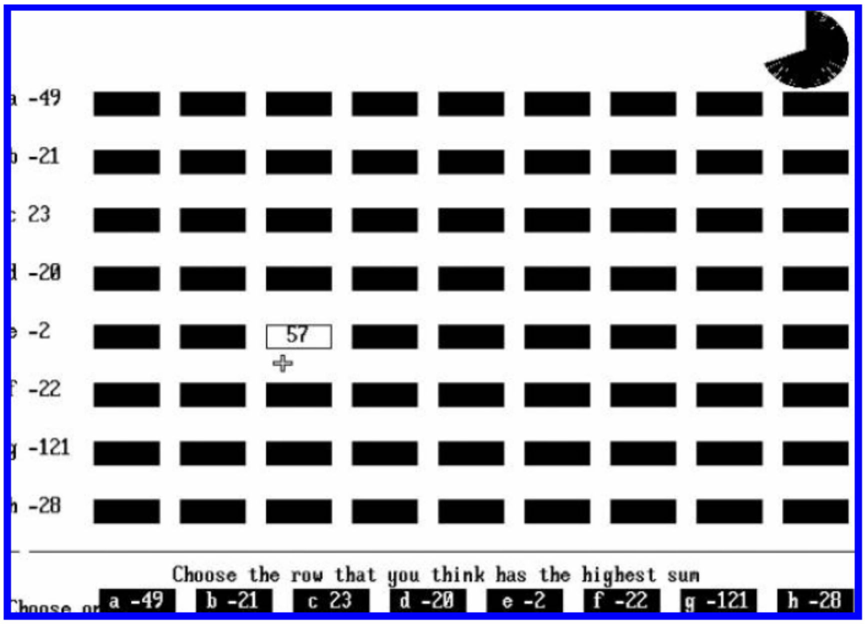

Figure 2.1: Search Problem From GL: Alternatives 1-8 run from top to bottom. Attributes 1-10 run from left to right. Attribute one values are completely observable for the duration of the problem. Of the remaining 72 attributes, only one can be observed at a time. In this particular moment of search the subject is observing attribute 4 of alternative 5. Choice boxes corresponding to each of the alternatives can be seen lining the bottom of the display. The amount of time remaining for each problem is continuously represented by the decaying disc in the top right corner of the display.

The experiment's 390 Harvard undergraduate subjects are given complete information of the distribution of all attributes, and the amount of time allocated for each problem. Subjects are paid the sum of attributes one through ten in the alternative they choose, regardless of how many of these attributes they searched. The average earnings per problem are $0.52 and subjects average 40.8 problems completed.

Building on work by Payne, Johnson, and Bettman (1993), GL use the MouseLab experimental interface in order to track the complete order and duration of all information acquisitions made by subjects. MouseLab is essentially a mechanical analog to eye-tracking, in which attribute values are contained within "boxes" on the computer screen that can be opened, one at a time, by clicking the left button of the computer mouse (see

Figure 2.1). Each box must be closed, with a right-click of the mouse, before the next attribute box can then be opened. In GL's design attribute one values for each alternative are fully observable for the entirety of the search problem, whereas attributes two through ten are covered (unless one is opened). At the bottom of the display are choice boxes-one for each alternative. In order to choose an alternative after search is completed, one must left-click once on the corresponding choice box to choose it, and then again to confirm the choice.

II. Basic Properties of Observed Search Behavior

In this section I perform a model-free analysis that reports the basic properties of search behavior in the average problem. Namely, most search transitions occur either within an alternative or within an attribute, and are adjacent. This simple finding is enough to eliminate large classes of candidate search models (of which the GL environment allows an enormous amount).[8]

I describe search in the average problem, first in terms of breadth, depth, and speed, then in terms of the types of transitions that occur from one attribute to the next. Because most search transitions are found to be adjacent and within an alternative or within an attribute, a simple, tightly fitting, classification of search "types" is made possible in Section III.

On average, subjects perform one attribute look-up per second, and 30.8 look-ups per search problem, which on average span 4.8 different rows and 7.8 different columns.

[8] There are 72^n possible search paths for every n boxes searched.

Tables 2.1 and 2.2 together summarize the characteristics of search transitions from one attribute to the next. Table 1 shows, for the "representative search problem," that subjects transition within alternative (horizontal) 13 times as frequently as they transition within attribute (vertical). Of the relatively small percentage of search transitions that occur neither within alternative nor within attribute, the majority are identical in one important respect- they are transitions to the first attribute of a transitioned to alternative as part of a "typewriter" acquisition pattern. This "typewriter" search pattern performs several sequential look ups in one alternative going left-to-right, then switches to the left-most attribute of the next searched alternative, followed by another run of left-to-right look-up transitions, and so on.

Table 2.1: Transitions: Horizontal, Vertical, and Other- in the Representative Problem

Transition Type	Mean
Within Alternative	24.6
Within Attribute	1.9
Other(typewriter)	3.4 (2.9)

Table 2 shows the frequency of search transitions to each of the four cardinal directions, along with the corresponding extent to which these transitions are spatially adjacent. The majority of all transitions occur from left-to-right (87%) and are adjacent (96%).

Table 2.2: Transitions: Direction and Degree of Adjacency

Transition Type	Mean
Right (adjacent)	22.9 (22.5)
Left (adjacent)	1.6 (1.4)
Down (adjacent)	1.2 (1.1)
Up (adjacent)	0.5 (0.3)
Stay	0.2

The results of this unstructured analysis are straightforward. In general, search sequences are smooth and systematic. 59 out of every 60 look-up transitions occur either within alternative, within attribute, or as an example of deliberate “typewriter” alternative switching. Thus, subjects are not “jumping” around the information matrix. Instead search tends to be “sweeping” and highly adjacent between transitions.

III. Classification of Heterogeneous Search Types

In this section I classify four simple and exclusive “search types,” that together describe 98% of all search behavior. Because subjects’ search patterns are noisy and highly heterogeneous, across subjects and even to some extent across problems within subjects, I study search on the individual problem level.

The “ALT” type begins search in the problem by performing at least two sequential within alternative transitions (three sequential look-ups in the same alternative). She can then transition once, either within attribute or neither within alternative nor within attribute, followed immediately by at least two sequential within alternative transitions, and so on.

Figures 2.2.1 and 2.2.2 show the two most common forms of ALT search: typewriter, and "boustrophedon," respectively. Figure 2.2.1 shows an ALT type whose transitions between alternatives are exclusively the typewriter type first discussed in Section II. 82% of transitions across alternatives for all Problems containing ALT search are typewriter. 14% are within attribute transitions. Of these within attribute transitions 81% are followed by an immediate reverse in direction of within alternative transitions, or boustrophedon type search.[9] Figure 2.2.2 shows an example of an ALT type in which transitions between alternatives are exclusively boustrophedon.

The positive integers in Figures 2.2.1 and 2.2.2 correspond to the order of look-ups in the exemplar subject's search sequence. Attributes 1-10 are listed from left to right, and alternatives 1-8 from top to bottom. Attribute one values are always observable to the subject, thus cannot be "looked-up." A choice box for each alternative can be found towards the bottom of each figure (C1-C8). After searching attributes, subjects choose an alternative- indicated here by the time at which the choice is made (in seconds, to one decimal place). Thus, in Figure 2.2.1, for example, the subject performs 16 attribute look-ups, then chooses alternative 4 without confirming, then chooses alternative 5 without confirming, before going back to choose and confirm alternative 4 at 35.2 seconds.

[9] According to Wikipedia, the etymology of boustrophedon is from the Ancient Greek βους, "ox" + στρεφειν, "to turn", because the hand of the writer goes back and forth like an ox drawing a plow across a field and turning at the end of each row to return in the opposite direction.

Alt\Att	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
1		1	2	3						
2		8	9	10						
3										
4		4	5	6	7					
5										
6										
7		11	12	13						
8		14	15	16						
		C1	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*	
					17,35.2	18				

Figure 2.2.1: ALT type: Typewriter Transitions Across Alternatives

Alt\Att	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
1		1	2	3	4					
2		8	7	6	5					
3		9	10	11	12					
4										
5		17	18	19	20					
6		16	15	14	13					
7										
8		23	22		21					
		C1	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*	
						27.1	24	26	25	

Figure 2.2.2: ALT type: Boustrophedon Transitions Across Alternatives

The “ATT” type begins search in the problem by performing at least two sequential within attribute transitions (three sequential look-ups in the same attribute). She can then transition once, either within alternative or neither within alternative nor within attribute, followed immediately by at least two sequential within attribute transitions, and so on.

Analogous typewriter and boustrophedon transitions between attributes (columns) exist for ATT search problems. 31% of all transitions across attributes are typewriter and 56% are within alternative. Of these within alternative transitions 69% are boustrophedon. Examples of ATT searchers with typewriter and boustrophedon transitions between attributes (Figures 2.3.1 and 2.3.2) are like the search patterns in Figures 2.2.1 and 2.2.2, rotated 90 degrees.

Alt\Att	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
1		1	9	18						
2		2	10	19						
3		3	11	20						
4		4	12	21						
5		5	13	22						
6		6	14	23						
7		7	15	24						
8		8	16	25						
		C1	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*	
		27	17,26,28.3							

Figure 2.3.1: ATT type: Typewriter Transitions Across Attributes

Alt\Att	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
1		1	16	17						
2		2	15	18						
3		3	14	19						
4		4	13	20						
5		5	12							
6		6	11							
7		7	10							
8		8	9							
		C1	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*	
				21, 29.9	22					

Figure 2.3.2: ATT type: Boustrophedon Transitions Across Attributes

The "ALT-ATT" and "ATT-ALT" types switch back and forth between ALT and ATT type search, within the same problem (with no restriction on how many times they switch back and forth). ALT-ATT types (Figure 2.4) first perform ALT search, whereas ATT-ALT types (Figure 2.5) first perform ATT search. [10]

[10] The average transition frequency from ALT search to ATT search ranges from 5-14% (depending on how many switches back and forth precede), and the average transition frequency from ATT search to ALT

Alt\Att	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
1		16								
2		17								
3		18								
4		19								
5		10,20	11	12	13	14	15			
6		1	2	3	4	5	6	7	8	9
7		21								
8		22								
		C1	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*	
							19.6			

Figure 2.4: ALT-ATT Type

Alt\Att	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
1			8							
2		6	7	16	17	18	19			
3		5	9	15	22	21	20			
4		4	10		23	24,29	25			
5		3	11		28	27	26			
6		2	12							
7		1	13							
8			14							
		C1	*C2*	*C3*	*C4*	*C5*	*C6*	*C7*	*C8*	
					31.4	30				

Figure 2.5: ATT-ALT Type

96.6% of subjects' 7552 search problems can be designated as ALT, ATT, ALT-ATT, or ATT-ALT search, based on the following simple classification scheme. Search problems are designated as ALT type if they contain at least one instance of two sequential within alternative transitions and no instances of two sequential within attribute transitions. Search problems are designated as ATT type if they contain at least one instance of two sequential within attribute transitions and no instances of two sequential within alternative transitions. Problems that classify as ALT type for the first

---

search ranges from 79-86%. Thus, most ALT-ATT problems are actually ALT-ATT-ALT problems, and most ATT-ALT problems are indeed ATT-ALT problems.

portion of search, then as ATT types in the immediately following portion of search are deemed ALT-ATT Types. Similarly, ATT-ALT types are ATT types for the first portion of search in a problem, then switch to being ALT types for the immediately following portion of search.

Table 2.3 shows the number of search problems selected to each search type along with their corresponding fits. Together, the four search types absorb 98.2% of subjects' look-ups.

Table 2.3: Fits of Search Types

Search Type	N	Fit (%)
ALT	6089	98.5
ATT	146	92.2
ALT-ATT	345	95.4
ATT-ALT	717	98.2

IV. Analysis of Behavior Using a Partial Characterization of Optimality

In this section I present a partial characterization of optimality in the form of five necessary conditions, using it as a partial rational benchmark against which I compare subject behavior. Pooling the data first by problems, I show that violation rates of optimality are significant and substantial. Then pooling problems by search Types, as classified in Section III, I show that violation rates differ significantly and substantially across types. This section also includes a brief discussion of how my partial characterization of optimality falls short of a full characterization, and it closes with an econometric analysis of an important feature of search- alternative switching, as related to one of my necessary conditions. Interpretations of violations to necessary conditions will, for the most part, be postponed until Section V.

I demonstrate the first two necessary conditions, which are non-trivial, in the appendix. All conditions assume risk neutrality, perfect reasoning (including memory), and costly search.[11] The first necessary condition is as follows.[12] (1) "If an attribute in alternative x is searched, it must be the case that there does not exist an alternative y that is both weakly less searched than x, and has a weakly higher cumulative revealed value than x, with at least one of these inequalities strict." Every violation of optimality identified by this condition signifies both conditional over-searching in one alternative, and conditional under-searching in another. The intuition behind this necessary condition is that because search can only end when one alternative has sufficiently separated itself in value above the others, it is best to first search alternatives with relatively high revealed values, and relatively more remaining uncertainty; the worst thing to do is search an alternative with a relatively low revealed value and relatively little remaining uncertainty. This necessary condition is important because violations of it partially address the important question of whether subjects' search too many or too few alternatives, and too many or too few attributes within alternatives- which is a natural extension to the question of whether subjects search too many or too few alternatives in single attribute search.[13]

Two qualitatively distinct search actions violate optimality via necessary condition one: transitioning within an alternative, and switching from one alternative to

[11]It is reasonable to assume GL's subjects are risk neutral given that payoffs per search problem are small ($0.52), along the line of Rabin (2000)'s argument. In addition GL encourage risk neutrality by telling subjects to "Choose the row you think has the highest sum," which can be seen in Figure 1, above the choice boxes.
[12]Necessary conditions are stated more formally in Appendix A.
[13]Camerer (1995) reports that a robust finding in single-attribute search problems is that subjects systematically search too few alternatives. Bearden et. al. (2007) find the same result in their two-attribute search problem.

another. In terms of payoff consequences in the data, violations that occur with alternative switching are particularly costly because subjects tend to search several attributes in the "wrong" alternative once they have switched to it, which further compounds the cost. Weitzman (1979) determines which (asymmetric) alternative should be searched next, if one is to be searched at all, in single attribute search. My condition is weaker in the sense that it does not always identify which alternative should be searched next, but it is related in the sense that it does identify cases where the switch from one alternative to another is sub-optimal.

My second necessary condition states (2) "An attribute cannot be searched if there exists another unsearched attribute with greater variance in the same alternative." This is the last of two necessary conditions that I demonstrate in the appendix. Because the attributes within each alternative are i.i.d., aside from a declining variance structure, searching the remaining unsearched attribute with highest variance provides the largest value of information.

Condition three states that (3) "the same attribute must never be searched more than once in the same problem." This "no repeated look-ups" condition follows directly from the perfect reasoning (memory) assumption and the existence of time costs. Condition four requires that subjects (4) "choose the alternative with highest revealed value," conditional on the attributes searched in each problem. This condition follows directly from the assumptions of risk neutrality and perfect reasoning (memory). Finally, condition five simply states that (5) "unconfirmed alternative choices cannot occur," and follows directly from the existence of time costs.[14]

Together, these five necessary conditions form a partial characterization of optimality. A complete characterization, on the other hand, would ensure that each searched attribute yield the highest (positive) expected payoff, with search ending only when the search of any further attribute would yield a negative expected payoff. Also, at the end of search, the alternative with the highest revealed value would be chosen. In terms of an observable search pattern, the optimal policy would begin search in attribute 2 of the alternative with the highest (observable) attribute one value. The (ex-post) optimal number of attributes in that alternative would then be searched, preceding the transition to attribute 2 of the remaining unsearched alternative with highest attribute one value. An (ex-post) optimal number of attributes would then be searched in this alternative before either switching to the previously unsearched alternative with highest attribute one value, or returning search to the highest variance unsearched attribute in one of the previously searched alternatives, and so on, until search ends. As such, the optimal search policy will probably contain a fair amount of "jumping" back and forth between previously searched alternatives. Sections II and III show that in general, subjects' search is not jumpy, but highly systematic, "sweeping" through alternatives. The analysis contained in this section will make it evident that a large fraction of behavioral violations of optimality occur precisely due to this type of highly non-

[14] As stated in Section I, in order to choose an alternative (to consume) the corresponding choice box for that alternative must be clicked on a first time with the computer mouse, then again a second time to confirm the choice.

contingent search, that sweeps not only across the attributes within alternatives, but across the alternatives as well.

Of the five necessary conditions I present, condition one identifies which alternative search should begin with, if it should begin at all, and condition two always identifies which attribute should be searched within an alternative given that the alternative is being searched. Given that a transition is occurring from one alternative to another, condition one identifies a list of alternatives that cannot be transitioned to (depending on the search history, this list may or may not be exhaustive). In addition, condition one identifies cases in which search should switch from one alternative to another. Of course, condition four identifies the optimal alternative choice at the completion of search.

That my partial characterization of optimality falls short of a complete characterization is primarily due to my inability to determine which alternative is better to search when, relative to another, one alternative has a higher cumulative revealed value and is more thoroughly searched. As such, I cannot always determine when search should switch out of an alternative, nor can I always make a complete prediction of which alternative should be switched to, when a switch does occur. Also, independent of this shortcoming, I do not know when search in a problem should stop.

However, even without a full characterization of optimality, my conditions allow useful inferences about several, and the most obvious, of subjects' systematic deviations from optimal search. In particular, 97% of all search problems, and 62% of all search actions within search problems violate at least one of my five necessary conditions. Where Section II shows that subjects' search patterns tend to be "smooth" and highly systematic, this section shows that they also tend to violate necessary conditions for optimality at high rates. Where Section III shows that 98% of all search behavior is explained by four tightly fitting search types, this section shows that violation rates of several necessary conditions (as well as other summary statistics describing search behavior and performance) vary significantly and substantially across the four search types.

Summary statistics for all search problems (pooled), as well as by search type, can be found in Table 2.5. From left to right columns contain (N)- the number of problems classified as each type, (Alt:Att)- the average ratio of within alternative to within attribute transitions, (Time)- the average end time of search in a problem, (LU's)- the average number of look-ups in a problem, and, the violation rates of (1) necessary condition one, ((1) ∩ Within)- necessary condition one violated by a within alternative transition, ((1) ∩ Switch)- necessary condition one violated by sub-optimal alternative switching, (Var)- necessary condition two (search maximum variance attribute in alternative), (RLU's)- necessary condition three (repeated look-ups), (Correct)- necessary condition four (choose highest revealed alternative), (Uncon)- necessary condition five (unconfirmed alternative choices), (Union NCO's)- search actions in which at least one necessary condition is violated, and (Exp. Score)- the average expected score, or the average revealed value of the chosen alternative. Throughout, (%) represents the percentage of violations relative to all look-ups performed in problems of each search type.

Using a two-sided Mann-Whitney U-test, frequencies (and averages) within each column, (Time) through (Exp. Score), are pair-wise statistically distinguishable, at the five percent level, in all but two cases. In Table 2.4 these two cases are numbered (1-2), in corresponding pairs.[15]

Table 2.4: Summary Statistics & Violations of Necessary Conditions: Pooled & by Type

Types	N	*Alt:Att*	Time	LU's	(1) (%)	(1) ∩ W/in (%)	(1) ∩ Switch (%)	(2) Var (%)	(3) RLU (%)	(4) Corr. (%)	(5) Uncon (%)	Union NCO's (%)	Exp. Score
Pooled (all 4 types)	7297	11.8:1	34.0	31.7	49.7	38.6	11.1	5.3	11.8	67.0	4.3	61.5	51.6
ALT	6089	1:0	33.8	31.8	48.9	40.2	8.7	4.8	10.7	68.4	4.4	60.2	53.4
ATT	146	0:1	18.2	13.4	62.6	6.6	56.0	11.0	8.2	56.8	6.1	75.8	30.4
ALT-ATT	345	4:1	41.2	38.5	51.8	35.0	16.8	7.3	20.0	60.0_1	3.1	66.8	44.1_2
ATT-ALT	717	1.8:1	35.4	31.8	54.0	29.6	24.4	8.2	17.2	62.7_1	4.1	69.0	45.8_2

(1) "If an attribute in alternative x is searched, it must be the case that there does not exist an alternative y that is both weakly less searched than x, and has a weakly higher cumulative revealed value than x, with at least one of these inequalities strict." [16]

A violation of this condition occurs in 50% of subjects' look-ups. These violations occur in the form of either attribute transitions within an alternative, or as switches across alternatives. Violations of each qualitatively distinct type of violation are explored below.

Within Alternative Transitions

The Majority of violations of necessary condition one, 77%, occur as within alternative transitions. These are cases when the currently searched alternative is

[15] Independence between search types is, of course, questionable here.

[16] Proof is in the appendix.

searched deeper, though the subject would have been better suited switching to a less searched alternative.

Violations of condition one that occur as a within alternative transition can either be preceded by a violation of the same type, or a look-up that does not violate the condition. It is interesting to look at violations of the latter variety, as a test of contingency in search. In other words, I will look at cases where the previous look-up did not violate optimality, but yielded a realization that necessitated switching out of that alternative on the next look-up. Specifically, this attribute realization makes the current alternative both weakly more searched than, and have a cumulative revealed value that is weakly less than, with one of these inequalities strict, another alternative. In 73% of the cases in which an attribute realization necessitates switching out of an alternative, in this way, subjects fail to do so.

Table 2.4 shows that as a fraction of all search transitions, ALT types violate condition one on within alternative transitions at a much higher rate (40%) than ATT types (7%). This is simply because a much lower fraction of ATT look-up transitions are within alternative. Not surprisingly ALT-ATT and ATT-ALT violation rates are roughly proportional to their relative amounts of embedded ALT search.

<u>Alternative Switching</u>

While only 23% of necessary condition one's violations occur on alternative switching, these violations are especially important because, once an alternative is switched to, subjects tend to then search several of its attributes in sequence, which exacerbates the cost of them having switched to the "wrong" alternative.

For the remainder of this section I will limit my consideration of necessary condition one's alternative switching violations to alternative switches that occur to previously unsearched alternatives. Doing so allows me to dispel of concerns associated with comparing alternative switches with switched-to alternatives that have been searched to different extents. In addition, this restricted sample still accounts for 85% of all violations of condition one that occur in alternative switching.[17] Violations of this type are of additional interest due to the transparently sub-optimal behavior they suggest- when deciding which unsearched alternative to switch to, the subject need only switch to the alternative with the "highest number in front of it-" its fully observable attribute one value (see Figure 2.1).

Despite the relative importance (and ease) of complying with necessary condition one in alternative switching, only 51% of subjects' alternative switches do so. Moreover, as the ex-post number of alternatives searched within a problem increases, the likelihood of compliance falls dramatically.

This phenomenon can be seen clearly in Table 2.5, where the (ex-post) number of alternatives searched in a problem lies on the vertical axis, and the sequential order of alternative switches, within the problem, lies on the horizontal axis. The table contains frequencies of optimal alternative switches in columns labeled 1-8, while the final two columns report the number (# Opt.) and fraction (% Opt.) of problems for which all alternative switches are optimal. The probability of complying by chance, for each sequential switch, is shown in the bottom row of the table. Using a two-sided Mann-Whitney U-test, frequencies within each column (1-8) are pair-wise statistically

[17] In general, 78% of all alternative switches occur to previously unsearched alternatives.

distinguishable, at the five percent level, in all but five cases. In Table 2.4 these five cases are shown to the left of fractions, in corresponding pairs (1-5).

Table 2.5: Frequencies of Sub-Optimal Alternative Switching. On the vertical axis are the total number of alternatives searched in the task, and on the vertical axis is the sequence those alternatives are searched in.

	N	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**# Opt.**	**% Opt.**
1	713	0.75								537	0.75
2	846	0.63	0.70							410	0.49
3	976	[1]0.57	0.54	0.65						261	0.27
4	937	[1]0.55	0.49	0.50	0.63					120	0.13
5	848	0.50	0.42	[4]0.43	[5]0.44	0.57				29	0.03
6	641	[2]0.43	[3]0.37	[4]0.39	[5]0.39	0.44	0.54			10	0.02
7	575	[2]0.43	[3]0.35	0.29	0.30	0.37	0.44	0.65		4	0.01
8	1905	0.22	0.19	0.21	0.22	0.30	0.37	0.57	1.0	2	0.00
Cha-nce		*0.13*	*0.14*	*0.17*	*0.20*	*0.25*	*0.33*	*0.50*	*1.0*		

If subjects' alternative switches always complied with necessary condition two, all frequencies in Table 2.5 would be equal to one. Clearly, this is not the case. Furthermore, the rate of compliance falls dramatically as the ex-post number of alternatives searched in a problem increases. This result is particularly pronounced in that problems with only one searched alternative yield a 75% compliance rate. However, when eight alternatives are searched in the problem, the first alternative transition complies only 22% of the time- just 9% above chance. In addition, a remarkably low 2 of 1905 problems maintain compliance to condition two throughout all alternative switches when 8 alternatives are searched. This finding is slightly surprising, given the relative importance of and ease with which searching the unsearched alternative with the biggest number in front of it should be.

That compliance rates decrease systematically with the ex-post number of alternatives searched seems to indicate that subjects know ahead of time whether they will search in a more or less contingent manner. Search may actually resemble a two-step process, where in the first step the subject decides on searching a particular "chunk" of attributes across alternatives, and in the second step searches that chunk in the preferred manner, which is often sub-optimal (assuming perfect reasoning and memory).[18] This type of two-step processing strategy seems particularly supported by problems in which subjects search all eight alternatives. In these problems compliance rates of alternative switches are scarcely above chance. It may be the case that subjects decide ahead of time to search all eight alternatives, so then proceed to search them in some manner that is (sub-optimally) preferred to transitioning to unsearched alternatives with highest revealed values. In any case, that subjects are violating compliance with condition two at high rates is clear, but why they are doing so remains unclear. Thus, an econometric analysis, reported at the end of this section, is conducted in order to gain insight into what factors, aside from optimality considerations, are linked to alternative switching behavior.

In terms of heterogeneity in behavior across types, Table 2.5 shows that ATT types switch alternatives sub-optimally (56%), much more often than ALT types (9%), as a percentage of all performed look-ups. This difference is mostly due to alternative switches being more relatively abundant in ATT search than in ALT search. However, even when conditioning on alternative switches alone, ATT types still violate the

[18] This would be a type of rational optimizer that lapses into Stiglerian (classic, pre-committed) search, before waking up every once in a while to re-optimize.
GL's behavioral 'Directed Cognition' model assumes that this type of non-contingent search can occur within but not across alternatives. (GL, p 1052)

condition more frequently (61% vs. 50%) than ALT types.[19] ATT types often switch alternatives in sweeping sequentially adjacent transitions up or down attribute columns (ALT's do this also, but intersperse search within alternatives along the way). Because alternatives' spatial positions are randomly ordered with respect to their attribute values, this type of sweeping search leads to necessary condition one being violated at high rates. Not surprisingly, ALT-ATT and ATT-ALT search types violate the condition at rates roughly proportional to their relative ratios of ATT to ALT search.

(2) "An attribute cannot be searched if there exists another unsearched attribute with greater variance in the same alternative."

Because attributes decline in variance from left to right, a subject should never search an attribute that has an unsearched attribute to its left, in the same alternative. This result is due to attributes with higher variance having more information value than attributes with lower variance. Condition two identifies sub-optimal attribute look-ups within an alternative, while condition one identifies that all attributes within certain alternatives are sub-optimal to search. Thus a look-up can violate both necessary conditions one and two.

Roughly 1.6 look-ups per problem (5% of all look-ups) violate this condition. Across search types this violation rate is higher in ATT search. This is because ATT types sometimes search an entire dominated attribute, in sequence. Most violations in ALT search, on the other hand, occur as a result of Boustrophedon search.

(3) "The same attribute must never be searched more than once in the same problem."

[19] This difference is significant at the 5% level, using the two-sided Mann-Whitney U-test.

Due to the scarcity of time in both experimental treatments, subjects should never repeat attribute look-ups[20]. However, on average 3.6 look-ups per problem, or 12% of all look-ups are repeats. Repeat look-ups can occur simultaneous with violations of necessary condition one and/or two.

Repeat look-ups occur most often in ALT-ATT (20%) and ATT-ALT (17%) search, due to frequent overlapping in the ALT and ATT components of their hybrid search patterns.

<u>(4) "Choose the alternative with highest cumulative revealed value."</u>

Subjects should always choose the alternative with the highest cumulative revealed value in order to maximize expected payoffs, but fail to do so in 33% of problems.

ATT types violate this condition at a slightly higher rate than ALT types (68% vs. 57%), and ALT-ATT and ATT-ALT types predictably violate it more than ALT's and less than ATT's.

<u>(5) "Unconfirmed alternative choices cannot occur."</u>

In order to choose an alternative, at the end of search one must scroll to the unique choice box representing one's desired choice, (see Figure 2.1) click once with the computer mouse to choose that alternative, then click a second time to confirm the choice. However, many times during and after the course of searching attributes, subjects commit unconfirmed choices. This behavior is clearly sub-optimal given that these actions waste time that could have been spent searching still unrevealed attribute values.

[20] There is one rare case in which it is not sub-optimal to repeat attribute look-ups. This is the case in which a subject in the Exogenous treatment has already searched every attribute and time still remains. This occurs in only 1 of the 7552 problems studied.

On average, this violation of optimality occurs in .97 look-ups per problem, or 3% of all look-ups. ATT types violate this condition at a slightly higher rate than ALT types, as do ATT-ALT's relative to ALT-ATT's.

Comparing Search Types Across Necessary Conditions

In comparing behavior across search types, for all necessary conditions, it is reasonable to focus on ALT and ATT types as they are the primitive search types that synthesize into ALT-ATT and ATT-ALT types. Accordingly, ALT and ATT are the two most behaviorally disparate search types, with ALT-ATT and ATT-ALT statistics representing a convex combination of ALT and ATT statistics in all but three cases. Namely, ALT-ATT and ATT-ALT types tend to commit more look-ups, take longer, and commit more repeats (due to overlap) than ALT and ATT types.

Relative to ATT types, ALT's take more time, perform more look-ups, violate necessary condition one less often, violate necessary condition one via within alternative transitions at a much higher rate, violate necessary condition one via alternative switching at a much lower rate, search attributes with lower relative variance less often, choose the correct alternative more often, score higher, and, in general, perform a higher rate of look-ups that violate none of the five conditions.[21]

Econometric Analysis of Alternative Switching:

[21] It should be noted that Table 2.5 pools data from the Exogenous and Endogenous treatments of GL's experiment. Although for the most part relative differences between types' behavior across treatments is constant, there is one notable exception- which occurs for ATT types. Relative to the Exogenous treatment, ATT types In the Endogenous search roughly half as many attributes in half as much time, and choose the correct alternative almost 30% more often. This anomalous result will be briefly addressed in Section V.

The econometric analysis of this section suggests that sub-optimal (spatial) considerations and optimal considerations are roughly equally present in search behavior. I run separate conditional logit regressions for each of the first three alternative switches of each subject, pooling data across all subjects and problems. Subsequent alternative switches are expected to yield similar results, so are omitted.

For the first alternative searched, the eight searchable alternatives differ by only two characteristics. The first difference is variation in the fully observable attribute one value of each alternative, which is the only information of importance for the rational searcher.[22] The second is the relative spatial location of each alternative, which is irrelevant to the rational searcher. Thus the two independent variables in the conditional logit model are the value of attribute one, drawn from a mean-zero normal distribution, and the height of the alternative, which can take any value between 1(top) and 8(bottom). I use this simple model in order to count the number of subjects for which the effect of either independent variable is statistically distinguishable from zero. Rational subjects should only consider the value of attribute one for each alternative, so only its corresponding estimated coefficient should be statistically distinguishable from zero.

Please see Appendix D for formal specifications of the systems of equations used to estimate conditional logit coefficients, here and later in this section.

Results from the first alternative transition regression in Table 2.6 show that the effect of attribute one value is statistically distinguishable from zero for 290 of the 390

[22] This statement would not always be true if searching one alternative rather than another conferred significant time gains, but it seems fair to assume that time gains of this type are small enough to ignore.

experimental subjects. However, the spatial, or height effect, is also distinguishable from zero for 275 of the 390 subjects.

Aside from the 373 subjects for whom the regression coefficients are estimated successfully, there are 17 subjects with a perfectly identified alternative transition for one of the two independent variables. 16 of these subjects always choose the first alternative (top), while one always searches the alternative with highest attribute one value. These respective subjects are added into Table 2.6 in order to report the combination of subjects for which each independent variable is either statistically distinguishable from zero or perfectly identified. Also reported in Table 2.6 are the number of subjects with both independent variable coefficients statistically distinguishable from zero, and the number of subjects with strictly one coefficient statistically distinguishable from zero, along with their respective average coefficient values.[23]

By simply counting the number of subjects for which each coefficient is statistically distinguishable from zero we observe that the effects of attribute one value and spatial location of the searched alternative are virtually identical.

[23] The interpretation of estimated coefficients in the conditional logit model is not entirely straightforward. Perhaps the easiest way of interpreting the coefficients is by using an equivalence easily derived from the original likelihood expressions for each alternative:

$$\log\left(\frac{P_m}{P_s}\right) = \beta_v (V_{tm} - V_{ts}) + \beta_H (H_{tm} - H_{ts})$$

where m and s are different alternatives, and t is the index identifying the search problem. All else equal, a change in the log ratio of the relative probabilities of searching any 2 given alternatives corresponds to an equal change in the product of either of the coefficients and the relative difference in its corresponding independent variable across the two alternatives. Thus, a natural way of comparing coefficients is to determine how much of a change in one independent variable is necessary to offset the effect of a one unit increase in the other.

Table 2.6: Conditional Logit Results: First Alternative Searched

	# of sig. subjects (+P	Avg. β_V	Avg. β_H
$\beta_{V\text{ sig.}}$	290+(1)=**291**	.112	-.378
$\beta_{H\text{ sig.}}$	275+(16)=**291**	.0606	-.795
β_V & $\beta_{H\text{ sig.}}$	**202**	.0789	-.516
$\beta_{V\text{ sig,}}$ $\beta_{H\text{ not}}$	88+(1)=**89**	.187	-.0632
$\beta_{H\text{ sig,}}$ $\beta_{V\text{ not}}$	73+(16)=**89**	.00996	-1.568

In order to specify a model representing factors related to the second alternative transition two additional independent variables are introduced. The first is the absolute distance (in rows) of the remaining 7 alternatives to the first alternative searched. The second is an indicator variable for down transitions. Thus the model for second alternative transition captures, for each of the remaining 7 unsearched alternatives, the effects of attribute one value, height of the alternative, absolute distance of the alternative from the first alternative searched, and whether the alternative is below or above the first alternative searched.

As in the case of the regression for the first alternative transition the number of subjects for which each variable yields a coefficient estimate statistically distinguishable from zero is reported along with average coefficient values for these subjects. The effect of attribute one value is statistically distinguishable from zero for 287 of the 390 subjects. Distance is statistically distinguishable from zero for 247 subjects, while height is for 127, and the down dummy for 43. However, in this regression there are also perfectly identified types and subjects with perfect multicollinearity and/or zero variation in the down indicator. Many subjects, for example, always search alternative one first and alternative 2 second, thus they are always perfect height types, perfect down types, and perfect distance types. Correspondingly, these coefficients are either dropped or estimated incorrectly in the regressions. There are 99 subjects perfectly identified in

Height, Distance, or Down[24]. Of the 99 cases, 83 always transition down, 51 always transition to an adjacent alternative, and 25 always transition to the (spatially) highest remaining alternative. Table 2.7 shows the combination of subjects for which each independent variable is either statistically distinguishable from zero or perfectly identified.

Table 2.7: Conditional Logit Results: Second Alternative Searched

	# of sig. subject (+PI)	Avg. β_V	Avg. β_D	Avg. β_H	Avg. β_I
β_V sig.	**287**	0.173	-1.423	0.672	-0.0166
β_D sig.	247+(51)=**298**	0.0739	-2.679	1.384	-1.471
β_H sig.	127+(25)=**152**	0.226	-4.903	2.125	-0.488
β_I sig.	43+(83)=**126**	0.0675	-1.268	0.216	0.760
β_V & β_D sig.	**204**	0.0868	-1.941	1.087	-1.096
β_V sig, β_D not	**83**	0.386	-0.150	-0.348	2.636
β_D sig, β_V not	43+(16)=**59**	0.0126	-6.182	2.794	-3.253

The specification of the regression for the third alternative transition is identical to that used for the second. The results are also similar. The effect of attribute one values are statistically distinguishable from zero for 250 subjects, as is distance for 238, height for 96, and the down dummy for 44. For the third alternative transition there are 80 subjects with regression coefficients dropped or inestimable. Of these 80 subjects 52 always transition down, 52 always transition to an adjacent alternative, and 30 always choose the (spatially) highest remaining alternative. These perfectly identified types are added into Table 2.8 to show the combination of subjects for which each independent variable is either statistically distinguishable from zero or perfectly identified.

[24] As long as at least 90% of transitions comply with perfectly identified behavior I consider the subject perfectly identified

Table 2.8: Conditional Logit Results: Third Alternative Searched

	# of sig. subjects (+PI)	Avg. β_V	Avg. β_D	Avg. β_H	Avg. β_I
β_V sig.	**250**	0.100	-0.895	0.140	0.967
β_D sig.	238+(52)=**290**	0.0530	-2.175	0.792	0.338
β_H sig.	95+(30)=**125**	0.0343	-6.504	3.236	1.122
β_I sig.	44+(52)=**96**	0.0500	-0.741	-0.561	1.808
β_V & β_D sig.	**165**	0.0696	-1.251	0.223	1.070
β_V sig, β_D not	**85**	0.160	-0.203	-0.0215	0.769
β_D sig, β_V not	73+(52)=**125**	0.0154	-4.261	2.0786	-1.315

Despite rationality dictating that subjects only consider attribute one values when deciding which unsearched alternative to transition search to, results of the conditional logit analysis show that subjects systematically incorporate specific non-optimal considerations into their alternative switching behavior. The first alternative transition analysis suggests that the spatial location (non-optimal consideration) of the alternative and its attribute one value are of roughly equal importance to subjects. This finding corresponds to a general proclivity of subjects in the GL data to start with the highest alternatives, then transition downward. 78% of all alternative transitions, in fact, are downward. Regressions for the second and third alternative transitions reveal that attribute one value and adjacency (non-optimal consideration) are the most prevalent transition considerations, and are roughly equally represented in subjects' search. The regressions also suggest that consideration of alternative height, as well as transitioning downward, are less prominent, but also present. An interpretation of these results is made in the next section.

V. Evidence of Working Memory Limitations

There is strong evidence in the GL dataset that working memory limitations play an important role in shaping search behavior. This section documents several- some obvious, others subtle- examples of such evidence.

The two clearest manifestations of working memory limitations in subjects' behavior are found in violations of necessary conditions three and four. Subjects, on average, repeat 11.8% of all look-ups, and choose the wrong alternative in 33% of problems. Limitations in working memory are responsible for these two systematic violations of optimality.[25]

Several other important search behaviors affected by working memory limitations, including violations necessary condition one, will be organized below as consistent with one of two results demonstrated in my accompanying paper. A description of the basic intuition behind these results will be sufficient for the purposes of this section.

Sanjurjo 2008b builds on the work of Newell and Simon (1972), Johnson, Bettman, and Payne (1993), and Crawford (2008) in creating a simple model of working memory load (henceforth WML) in search. The first result to be used here is that (1) ATT search is more WML intensive than ALT search. The basic intuition behind this result is that, because alternative values are summed within alternatives, even after several alternatives have been searched, ALT search requires only two sums to be stored in memory at one time: the highest alternative sum value so far observed, and the sum value of the currently searched alternative. ATT search, on the other hand, requires

[25] I use the general term "limitations" here, which can translate into costs, or literally refer to hard-wired short term memory limits.

memory storage of running sum values for each alternative searched (in multiple-attribute search). Thus, WML in ATT search is not only larger, but it increases with the number of alternatives searched, whereas ALT search does not. The second result is that (2) more systematic (rehearsed) search patterns are less WML intensive than unsystematic (non-rehearsed) search patterns. The intuition behind this result can be gained by imagining a simple example in which several (ten, for example) attributes within an alternative can be searched in two different patterns. The first pattern follows a simple (rehearsed) convention, such as left-to-right adjacent, for the entire sequence of ten look-ups. In the second pattern attributes are searched in a (non-rehearsed) randomly selected order. WML is higher in the second pattern than in the first, because it requires not only that the observed attribute values be remembered, but that the searched locations of those values be remembered as well.

The first search behavior to be reported relates to result (1)'s assertion that the difference between WML in ALT and ATT search increases with the number of alternatives searched. The ratio of within alternative to within attribute search transitions is considerably higher in the GL dataset than in similar, but dimensionally smaller, experiments. Payne, Bettman, and Johnson (1988) designed a series of MouseLab experiments which differed from GL in that subjects chose one of four four-outcome lotteries, but was otherwise virtually identical in terms of the rules of search. In their experiments the average ratio of within alternative to within attribute search transitions ranged from roughly 1:2 to 2:1. The average ratio in GL's data is 13:1. Because the WML for ATT doubles with a doubling in the number of alternatives, while the WML for

ALT remains the same, it should not be surprising that the relative presence of ALT search is much higher in the larger design.

Another possible effect of working memory limits, also related to result (1), is found in a curious behavioral disparity between ATT types in each of GL's two experimental conditions: Exogenous (explicit time limits) and Endogenous (25 minute problem buffet). ATT types in the Endogenous treatment perform an average of 8.7 (roughly one column of attributes) look-ups and choose the highest revealed alternative 72.6% of the time, whereas ATT types in the Exogenous treatment perform an average of 16.9 (roughly 2 columns of attributes) and choose the highest revealed sum only 45.4% of the time.[26] Predicted WML is much higher for multiple-attribute ATT search than for single-attribute ATT search (because the multiple attribute version requires storing all values in the first searched column of attributes), a fact that suggests that WML might be related to errors in alternative choice.[27]

Result (2) provides a possible explanation for subjects' high violation rates of necessary condition one- both in the form of alternative switching and within alternative transitions. Optimal search in GL's problem contains an a-priori unpredictable ordering of optimal alternative switching. In addition, optimal search probably contains a fair amount of switching back and forth between alternatives on sequential look-ups; in other words it is probably much more contingent (spatially jumpy across alternatives) than the

[26] Because this was the only example of truly disparate behavior between the two treatments, for my four search types, I refrained from posting separate summary tables for each of the treatments, in Section IV.
[27] The other obvious potential explanation is that subjects in the Endogenous treatment are searching less and choosing the right alternative more frequently due to a problem selection effect. However, a simple analysis of spreads between best and second best alternatives before and after subjects' search shows no evidence of the alternative choice in the Endogenous treatment being any easier; in fact they are slightly more difficult (closer in value). Thus there appears to be no evidence of a selection effect driving this disparate behavior across treatments.

subjects' behavior we observe. The econometric analysis of Section IV shows that subjects have strong spatial search biases; they tend to start high and switch (downward) to adjacent alternatives. These biases clearly lead to high violation rates of necessary condition one because alternatives are spatially unordered with respect to their attribute one values. Although these biases are costly, they also reduce the burden of WML- due to the same intuition contained in the example of random vs. systematic search of ten attributes in an alternative. Likewise, violations of necessary condition one that occur as within alternative transitions may be part of a "one-time-through" approach, where subjects search an alternative, move on, and never come back to that alternative. This type of "one-time-through" approach severely reduces WML relative to optimal search, which requires the subject to remember current sums of all alternatives (because each one might be returned to and searched further), as well as which attributes have already been searched in those alternatives (so that they are not searched again and mistakenly added to the sum).

Two other possible explanations for several of the search behaviors presented in this section are reading bias and time-saving considerations. The standard reading bias (for American students) predicts that subjects search from left to right (adjacent) within alternatives, and top to bottom (adjacent) across alternatives, which is highly consistent with several of the search behaviors described above. However, the strength of this argument as a primary cause of observed behaviors is severely weakened by the tendency of subjects not to show the same extent of reading-type search in other similar experiments, such as Payne, Bettman, and Johnsons' (1988). In addition, the econometric analysis of alternative transitions shows that adjacency is significant for more than twice as many students as the dummy for downward transitioning across alternatives, which further undermines a strict reading bias interpretation. Nevertheless, to the extent that working memory limitations are playing a role in the determination of search patterns, the standard reading-type search pattern is the most familiar and rehearsed WML minimizing search convention that subjects can adopt, so it is not surprising that it is consistent with much of subjects' observed search, even if the search patterns themselves are actually being driven fundamentally by working memory limitations.

Another obvious possible explanation for the over-searching of alternatives, and sub-optimal alternative switching, is time-saving considerations. In fact, any tendencies that subjects might have to search in a more systematic and rehearsed fashion will not only reduce WML, but also reduce search time. Though the time-saving effect is real, it is questionable how significant a role it plays in determining search patterns; given the physical speed with which a computer mouse can move across the computer display, the gains are slight.

VI. Conclusion

In this paper I build on the work of GL in two ways. First, I study highly heterogeneous search behavior on an individual problem level. Second, I provide a partial characterization of optimality in their multiple attribute, multiple alternative full recall search problem, which I use as a partial rational benchmark, against which I test subject behavior. Although my characterization of optimality is only partial, it successfully identifies theoretically important violations of optimality in search, and empirically catches significant and substantial rates of violations in subjects' behavior. In demonstrating two non-trivial necessary conditions I make numerical computation of the optimal policy in smaller versions of the GL problem much easier. This advance, in turn, allows for the design of future experiments in which subjects' behavior can be compared against a fully rational benchmark- one which will include calculability of the exact costs of all deviations from the optimal search path.

My analysis of GL's data on the individual problem level reveals that nearly all problems fall into one of four surprisingly simple and tightly fitting search types. Violation rates of necessary conditions, as well as other statistics describing search, are shown to vary significantly and substantially across these types.

A common explanation for violations of three of my five necessary conditions for optimality, as well as several other seemingly unrelated observed search behaviors, is working memory limitations- though reading bias and time-saving considerations likely also play a confounded role. Future experimentation on multiple attribute search will allow for these three effects to be separated.

My analysis yields two general findings that cannot yet be reconciled. On the one hand, subjects clearly violate optimal search, assuming perfect reasoning (and memory), at high rates. On the other hand the GL search problem is clearly working memory load intensive, and subjects show that they are seriously affected by working memory limitations. Thus, the question becomes, "to what extent is the otherwise sub-optimal search behavior we observe an optimal response to working memory limitations?" I hope to answer this question as much as possible in future work, starting with a simple model of working memory load (Sanjurjo 2008b), and with further related experimentation.

This paper follows the lead of Newell and Simon (1971), Payne, Johnson, and Bettman (1993), Gabaix and Laibson (2006), and Crawford (2008) in taking a procedural cognition approach to study decision making. By using an experimental method that allows for the recording of both revealed preferences by way of choice, and the entire order and duration of each preceding step of information acquisition, a type of dual analysis is made possible that yields new insights into the cognitive processes actually underlying decision making in search. In this paper, these insights are not only useful in and of themselves, but they also lead to a study of the effects working memory load (Sanjurjo 2008b) on search strategy, thus choice. Insights gained from the procedural cognition approach can only serve to improve models of decision making. At worst they will provide additional information about behavior that will serve to refine experimental designs and assist theorists in the way they think about building models. At best, they will lead to the formation of general models of decision making increasingly rooted in well-defined cognitive primitives.

Appendix

Appendix A: Necessary Conditions 1,2, and 4, More Formally Stated

[28] Necessary conditions 4 and 6 are completely transparent, so are not included here.

Each realized attribute is defined as a_{rc}, where a is equal to the attribute value, r is the alternative row (1,…,8), and c is the attribute column (1,…,10). Note that $E(\tilde{a}_{rc}) = 0$, so unsearched attributes disappear in the summations below.

(1) "If an attribute in alternative x is searched, it must be the case that there does not exist an alternative y that is both weakly less searched than x, and has a weakly higher cumulative revealed value than x, with at least one of these inequalities strict."

$$\textit{If an attribute in alternative x is searched, then } \not\in \text{ y } s.t.\ a_{x1} + \sum_{c=2}^{10} a_{xc} \le a_{y1} + \sum_{c=2}^{10} a_{yc}$$
$$\textit{and .y is weakly less searched, with one of these inequalities strict.}$$

(2) "Within any alternative, an attribute cannot be searched if there remains another unsearched attribute with greater variance."

$$\textit{If } a_{rc} \textit{ is searched, for any } r \textit{ and } c, \textit{ then } \neg\exists\, p\ s.t\ a_{rp} \textit{ is unsearched and } Var(\tilde{a}_{rp}) > Var(\tilde{a}_{rc}).$$

(4) "Choose the alternative with highest revealed value."

$$\max_{\{r\}} C(r), \textit{where } C(r) = a_{r1} + \sum_{c=2}^{10} a_{rc}$$

Appendix B: Demonstration of Necessary Conditions 1 and 2 (In Construction-have proof sketch for condition 2)

In this section I demonstrate necessary condition two analytically (and will try to do the same for condition one). I now have a proof sketch for necessary condition two, which I include below, but do not yet have a proof for necessary condition one. If I am not able to demonstrate condition one analytically then I will use the method of numerical computation to show, that for a wide variety of simulated search problems, the condition holds. Necessary condition one will be demonstrated using a 3x3 version of the GL 8x10 search problem because it is the dimensionally smallest version of the problem that contains all of the fundamental features of the larger GL problem. Namely, having at least three attributes in multiple alternatives allows the searcher, after searching an attribute in an alternative, to then decide to stop search, search again in the same alternative, or switch to another alternative and search an attribute there. Having at least three alternatives breaks an interesting "indifference by symmetry" result that holds in the two alternative problem (as demonstrated in Appendix C). 3x3 is also the maximum size problem I consider here because, numerically, GL's search problem becomes intractable "quickly" as the dimensionality of alternatives and/or attributes increases. The expressions used to estimate expected values of each relevant search strategy are greatly simplified by similar numerical proofs, as well as analytical proofs, on versions of the GL problem smaller than 3x3. Appendix C contains several examples of these analytical proofs.

For the proof of necessary condition two assume the following (slightly more general than above): One of N alternatives must be chosen, where N is at least two. The value of any alternative is equal to the sum of an initially known cumulative revealed value (depending on preceding search) v_i (alternative i) and however many unknown attributes remain in that alternative. Within each alternative, these unknown attributes are independent, normally distributed, mean zero, and differ in variance. At least one of these alternatives contains at least two unrealized attributes. The cost of realizing the value of any unknown attribute is $c > 0$.

Theorem: Given the assumptions above, for any search policy that begins by searching a lower variance attribute within an alternative, there exists a search policy that begins by searching a higher variance attribute in the same alternative, that yields a higher expected payoff.

Proof Sketch:

First I need to prove lemma 1 (though it seems obvious), which will be used throughout to prove the theorem. Lemma 1 states that for any unique alternative (i) and attribute (j) combination, given any history of search (H), $\exists\, N_{ij} | H$ s.t. if $x_{ij} > N_{ij} | H$ then search stops and that attributes' corresponding alternative is chosen.

In the following analysis I compare the expected values of two search strategies, conditional on any preceding search history (identical for both strategies).

Consider a search policy that begins by searching a lower variance attribute in an alternative, call it alternative 1 (w.l.o.g.), then proceeds optimally. Call the lower variance attribute $\tilde{x}$ and the strategy $\gamma_x(\tilde{x})$.

Now consider a new strategy that begins by searching a higher variance attribute in alternative 1 instead. Call this attribute $\tilde{x}'$. For each realization x of $\tilde{x}$ there is a unique realization x' of $\tilde{x}'$ s.t. $p_x(\tilde{x} < x) = p_{x'}(\tilde{x}' < x')$. We will denote the correspondence x' to x by the function $h(x') = x$, and because $h(x')$ is invertible $h^{-1}(x) = x'$.

The new search policy strictly follows the decision rules of $\gamma_x(\tilde{x})$, using $h(x')$ in place of x, until it reaches x in search (if it does). If it does reach x then $\gamma_x(\tilde{x})$ would have reached x'. In this case, once x is realized, the new policy switches back to the original strategy ($\gamma_x(\tilde{x})$) from that point on, yielding the same expected payoffs for the remainder of search. Call the new search policy $\tilde{\gamma}_x(\tilde{x}')$.

Thus, the only way the expected payoffs of $\gamma_x(\tilde{x})$ and $\tilde{\gamma}_x(\tilde{x}')$ differ is if the originally searched alternative (alternative 1) is chosen before x is reached under $\tilde{\gamma}_x(\tilde{x}')$. Consider the differences in payoffs earned by each strategy in this case.
If $x > 0$ then $h^{-1}(x) = x' > 0$, and $h^{-1}(x) - x > 0$ (by definition of $h(x)$). Likewise, if $x < 0$ then $h^{-1}(x) = x' < 0$, and $h^{-1}(x) - x < 0$. These differences in payoffs are symmetric around $x = 0$, so $h^{-1}(x) - x = -\left(h^{-1}(-x) - (-x)\right)$. Now consider the set of combinations of attribute realizations that lead to the original alternative being chosen. Ceteris paribus, this set is strictly increasing in x, so the probability of stopping search to choose the originally searched alternative is also strictly increasing in x. Thus, if we define $f(x)$ as the probability of the originally searched alternative being chosen, then $\int_{-\infty}^{\infty} f(x)\left(h^{-1}(x)-x\right)dx > 0.$ □

Show results of numerical estimations here (if cannot get analytical proof) for necessary condition one (still need a 4-5 days (of work) to finish these, but I already have usable expressions. It's just a matter of pruning them down a bit further so I can get the most iterations, and richest estimations (more draws from the distributions) for the time I'll spend actually running them. I've already run, and confirmed all three necessary conditions on smaller, nested search problems (which I am using to prune the larger 3x3 expressions))

Appendix C: Demonstration of Necessary Conditions on Smaller and/or Simpler Versions Than 3x3 (In Construction)

This section contains several analytical results for necessary condition one, on versions of the GL search problem that are smaller and/or simpler than the 3x3 versions demonstrated in Appendix B. Many of these results are used in order to simplify the numerical expressions used for proofs in Appendix B, which greatly speeds up the necessary computations. **(Still need to clean up language, add a couple more proofs, and organize better)**

As in Appendix A, Each realized attribute is defined as a_{rc}, where a is equal to the attribute value, r is the alternative row where $r \in R, R = \{1,\ldots,\text{N}\}$, and c is the attribute column where $c \in C, C = \{1,\ldots,\text{M}\}$. As in GL's problem, the first column of attributes are fully observable (for free). Further, each unknown attribute is searched at a constant cost c>0. Agents are risk neutral and must choose one of the offered alternatives. The value of each alternative is equal to the sum of its attributes. Unsearched attributes are i.i.d. within columns, $E(\tilde{a}_{rc}) = 0$ (mean zero), and $E(\tilde{a}_{rc}) = E(-\tilde{a}_{rc})$ (symmetric). Also, without loss of generality, R is the set of alternatives $\{1,2,3,\ldots\}$ that one alternative is chosen from, and $a_{i1} > a_{j1}$ *for* $i < j$ (alternatives are ordered by the value of attribute 1). The variance of attributes within an alternative are also ordered, with $Var(a_{ri}) > Var(a_{rj}),\ for\ i > j$.

Finally, let $\gamma(a_{rc})$ be the expected value of a policy which searches a_{rc} now and then proceeds optimally.

Two trivial results follow from the assumptions given above. First, if $C = \{1\}$ choose a_{11} for any R- if the value of each alternative is fully determined by attribute 1, alternative 1 should be chosen immediately. The second is that if $R = \{1\}$ no search should occur, for any C. Alternative 1 should be chosen immediately.

The simplest nontrivial search comparison occurs when there are two alternatives, two attributes for each alternative, and only one attribute can be searched. Result one includes this case, while extending beyond it. Results two and three allow as many searches as there are attributes.

Result 1: *If* $\{1,2\} \subset R$ *and* $\{1,2\} \subset C$, *where* $C = \{1,2,...,n\}$ *and only* 1 *attribute can be searched*, *then* $\gamma(a_{1c}) = \gamma(a_{2c}) > \gamma(a_{3c}) > \gamma(a_{4c}) > ... > \gamma(a_{nc})$ *for any c.*

Proof:

Equality:

$$\gamma(a_{1c}) = E_{a_{1c}}[\max\{a_{11} + \tilde{a}_{1c} - c, a_{21} - c\}]$$
$$= E_{a_{2c}}[\max\{a_{11} + \tilde{a}_{2c} - c, a_{21} - c\}] \text{ (by symmetry)}$$
$$= E_{a_{2c}}[\max\{a_{11} - c, a_{21} - \tilde{a}_{2c} - c\}] \text{ (by translation)}$$
$$= E_{a_{2c}}[\max\{a_{11} - c, a_{21} + \tilde{a}_{2c} - c\}] \text{ (by symmetry)}$$
$$= \gamma(a_{2c}) \square$$

Inequality:

$$\gamma(a_{2c}) = E_{a_{2c}}[\max\{a_{11} - c, a_{21} + \tilde{a}_{2c} - c\}]$$
$$> E_{a_{2c}}[\max\{a_{11} - c, a_{31} + \tilde{a}_{2c} - c\}]$$
$$= E_{a_{3c}}[\max\{a_{11} - c, a_{31} + \tilde{a}_{3c} - c\}] \text{ (by symmetry)}$$
$$= \gamma(a_{3c})$$

Now, assume $\gamma(a_{ic}) > \gamma(a_{(i+1)c})$ for $i \geq 2$, then

$$\gamma(a_{ic}) = E_{a_{ic}}[\max\{a_{i1} + \tilde{a}_{ic} - c, a_{11} - c\}]$$
$$> E_{a_{(i+1)c}}[\max\{a_{(i+1)1} + \tilde{a}_{(i+1)c} - c, a_{11} - c\}]$$
$$> E_{a_{(i+1)c}}[\max\{a_{(i+2)1} + \tilde{a}_{(i+1)c} - c, a_{11} - c\}]$$ [29]
$$= E_{a_{(i+2)c}}[\max\{a_{(i+2)1} + \tilde{a}_{(i+2)c} - c, a_{11} - c\}] \text{ (by symmetry)}$$
$$= \gamma(a_{(i+2)c})$$

So the inequality result holds by induction. □

Result 2: *If* $R = \{1,2\}$ *and* $C = \{1,2\}$ *and at most* 2 *attributes can be searched*, *then* $\gamma(a_{12}) = \gamma(a_{22})$.

Proof:

$$\gamma(a_{12}) = E_{a_{12}}[\max\{a_{11} + \tilde{a}_{12} - c, a_{21} - c, E_{a_{22}}[\max\{a_{11} + \tilde{a}_{12} - 2c, a_{21} + \tilde{a}_{22} - 2c\}]\}]$$
$$= E_{a_{22}}[\max\{a_{11} + \tilde{a}_{22} - c, a_{21} - c, E_{a_{12}}[\max\{a_{11} + \tilde{a}_{22} - 2c, a_{21} + \tilde{a}_{12} - 2c\}]\}] \text{ (by symmetry)}$$
$$= E_{a_{22}}[\max\{a_{11} - c, a_{21} - \tilde{a}_{22} - c, E_{a_{12}}[\max\{a_{11} - \tilde{a}_{12} - 2c, a_{21} - \tilde{a}_{22} - 2c\}]\}] \text{ (by translation)}$$
$$= E_{a_{22}}[\max\{a_{11} - c, a_{21} + \tilde{a}_{22} - c, E_{a_{12}}[\max\{a_{11} + \tilde{a}_{12} - 2c, a_{21} + \tilde{a}_{22} - 2c\}]\}] \text{ (by symmetry)}$$
$$= \gamma(a_{22}) \square$$

---

[29] Thanks to Patrick Fitzsimmons for suggesting this method of proof. Integrating out each expectation, and comparing, also works.

Result 3: Result 2 *holds for* $R = \{1,2\}$ *and for any* C *s.t.* $\{1,2\} \subset C \subset \square$ *when at most* d *attributes can be searched, for any* $d \in [1, 2C]$.

Proof:

In the case of 2 alternatives, the expected value of first searching an attribute in one alternative can always, by construction, be identically reproduced by transforming the expected value of first searching the other with the same 3-step transformation used in the proof of Result 2 (symmetry result).

The proven equalities in Results one through three at first appear counterintuitive. Indeed the intuition behind costly search (for economic goods) is to search until one alternative has sufficiently separated itself above the others, so that search may be stopped and that alternative may be chosen. Thus, it seems intuitive that searching alternative 1 first should generate a larger expected value than searching alternative 2 first. However, the equality results exist due to a fundamental symmetry in information revelation. If a draw of x from an attribute in alternative 1 is sufficiently large as to induce the termination of search and a choice of alternative 1, then a draw of –x from an attribute in alternative 2 yields the exact same result.

Interestingly, Result one shows that any truly myopic searcher, who treats each searched attribute as the last, is always indifferent between searching either the alternative with the highest or second highest attribute one value, given that the two alternatives are equally unsearched. Thus, models such as GL's directed cognition model, require a tie-breaking assumption for the many cases in which this type of search decision is made throughout the course of a task.

Appendix D: Econometric Specifications

This section shows the systems of equations used to estimate conditional logit coefficients in Section IV. Three different sets of coefficients are estimated for each subject: one for each of the first three alternative transitions.

For regressions, by individual subject, on the first alternative searched:

$$\ln\left(\frac{P_{it1}}{P_{it8}}\right) = \beta_{vi}V_{it1} + \beta_{Hi}H_{it1} + \varepsilon_{it1},$$

…,

$$\ln\left(\frac{P_{it7}}{P_{it8}}\right) = \beta_{vi}V_{it7} + \beta_{Hi}H_{it7} + \varepsilon_{it7}$$

where i=individual subject, i $\in I$ *where* $I = \{1,2,3,...,390\}$

t=task, t $\in T$ *where* $T = \{1,2,3,....\}$

$P \equiv$ 1 if alternative is searched, 0 otherwise

$V \equiv$ Value of Attribute 1 for given alternative, V N$(0,\sigma^2)$

$H \equiv$ Spatial height of Alternative (1-top,..., 8-bottom)

The same specified system of equations is used for regressions, by individual subject, for switches from the first alternative searched to the second, and for switches from the second to the third.

$$\ln\left(\frac{P_{it1}}{P_{it8}}\right) = \beta_{vi}V_{it1} + \beta_{Hi}H_{it1} + \beta_{Di}D_{it1} + \beta_{Ii}I_{it1} + \varepsilon_{it1},$$

…,

$$\ln\left(\frac{P_{it7}}{P_{it8}}\right) = \beta_{vi}V_{it7} + \beta_{Hi}H_{it7} + \beta_{Di}D_{it7} + \beta_{Ii}I_{it7} + \varepsilon_{it7}$$

where i=individual subject, i $\in I$ *where* $I = \{1,2,3,...,390\}$

t=task, t $\in T$ *where* $T = \{1,2,3,....\}$

$P \equiv$ 1 if alternative is searched, 0 otherwise

$V \equiv$ Value of Attribute 1 of alternative, V N$(0,\sigma^2)$

$H \equiv$ Spatial height of Alternative (1-top,..., 8-bottom)

D $\equiv$ Absolute distance of searched alternative from previously searched alternative, D $\in \{1,2,3,...,7\}$

I $\equiv$ 1 if alternative is below previously searched alternative, 0 otherwise

References

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

Camerer, Colin (1995): “Individual Decision Making,” in Alvin Roth and John Kagel, editors, *The Handbook of Experimental Economics,* Princeton University Press, Ch. 8.

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.

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.

Lim, C., Bearden, Neil, & Smith, C. (2005): "Sequential Search with Multiattribute Options," *Decision Analysis, 3,* 3-15.

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

Newell, Allen, and Herbert Simon (1972): *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Payne, John (1976): "Task Complexity and Contingent Processing in Decision-Making – Information Search and Protocol Analysis," *Organizational Behavior and Human Performance*,16, 366-387.

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

Payne, John, James Bettman, and Eric Johnson (1988): "Adaptive strategy selection in Decision Making," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 534-552.

Rabin, Matthew (2000): "Risk Aversion and Expected-Utility Theory: A Calibration Theorem," *Econometrica,* 68, 1281-1292.

Sanjurjo, Adam (2008b): A Characterization of Working Memory Load in Multiple Attribute, Multiple Alternative Search," In Progress.

Sanjurjo, Adam (2008c): The Effects of Variation in Time Pressure and Stakes in Multiple Attribute, Multiple Alternative Search," In Progress.

Stigler, George (1961): "The Economics of Information," *The Journal of Political Economy*, 69, 213-225.

Tversky, Amos (1972): "Elimination by Aspects: A Theory of Choice," *Psychological Review*, 79, 281-299.

Weitzman, Martin (1979): "Optimal Search for the Best Alternative," *Econometrica*, 47, 641-654.

This chapter is currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this material.

Can Working Memory Limits Explain the Fact Patterns in Multiple Attribute Search?

Adam Sanjurjo[1]
University of California, San Diego
8 Sept 2009

Abstract: In Sanjurjo (2008) I provide a partial characterization of optimal search in Gabaix and Laibson's (2006) multiple attribute/multiple alternative search experiment. My analysis reveals three puzzling patterns: (1) subjects often switch to searching clearly less promising alternatives, (2) search too deeply within alternatives, and (3) exhibit a strong scale effect in relying more heavily on within alternative intensive search (as opposed to across) here than they do in isomorphic experiments with fewer alternatives and attributes. In this paper I demonstrate that all three of these puzzles systematically reduce working memory load (WML), which suggests that a simple model in which working memory is limited but subjects otherwise behave optimally can explain all three patterns. My results demonstrate that more systematic search sequences require less WML (explains (1) and (2)) and that a within alternative intensive search sequence is the unique minimum WML sequence (explains (3)). By operationalizing WML I illustrate a fundamental component of information overload and complexity, provide a structural justification for peoples' use of heuristic problem solving strategies, and reveal various forms of strategic manipulation possible through design effects.

[1] Department of Economics, University of California, San Diego. 9500 Gilman Drive, La Jolla, CA 92093-0508 (email: asanjurjo@ucsd.edu). I thank John Conlisk, Richard Carson, Patrick Fitzsimmons, Hal Pashler, Angel Sanjurjo, David Schkade, and Joel Sobel for helpful comments, Xavier Gabaix and David Laibson for generously sharing their dataset, and to Vincent Crawford for invaluable guidance throughout the entire process.

Introduction

Optimal search policies have been fully characterized for a wide variety of search problems 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 these analyses yield substantial insights, many important applications have alternatives whose values are determined by multiple attributes: we consider more than wage when choosing a job, and more than price when purchasing a home.

Gabaix and Laibson (2006; henceforth "GL") study search with multiple attributes (ten) and alternatives (eight) experimentally, with full recall and no order restrictions. Each of their subjects faces a series of search problems in which search is selective due to exogenously imposed time limits. GL's analysis is of particular interest because the richness of their search environment approximates human cognition in less structured settings more closely than most other models, and because their experimental interface bears a close family resemblance to the kind of information displays commonly used in internet commerce.[1]

Although optimal search would be a natural benchmark with which to compare subjects' behavior, GL find that the high dimensionality of their problem makes characterizing optimal search analytically and numerically intractable (p. 1066). Instead they focus on comparing their "Directed Cognition" model of search, which is myopic in

[1] This display is similar to those found for consumer products from cellular phones, to automobiles, to homes, to Medicare part D, and even for information tables in academic papers.

that it ignores option value, but otherwise fully rational, with “naïve” heuristics taken from the psychology literature (Tversky, 1972).

In Sanjurjo (2008) I build on GL’s work in three ways. First, I conduct a detailed analysis of GL’s search data on the individual level. Second, although a full characterization of optimal search (assuming perfect cognitive abilities) in GL’s environment does indeed appear to be intractable, I give a partial characterization in the form of several necessary (but not sufficient) conditions for optimality. Third, I compare subjects’ behavior to this partial rational benchmark.

I find that while subjects’ search behavior is largely uncontingent (independent of previously revealed information) it is also highly systematic. 98% of all subjects’ search behavior can be described by two archetypal search patterns- within alternative (row) transition intensive (ALT; 92%) and within attribute (column) transition intensive (ATT; 6%).[2] Remarkably, the ratio of GL’s subjects’ within alternative to within attribute transitions is thirteen times as large as it is in isomorphic experiments with fewer alternatives and attributes (Payne et. all, 1988).[3] This surprising scale effect is puzzle #1.

My necessary condition on the optimal tradeoff between the depth and breadth of search across alternatives and attributes allows me to identify instances of conditional over-search and under-search within alternatives, which include violations that occur both when the searcher switches alternatives to a wrong alternative, and when the searcher continues to search in the same alternative (too deeply). I find that this

[2] I define both ALT and ATT clearly in Section II.
[3] The key difference between the PBJ (four by four) and the GL (eight by ten) experiments is that the object of choice in PBJ is a lottery whereas it is an additive sum in GL. If assuming risk neutrality this reduces to the difference between a weighted average and a rescaled weighted average. Aside from this difference the design uses a virtually identical MouseLab software interface.

necessary condition for optimality is violated often- in 50% of all subjects' search actions. In 39% of all search actions subjects search conditionally too deeply within an alternative (puzzle #2), and in 49% of all alternative transitions, subjects make the costly error of transitioning to a "wrong" alternative (puzzle #3). To my knowledge there is no existing theory that can explain these systematic deviations from optimality.

A searcher with no cognitive limitations normally does best searching contingently (Kohn & Shavell, 1974; Stigler, 1961), and trades off depth of search within an alternative and breadth of search across alternatives optimally. Here I explore the possibility of explaining both of subjects' main systematic deviations from optimal search, as well as the puzzling scale effect, via the single hypothesis that they have limited working memory.[4]

I demonstrate, in three theoretical results, that all three of subjects' puzzling behaviors systematically reduce working memory load (WML), which suggests that a simple model in which working memory is limited but subjects otherwise behave optimally can explain all three puzzles.

My first result demonstrates that for any rectangular information display with at least two alternatives and two attributes, a within alternative transition intensive (ALT) search sequence is the unique minimum WML sequence. Because GL's subjects display a mixture of contingent and uncontingent search behavior I demonstrate my result under both extreme conditions. Under uncontingent search I show that ALT is the unique

[4] After two of my other necessary conditions for optimality demonstrated unambiguous working memory failure- 11% of all look-ups are repeated, and 33% of final alternative choices were incorrect (based on previously observed information)- it was natural to question whether working memory limitations, a fundamental bottleneck in human problem solving (Newell and Simon, 1972), might also explain the three empirical puzzles just described.

undominated WML search sequence. My definition of WML dominance is like Pareto dominance; one sequence WML dominates another if it everywhere requires weakly less WML, and requires strictly less WML for at least one step in the search sequence. For contingent search I demonstrate the slightly weaker result that for a rectangular information matrix with at least two alternatives and two attributes ALT is the unique minimum aggregate WML search sequence, where aggregate refers to all steps in the search sequence.

For both contingent and uncontingent search I also show that the differences in the aggregate WML requirements between the ALT and ATT search sequences increase quadratically, while the differences in the maximum WML of the search sequences increase linearly, with the size of the information matrix.[5]

Thus, my first result provides a structural explanation for the tendency of subjects to systematically shift to a higher ratio of within alternative to within attribute transitions as the dimensions of the information display increase, for subjects searching too deeply within alternatives (I will explain how in Section II), and is also generally consistent with 92% of all subjects' search behavior being ALT.

My second result simply establishes that, under reasonable assumptions, for any given search sequence uncontingent search WML dominates contingent search. This result provides concrete evidence of the non-trivial tradeoff between searching contingently, and searching uncontingently in a way that reduces WML.

[5] Whether the aggregate or maximum WML of a sequence has a greater affect on decision making is an empirical question that requires more attention. Aggregate WML would seem to matter more if the costs of storing WML is the primary issue; maximum WML would seem to matter more if the primary issue is increased error rate due to exceeding one's instantaneous WML threshold.

For my third result I enrich my model of WML to more fully capture the presence of systematicity in uncontingent search sequences. I demonstrate the general and highly intuitive notion that more systematic search rules require less WML than less systematic search rules (ones that look as if they're generated from a random process). By more systematic, I specifically mean that a search rule requires comparatively fewer unique instructions in order to be executed correctly. It follows from this result that a "refined" (specific) version of ALT is the unique undominated uncontingent search sequence- so a refined version of my first result still holds under this enriched model of WML.[6]

This, my third result, provides a structural explanation for the two main systematic deviations from optimal search found in the GL dataset. Subjects' costly tendency to disproportionately favor adjacent alternatives when transitioning from one alternative to another increases the degree of systematicity in their search sequences. Similarly, "over-searching" in a particular alternative is part of GL subjects' "one time through" approach that makes search more systematic and thus reduces WML relative to a sequence in which the searcher instead switches back and forth between partially searched alternatives.[7] Further, this result also suggests an explanation for why ATT search is more common in GL subjects' behavior than otherwise lower WML search sequences (ATT is weakly WML dominated by all search sequences in the non-enriched model of WML)- because the maximum systematicity of the sequence itself is now acknowledged. This result is therefore also consistent with 98% of all search behavior in

---

[6] In section III I will explain why whether a version of the second main result still holds depends on what specific assumptions are made.

[7] As mentioned, this type of "over-searching" within alternatives is also consistent with the first main result, which is based on the intuition that minimum WML search sequences have the fewest possible number of alternatives open to further search at any point in the search sequence.

the GL dataset being either ALT or ATT, and 96% of all within alternative or attribute search transitions being adjacent.

Aside from my theoretical results themselves, I demonstrate that working memory load for different decision-making approaches can be clearly defined and measured. In fact, I provide formulas that can be used as diagnostic tools to compute the WML required for different search sequences in any real world multiple alternative/multiple attribute information display, step by step, yielding summary statistics such as the maximum and aggregate WML's of these search sequences.[8]

This paper is motivated by the well established finding that working memory capacity is limited (Miller, 1956; Newell & Simon, 1972) and plays a crucial role in decision making (Cowen 2001), adversely affecting reasoning abilities. Everyday events that corroborate with these findings are the difficulty we might experience in attempting to remember a ten digit number or to work out the product of two three digit numbers in our heads. Notions of Complexity, Information overload, errors in decision making, and selection of simplifying decision heuristics all in fact seem to be closely linked to (small) finite WM capacity.

This paper most closely relates to the process-tracing literature which began with the work of Newell and Simon (1972) and continued with the work of Payne, Bettman, and Johnson (PBJ; 1993), GL (2006), and Crawford (2008), among others.

It differs fundamentally from a class of other papers in the Economics Literature that address memory limitations in decision making, such as Bernheim and Thomadsen

[8] More abstract notions such as "Cognitive costs," on the other hand, are not well defined so they are difficult to measure.

(2005), Piccione and Rubenstein (2003), Wilson (2003), Benabou and Tirole (2002), and Mullainathan (2002), to name just a few, in that here I take a more primitive approach to memory. Whereas the models in these mentioned papers each assume some specific memory technology, then proceed with an analysis of optimal decision-making subject to that technology, my theoretical results come from simply computing and comparing the number of representations that, logically, must be stored in working memory as a function of different information search strategies. By operationalizing a fundamental component of complexity and information overload, and providing a clear, measureable, justification for the use of heuristic problem solving strategies in general, this paper is also closely related to those three literatures as well.

In Section II I present my first two results, embedded in a series of examples meant to provide necessary intuition. In Section III I present my third result. Following the Conclusion in Section IV is an appendix containing proofs of all Propositions presented in Sections II and III, formal definitions of ALT and ATT search, and formulas for the maximum and aggregate WML's of ALT and ATT search.

II. The WML minimizing search sequence (with illustrative example), and the WML benefits of uncontingent search

In this section I prove my first main result; for any rectangular information display with at least two alternatives and two attributes, a within alternative transition intensive (ALT) search sequence is the unique minimum WML sequence. Because GL's subjects display a mixture of contingent and uncontingent search behavior I demonstrate my result under both extreme conditions. Under uncontingent search I show that ALT is the unique undominated WML search sequence. My definition of WML dominance is like Pareto dominance; one sequence WML dominates another if it everywhere requires weakly less WML, and requires strictly less WML for at least one step in the search sequence. For contingent search I demonstrate the slightly weaker result that ALT is the unique minimum aggregate WML search sequence, where aggregate refers to all steps in the search sequence. If subjects use a combination of contingent and uncontingent sub-sequences within the same sequence, the same results hold locally over-subsequences, so mixture models should be in line with the results reported here.

My first main result holds under my benchmark specification of total WML, which is simply the computed total WML logically required for each step of any search sequence- the sum of all necessary value and location representations. For example, if the value of each alternative is the sum of its attribute values (as in GL's design), then running sums of each alternative must be held as they are being searched, as well as the spatial location of each running sum.

In order to provide additional intuition for my first main result I prove the result for each component of total WML separately- the WML of values and the WML of locations. Also, I precede each general result with a specific example- comparing the WML's required for ALT and for ATT- the two archetypal search sequences, which together describe 98% of all search behavior in GL's data. In the examples I use an information display with four alternatives and four attributes, as in Payne et. al (1988).

For both contingent and uncontingent search, and for the WML's of both values and locations, I also show that the differences in the aggregate WML requirements between the ALT and ATT search sequences increase quadratically, while the differences in the maximum WML of the search sequences increase linearly, with the size of the information matrix.

First I define archetypal ALT and ATT search sequences, which will be used in examples provided throughout the section. ALT search begins in any alternative (which I will assume, as in GL, will always be a row) of the information display. It sequentially searches previously unsearched attributes (in any order) in that alternative until all have been searched once. It then transitions to any other previously unsearched alternative and repeats the instructions from the previous sentence. This process iterates until all alternatives have been searched exhaustively. ATT search is the same, but with alternatives replaced by attribute columns in the definition of ALT (ATT is ALT rotated 90 degrees). Clearly, ALT and ATT are actually relatively large classes of search sequences. In the following example I use the most systematic version of each, where ALT proceeds left-to-right adjacent within alternatives and top-to-bottom adjacent across alternatives- as in the way the text on this page is read. ATT proceeds top-to-bottom adjacent within attribute columns and left-to-right adjacent across attribute columns.[9]

In each example I report the WML required for each step of the search sequence in its corresponding position of the information display. For each case (uncontingent or contingent, WML of values or locations) I also provide the appropriate formulas for maximum and aggregate WML's of the search sequence, for an information display of any dimensions, in the appendix. To explain a few bits of necessary notation: t is a counter of the step in a given search sequence, M is the number of alternatives in the

---

[9] In this section all versions of ALT search yield the same WML under the assumption that uncontingent search sequences are "hard-wired," thus costless. However, when this assumption is relaxed in Section III I will show that this maximally systematic version of ALT is the unique undominated search sequence-WML dominating all other (less systematic) versions of ALT.

information display, N is the number of attribute columns, WML^v is the WML of values, WML^l is the WML of locations, and WML^T is total WML ($WML^v + WML^l$).

After presenting the result for both uncontingent and contingent search I discuss its implications, and in particular, its ability to provide a structural explanation for two previously unexplained essential features of GL's subjects' search behavior.

Uncontingent Search:

Search is uncontingent if current and future search actions are independent of previously observed information, as in Stigler's (1961) classic search. Because GL's subjects systematically violate contingent search, and opt for highly systematic uncontingent sequences while doing so, I compute the WML's required for different uncontingent search sequences and compare them. The following two sub-sections show that ALT is the unique undominated search sequence in terms of both the WML of values and locations, thus total WML- my first result for uncontingent search.

WML of Values:

0	1	1	1
1	2	2	2
1	2	2	2
1	2	2	2

Figure 3.1.1: **<u>ALT</u>**

$\max_{\{t\}} \{WML^v(t \mid ALT)\} = 2 = 2$

$\sum_{t=1}^{MN} WML^v(t \mid ALT) = 2MN\text{-}N\text{-}M = 24$

0	4	4	4
1	4	4	4
2	4	4	3
3	4	4	2

Figure 3.1.2: **<u>ATT</u>**

$\max_{\{t\}} \{WML^v(t \mid ATT)\} = M = 4$

$\sum_{t=1}^{MN} WML^v(t \mid ATT) = M^2(N-1)+(M-1)=51$

For the particular ALT sequence depicted in Figure 3.1.1 search starts in the left-most attribute of the top alternative. Initially, zero WML^v is required because the value of the first searched attribute is visible. However, once search proceeds one attribute to the right, now the left-most attribute value must be remembered because it is no longer visible.[10] As this second attribute is being observed its value is added to the remembered value of the left-most attribute, then search proceeds one attribute to the right, but now with the single summed value of the first two searched attributes remembered; this process iterates to the right-most attribute in the top alternative.[11] When the left-most attribute of the second from the top alternative is next searched the summed value of the top alternative must be remembered. Search then transitions one attribute to the right, and now the value of the attribute just searched must be remembered in addition to the value of the top alternative, and so on. Crucially, regardless of how many alternatives have been completely searched, only one completed alternative value need be recalled- the highest.

The maximum WML^v for any moment in ALT search sequence is two, and this result is general for arbitrarily large numbers of alternatives and attribute columns. The aggregate WML^v, in general, is 2MN-N-M, which in the four by four is 24.

For the particular ATT sequence depicted in Figure 3.1.2, search also starts in the left-most attribute of the top alternative. Search then proceeds one attribute down. Now that the first searched attribute value is no longer observed, it must be remembered. Search then again transitions one attribute down, and now each of the two previously

[10] In GL's design only one attribute is visible at a time. In an environment where all attributes are visible at once my approach models a human being that can attend to at most one attribute in any given instant.
[11] For simplicity of exposition I stick to the GL design, but my results are general to a broader class of multiple alternative/multiple attribute information displays. For example, the WML of values reported here would also hold for any form of within alternative attribute integration, not just summing.

searched attribute values must be remembered. When the search sequence arrives to the top attribute of the second from the left attribute column four values must be remembered. This WML^v persists until the second to last searched attribute in the ATT sequence, where two alternatives have now been searched completely, so the lower completed alternative value can be discarded, which means a WML^v of three rather than four, and for the last searched attribute, two rather than three.

The key intuition for the differences in WML^v between the ALT and ATT sequences is that WML^v is equal to the number of alternatives that have been partially, but not exhaustively searched. If this is true of an alternative I say that it is "open." At each step of ALT search the least possible number of alternatives are open, and for each step of ATT search the maximum number of alternatives are open.

The maximum WML^v for any step in the ATT search sequence is M, in general, thus four in the four by four sized information display. The aggregate WML^v is $M^2(N-1)+(M-1)$, in general, thus 51 in the four by four display.

General result:

Proposition UV: ALT is the unique undominated WML^v search sequence for $M, N \geq 2$.[12]

In the appendix I also show that the difference in maximum WML^v between the ATT and ALT search sequences increases linearly in M- the number of alternatives, while the difference in aggregate WML^v between ATT and ALT increases linearly in N and quadratically in M.

[12] Proofs for all Propositions are contained in the appendix.

WML of Locations:

0	**0**	**0**	**0**
0	**1**	**1**	**1**
1	**1**	**1**	**1**
1	**1**	**1**	**1**

Figure 3.2.1: <u>ALT</u>

$$\max_{\{t\}}\{\mathrm{WML}^{l}(t \mid \mathrm{ALT})\} = 1 = 1$$

$$\sum_{t=1}^{MN} \mathrm{WML}^{l}(t \mid \mathrm{ALT}) = (M\text{-}1)N\text{-}1 = 11$$

0	**3**	**3**	**3**
0	**3**	**3**	**3**
1	**3**	**3**	**2**
2	**3**	**3**	**1**

Figure 3.2.2: <u>ATT</u>

$$\max_{\{t\}}\{\mathrm{WML}^{l}(t \mid \mathrm{ATT})\} = M\text{-}1 = 3$$

$$\sum_{t=1}^{MN} \mathrm{WML}^{l}(t \mid \mathrm{ATT}) = (N\text{-}1)(M\text{-}1)M = 36$$

Because the search sequence is uncontingent, for ALT (Figure 3.2.1), the searcher need not remember any locations until there are two distinct alternative values to be recalled simultaneously.[13] This occurs for the first time as she searches the sixth attribute in the sequence. Remembering which alternative corresponds to one of the two values is sufficient for her to know which alternative the other value belongs to.[14]

For ATT (Figure 3.2.2), The WML of locations (WML^{l}) is similar to its WML^{v} (see Figure 3.1.2), but systematically one unit lower due to a searcher's ability to logically identify the values associated with each of four alternatives by remembering the locations of only three of those values, for example.

The basic intuition regarding the WML^{l}'s of ALT and ATT is the same as that for WML^{v}; ALT has the lowest possible number of alternatives open for each step of the search sequence, while ATT has the most.

[13] Uncontingent search can be thought of as predetermined, thus by knowing where she is currently searching, the searcher knows where she has already searched.
[14] Throughout, I assume that searchers only store necessary information.

General result:

<u>Proposition UL</u>: ALT is the unique undominated WML^l search sequence for M, N ≥ 2.

In the appendix I also show that the difference in maximum WML^l between the ATT and ALT search sequences increases linearly in M- the number of alternatives, while the difference in aggregate WML^l between ATT and ALT increases linearly in N and quadratically in M.

Total WML($WML^v + WML^l$**):**

0	**1**	**1**	**1**
1	**3**	**3**	**3**
2	**3**	**3**	**3**
2	**3**	**3**	**3**

Figure 3.3.1: <u>ALT</u>

$\max_{\{t\}} \{WML^T(t \mid ALT)\} = 3 = 3$

$\sum_{t=1}^{MN} WML^T(t \mid ALT) = 3MN\text{-}2N\text{-}M\text{-}1 = 35$

0	**7**	**7**	**7**
1	**7**	**7**	**7**
3	**7**	**7**	**5**
5	**7**	**7**	**3**

Figure 3.3.2: <u>ATT</u>

$\max_{\{t\}} \{WML^T(t \mid ATT)\} = 2M\text{-}1 = 7$

$\sum_{t=1}^{MN} WML^T(t \mid ATT) = 2M^2(N\text{-}1)+M(2\text{-}N) = 87$

Here I show the total WML ($WML^T = WML^v + WML^l$) for uncontingent search patterns ALT (Figure 3.3.1) and ATT (Figure 3.3.2) combining the WML's required for both values and locations. This is the first main result for uncontingent search.

General result:

<u>Proposition 1UT</u>: ALT is the unique undominated WML^T search sequence for M, N ≥ 2.

In the appendix I also show that the difference in maximum WML^T between the ATT and ALT search sequences increases linearly in M- the number of alternatives- with a slope of two, and the difference in aggregate WML^T between ATT and ALT increases linearly in N and quadratically in M, twice as fast as for WML^v or WML^l alone.

Contingent Search:

Contingent search is optimal, basing each current action on all previously observed information. Because real searchers in the GL data use some mixture of contingent and uncontingent search it is important to compute the WML implications under both extremes.[15] In this sub-section I show that for information matrices with at least two alternatives and two attribute columns ALT is the unique minimum aggregate WML^T search sequence. Thus the results for contingent search are qualitatively similar to those for uncontingent search, though slightly weaker.[16] The results for the WML of values are identical under the uncontingent and contingent conditions. This is because the same values must be remembered for the same sequence of search, regardless of whether it occurred in a contingent or uncontingent fashion. Though it is unlikely that the archetypal search sequences ALT or ATT would result from contingent search, it is important to draw attention to the relative increases in WML that occur when one strays from ALT search, allowing the relative "randomness" of contingent search to dictate their search sequence. By understanding these dynamics a clear explanation for subjects'

[15] If subjects use a combination of contingent and uncontingent sub-sequences within the same sequence, the same results hold locally over-subsequences, so mixture models should provide no theoretical surprises.
[16] That a search sequence is uniquely minimizing in aggregate WML follows logically from it being the unique undominated WML sequence.

tendencies to stay behaviorally "close" to the uncontingent ALT search sequence emerges. The analysis for the WML of locations here is different than it was in uncontingent search, in that in addition to remembering the locations of different running alternative values, one must also remember where one has searched- in order to avoid repeating previous searches, or missing other attributes altogether- either of which will bias the choice of alternative.

WML of Values:

0	**1**	**1**	**1**
1	**2**	**2**	**2**
1	**2**	**2**	**2**
1	**2**	**2**	**2**

Figure 3.4.1: ALT

0	**4**	**4**	**4**
1	**4**	**4**	**4**
2	**4**	**4**	**3**
3	**4**	**4**	**2**

Figure 3.4.2: ATT

$$\max_{\{t\}} \{WML^v(t \mid ALT)\} = 2 = 2$$

$$\max_{\{t\}} \{WML^v(t \mid ATT)\} = M = 4$$

$$\sum_{t=1}^{MN} WML^v(t \mid ALT) = 2MN\text{-}N\text{-}M = 24$$

$$\sum_{t=1}^{MN} WML^v(t \mid ATT) = M^2(N\text{-}1)+(M\text{-}1) = 51$$

WML^v is identical under uncontingent and contingent search.

General result:

Proposition CV: ALT is the unique undominated WML^v search sequence for $M, N \geq 2$.

In the appendix I also show that the difference in maximum WML^v between the ATT and ALT search sequences increases linearly in M- the number of alternatives, and the difference in aggregate WML^v between ATT and ALT increases linearly in N and quadratically in M.

Locations WML:

Now that search is contingent, a searcher must remember not only the locations of the running values of alternatives, but also the locations of the attributes she has already searched. In order to compute precisely how much WML is required to remember the locations of searched attributes, an assumption must be made about the "chunking" (Miller, 1956) capabilities of searchers. Zero chunking capabilities would mean that one unit of memory is required for each searched attribute. Indefinite chunking capabilities would lead to just one memory unit being required for any pattern of previous search. An additional issue to consider with contingent search is whether searchers have the ability to switch from remembering what they have seen to what they have not seen, when it reduces WML to do so.

In this section I assume that subjects can only remember what they have seen and that they can chunk entire rows or columns as one memory unit, but no other combinations of attributes can be chunked. Changing these two assumptions in any reasonable way does not qualitatively change the results that follow.[17] It is important to note, however, that taking either extreme stance on chunking would allow for the stronger theoretical result from the uncontingent section- that ALT is the unique WML undominated search sequence- to go through. In the following sub-section I discuss the chunking issue a bit further.

---

[17] Assuming either extreme- zero chunking or perfect chunking lead to the seemingly unrealistic result that no previous search sequence is easier or more difficult to recall than any other (they all require the same amount of WML). One would anticipate that more systematic previous search sequences are easier to chunk than ones that look like they occurred due to some random process. Here I assume a chunking technology somewhere in between the extremes, where more systematicity in search does reduce WML.

WML of Locations:

$$\begin{array}{|c|c|c|c|}\hline 0&1&2&3\\\hline 1&2&3&4\\\hline 2&3&4&5\\\hline 3&4&5&6\\\hline\end{array} + \begin{array}{|c|c|c|c|}\hline 0&0&0&0\\\hline 0&1&1&1\\\hline 1&1&1&1\\\hline 1&1&1&1\\\hline\end{array} = \begin{array}{|c|c|c|c|}\hline 0&1&2&3\\\hline 1&3&4&5\\\hline 3&4&5&6\\\hline 4&5&6&7\\\hline\end{array}$$

Figure 3.5.1: <u>ALT</u>

$$\begin{array}{|c|c|c|c|}\hline 0&1&2&3\\\hline 1&2&3&4\\\hline 2&3&4&5\\\hline 3&4&5&6\\\hline\end{array} + \begin{array}{|c|c|c|c|}\hline 0&3&3&3\\\hline 0&3&3&3\\\hline 1&3&3&2\\\hline 2&3&3&1\\\hline\end{array} = \begin{array}{|c|c|c|c|}\hline 0&4&5&6\\\hline 1&5&6&7\\\hline 3&6&6&7\\\hline 5&7&8&7\\\hline\end{array}$$

Figure 3.5.2: <u>ATT</u>

$\max_{\{t\}}\{WML^{v}(t \mid ALT)\} = (M-1)+(N-1)+1 = 7$

$\max_{\{t\}}\{WML^{v}(t \mid ATT)\} = N+2M-4 = 8$

$\sum_{t=1}^{MN} WML^{v}(t \mid ALT) = M(N-1)N/2 + N(M-1)M/2+(M-1)N-1=59$,

$\sum_{t=1}^{MN} WML^{v}(t \mid ATT) = M(N-1)N/2+N(M-1)M/2+(N-1)(M-1)M=84$

The WML of locations under contingent search can be broken into two separate components- the locations of values that need to be stored in memory, and the locations of the attributes that have already been searched. The location of values WML was shown in the uncontingent search analysis, and is reproduced here for completeness. The location of searched attributes WML is specific to the contingent analysis, and is identical for both ALT and ATT search sequences, due to their inherent symmetry in this component of the memory task.

General result:

<u>Proposition CL</u>: ALT is the unique minimum aggregate WML^{l} search sequence for M, N ≥ 2 (not proven).

This result is weaker than UL because it is no longer necessarily the case that ALT is the unique undominated WML^{l} sequence. One way of seeing this is that ALT and ATT can each dominate each other over certain subsequences of search for WML^{l2} - the component of WML^{l} for locations already searched. However, it is clearly true that the WML^{l2} required for ALT and ATT (and a restricted set of combinations of ALT and ATT) are unique WML^{l2}-undominated, if one has freedom to rearrange t's. This is why Proposition CL holds straightforwardly.

In the appendix I also show that the difference in maximum WML^{l} between the ATT and ALT search sequences increases linearly in M- the number of alternatives, and the difference in aggregate WML^{l} between ATT and ALT increases linearly in N and quadratically in M.

Notice that for contingent search, although ATT is a high WML search sequence, it is no longer weakly WML^{T} dominated by all search sequences. This is because in contingent search, ATT yields a minimum aggregate WML^{l2} (WML of searched attributes component), while it remains the maximum WML^{l1} (WML of the locations of value sums) sequence.

Total WML($WML^{v} + WML^{l}$ **):**

0	**2**	**3**	**4**
2	**5**	**6**	**7**
4	**6**	**7**	**8**
5	**7**	**8**	**9**

Figure 3.6.1: ALT

0	**8**	**9**	**10**
2	**9**	**10**	**11**
5	**10**	**11**	**10**
8	**11**	**12**	**9**

Figure 3.6.2: ATT

$\max_{\{t\}} \{WML^{v}(t \mid ALT)\} = N+M+1 = 9$

$\sum_{t=1}^{MN} WML^{v}(t \mid ALT) = M(N-1)N/2 + N(M-1)M/2+3MN-2N-M-1 = 83$

$\max_{\{t\}} \{WML^{v}(t \mid ATT)\} = 3M+N-4 = 12$

$\sum_{t=1}^{MN} WML^{v}(t \mid ATT) = M(N-1)N/2 + N(M-1)M/2+2M^2(N-1)+M(2-N) = 135$

I now present the first main result for contingent search.

General result:

Proposition 1CT: ALT is the unique minimum aggregate WML^T search sequence for M,N $\geq$ 2.

In the appendix I also show that the difference in maximum WML^T between the ATT and ALT search sequences increases linearly in M- the number of alternatives- with a slope of two, and the difference in aggregate WML^T between ATT and ALT increases linearly in N and quadratically in M, twice as fast as for WML^v or WML^l alone.

One of the puzzling regularities in GL's subjects' search behavior is that in search transitions from one attribute to the next, the ratio of within alternative transitions to within attribute transitions is 13:1, whereas the ratio is roughly 1:1 in Payne et all's (1988) isomorphic, but dimensionally smaller (four alternatives by four attributes) design. By showing that ALT is the unique minimum WML^T search sequence, whose WML^T grows much slower than ATT's as the dimensions of the information matrix increase, I provide a structural explanation for subjects' tendency to systematically shift towards ALT search in larger information matrices, which is also generally consistent with the fact that roughly 92% of GL subjects' search behavior is ALT.[18]

My first result also provides an explanation for one of the previously unexplained main systematic deviations from optimality observed in GL's subjects' behavior (Sanjurjo, 2008); on 39% of all search actions subjects search conditionally too deeply within an alternative. This violation is part of subjects' "one time through" approach in

---

[18] We do not yet know exactly what optimal search looks like, but it almost surely looks something like ALT, but with lots of jumping back and forth between alternatives, so I do not mean to imply that the high frequency of ALT search in GL's data is solely due to WML considerations.

which they over-search alternatives on the first pass through, then never return to search them further, rather than performing the WML-intensive exercise of switching back and forth between partially searched alternatives.

It is important to mention that ALT's WML^T superiority should be interpreted slightly differently for the uncontingent and contingent conditions. In uncontingent search the searcher's ALT sequence can be thought of as deterministically chosen ex-ante, as in classic search. Contingent search, on the other hand, will very rarely lead to a pure ALT sequence. Thus, deviating from ALT increases the WML^T under both uncontingent and contingent search, but in uncontingent search the searcher can choose ex-ante not to deviate from ALT, whereas contingent searchers have no choice but to deviate from ALT.

My second result explicitly compares the WML^T's of uncontingent and contingent search, holding the search sequence constant. The result follows directly from my benchmark definition of total WML.

Proposition 2UC: For any given search sequence, uncontingent search WML^T dominates contingent search.

The proof is immediate and obvious when one considers that contingent search is identical to uncontingent search, but with the added burden of remembering where one has searched. In Section III, under an enriched model of total WML, I show that while a version of the first result still holds the second result need not.

III. The WML benefits of systematic search

In this section I first motivate why it is important to enrich the benchmark model of WML from Section II in order to more fully reflect the WML benefits of systematicity. Under this enriched model, that only affects the model of WML for uncontingent search, I demonstrate that a refined version of my first result still holds; the unique WML^T undominated uncontingent search sequence is a refined (maximally systematic) version of ALT. I will then show how this result provides a structural explanation for both of GL's subjects' previously unexplained main deviations from optimality (Sanjurjo 2008), while validating the general (and obvious) intuition that more systematic uncontingent search sequences require less WML to execute correctly.

Under Section II's benchmark model of total WML ($WML^T = WML^v + WML^l$) lies the latent assumption that uncontingent search sequences are "hard-wired," thus WML-costless to execute. One implication of this model is that the two search sequences in figures 3.7.1 and 3.7.2, respectively, have identical streams of total WML, where the numbers in each cell of the matrix here represent the order that cell is searched in:

1	3	4	2
8	5	7	6
10	12	11	9
14	15	13	16

Figure 3.7.1: <u>ALT</u>

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

Figure 3.7.2: <u>ALT (refined)</u>

While it appears possible to likely that Figure 3.7.2's ALT (refined) uncontingent sequence is WML-costless to execute correctly, the same can not be said for the unrefined ALT search sequence in Figure 3.7.1, which will clearly require WML in the form of non-trivial instructions in order to be executed correctly.

Although the argument can be made that GL subjects' search behavior looks more like ALT (refined) than ALT (96% of within alternative or attribute search transitions are adjacent), so it is reasonable to assume that subjects costlessly execute these more systematic uncontingent sequences, this way of thinking misses the point. By demonstrating that Figure 3.7.1's uncontingent ALT sequence requires more WML to execute than Figure 3.7.2's ALT (refined) sequence, I provide a WML-based structural explanation for why subjects' search looks more like ALT (refined) than ALT in the first place.[19]

ALT (refined), from Figure 3.7.2, can be thought of as requiring memory of instructions (stated informally) "left-to-right, top-to-bottom" in this order. Similarly, instructions for an analogous ATT (refined) could be "top-to-bottom, left-to-right." One can argue over exactly how many memory items are required for such a set of instructions, however, the approach here will be to first settle on a natural definition of WML^{US} (the WML required to correctly execute the *uncontingent sequence* itself), which implies some level of cardinality- consistent across uncontingent search sequences, and

[19] I want to make it clear that this is certainly not the only reason that GL's subjects almost always make adjacent transitions within alternatives or attributes. A declining variance structure across attribute columns, in fact makes adjacent transitions within alternatives, in the direction of the declining variance, necessary for optimality. However, this type of behavioral systematicity is similar in flavor to subjects' adjacency bias in alternative switching, which occurs on 49% of all switches and is sub-optimal. In addition, about 5% of the time subjects transition against the declining variance structure, violating optimality, but almost always making adjacent transitions while doing so, thus it may indeed be the case that adjacency considerations for the sake of reducing WML rival if not dominate the here-confounded benefits of searching optimally.

then to focus on the ordinal comparison of WML^{US} across different search sequences. I demonstrate that together ALT (refined), ATT (refined), and diagonally tilted versions of these same sequences are the unique WML^{US} minimizing search sequences. Stated informally- the more "random" a search sequence looks to an observer, the longer the list of instructions necessary to remember it, thus the higher the number of memory items required to execute it accurately.

As just suggested, an uncontingent search sequence can be represented by a set of instructions. The systematicity of an uncontingent sequence can be thought of as the degree to which it takes advantage of iterating a small number of unique instructions. I assume that until necessary requirements for the next instruction are met, or in the absence of additional instructions, the previous instruction is iterated indefinitely. An uncontingent search sequence that has no systematicity, or iterated instructions, will require a unique instruction for the search of each attribute in the entire information matrix. A maximally systematic uncontingent search sequence can be fully characterized, for a two dimensional information matrix of any size, by just a starting point and two additional instructions.

In order for an uncontingent search sequence to be non-redundant and exhaustive, I assert that it must contain

1) a starting point

2) an instruction for each subsequent spatial search transition, until all attributes in all alternatives have been searched

Additionally, in order to focus the analysis (in an unbiased way), I assume that the rule must require no memory of the locations of previous searched attributes, only knowledge of the location of the currently searched attribute.

Proposition USWML: ALT (refined), ATT (refined), and diagonal analogues are together the unique minimum WML^{US} uncontingent search sequences for any information matrix larger than M=1 & N = 2 (or visa versa). They are the non-unique minimum WML^{US} uncontingent sequences for all smaller information matrices.[20]

This result has several key implications. First, it refines the first main result in Section II, by acknowledging that more systematic uncontingent search sequences require less WML to execute accurately. It follows immediately from Proposition USWML that ALT (refined) is the unique undominated WML^{T} search sequence.[21] This result is important because it supports the finding that GL's subjects' search behavior is highly systematic- 96% of within alternative or attribute search transitions are adjacent- thus look more like ALT (refined) than ALT.

Second, this result provides a structural explanation for the two previously unexplained main systematic deviations from optimality in GL subjects' search behavior. In 49% of all transitions from one alternative to another subjects make the costly error of

---

[20] Implicitly, this result shows that for very small (i.e. 1 x 2) information matrices the penalty in WML^{US} for following a non-systematic search sequence is zero, but as the dimensions become at least 2 x 2 the difference becomes non-zero (WML of 4 items vs. 3), and then this difference continues to grow along with the dimensions of the information matrix. Thus, the use of a systematic uncontingent search sequence becomes increasingly important with the size of the information matrix, if one wishes to keep down the WML required to implement it.

[21] Where total WML now includes the WML required to correctly execute the uncontingent sequence, in addition to the WML's of values and locations.

switching to an alternative that is less promising (in expectation) than at least one other available alternative. This is particularly puzzling behavior given the presumable ease with which compliance to this necessary condition should occur.[22] That it often does not, and that the violations of it are highly systematic, suggests strongly that subjects are deliberately adapting their behavior due to some underlying factor. Indeed, an econometric analysis in Sanjurjo (2008) reveals that subjects exhibit a strong adjacency bias in their transitions from one alternative to the next. This bias increases the systematicity in the search sequence, and as demonstrated in (the proof of) Proposition USWML, thus reduces WML.

In 39% of all search transitions in GL's data subjects search conditionally too deep within an alternative. This violation of optimality occurs almost exclusively when directly preceded by other search transitions within the same alternative. In fact, subjects almost never return to further search a previously partially searched alternative. Result one shows that the more alternatives that subjects leave simultaneously "open" to further search, the higher their WML will be. Thus, an adaptive strategy sensitive to the "openness" of alternatives is one that simply over-searches in the currently searched alternative knowing that it will never be returned to; this type of "one time through" approach is effective in reducing WML via increasing the systematicity of the uncontingent search sequence (in addition to allowing fewer open alternatives).

Lastly, this result justifies the otherwise disproportionately large prevalence of ATT search relative to other possible sequences. Without considering $\mathrm{WML}^{\mathrm{US}}$, ATT is weakly dominated by all other search sequences, yet accounts for 6% of the 8% not

---

[22] In the GL design compliance with this necessary condition for optimality literally translates into switching to the unsearched alternative with the largest visible number displayed in front of it.

attributed to ALT search. However, when WML^{US} is included in the model of total WML ATT is no longer weakly WML^{T} dominated by all search sequences, as ATT (refined) is actually a minimum WML^{US} uncontingent search sequence. Thus ATT (refined)'s maximal systematicity is now acknowledged and the refined result is more consistent with subjects' behavior. This result is therefore also consistent with 98% of all search behavior in the GL dataset being either ALT or ATT, and 96% of all search transitions being adjacent.

In Section II's analysis, which assumes WML^{US} to be zero for all search sequences, my second main result follows immediately: that for any given search sequence the uncontingent version WML dominates the contingent version. Once the zero WML^{US} assumption is relaxed, however, the theoretical comparison of total WML between uncontingent and contingent search becomes non-trivial. Holding the WML of values and locations constant, whereas the contingent searcher must remember which attributes he has searched, the uncontingent searcher must remember the instructions of his uncontingent search sequence. Thus, whether total WML is higher for a contingent or uncontingent version of the same sequence reduces to the assumption made about the WML^{US} versus the assumption made about the "chunking" ability of contingent searchers. With minimal chunking ability a searcher will require one unit of WML for each attribute searched. Thus the information display could always be made large enough dimensionally that the contingent version of the sequence eventually systematically requires more WML than the uncontingent. Under maximum chunking ability a searcher would use one unit of WML for any preceding sequence of attribute look ups, thus contingent search could, in theory, WML dominate uncontingent search in this case. It is beyond the scope of this paper to go any further into this issue, but what is clear is that unsophisticated chunkers would clearly be better off with uncontingent search, and that the more sophisticated the chunker the relatively better off she will be with contingent search. Relatedly, Chase and Simon (1973) find that, in a range of tasks including chess, degree of expertise can be almost exclusively explained by chunking ability.

IV. Conclusion

A searcher with no cognitive limitations normally does best searching contingently, and trades off depth of search within an alternative and breadth of search across alternatives optimally. In my (Sanjurjo, 2008) analysis of Gabaix and Laibson's (2006) multiple attribute/multiple alternative search experiments I found three puzzling patterns in subjects behavior; two were main systematic deviations from optimal search, and the third was a puzzling scale effect. Inspired by two unambiguous examples of subjects' working memory limitations in the dataset, as well as the well-documented finding (Miller 1956, Newell & Simon 1972) that WML is a fundamental bottleneck in decision making, I decided to explore the possibility of explaining both of subjects' main systematic deviations from optimal search, as well as the puzzling scale effect, via the single hypothesis that they have limited working memory.

Through three theoretical results I provide an explanation for all three puzzles. My first result demonstrates that ALT, a within alternative transition intensive search sequence, is the unique minimum WML search sequence. Because search behavior in GL's dataset is a combination of both contingent and uncontingent search I perform my analysis under both conditions. For uncontingent search I demonstrate that ALT is the unique undominated WML search sequence. For contingent search I demonstrate the slightly weaker result that ALT is the unique minimum aggregate WML search sequence. In addition, I show that the difference between the WML's required for the ATT (within attribute column transition intensive) and ALT search sequences increases rapidly with the dimensions of the information display. Therefore, this result provides an explanation for why subjects in the eight alternative by ten attribute GL display perform within alternative to within attribute search transitions at a ratio of 13:1, whereas subjects from Payne et al. (1988)'s isomorphic four alternative by four attribute display search at a ratio of roughly 1:1.

If the systematic shifting from ATT to ALT search in larger information matrices proves to be robust, then this finding has important implications regarding an ability of designers to strategically manipulate searchers through matrix size effects. Tversky and Simonson (1993) and Hardie, Johnson, and Fader (1993) discuss loss aversion and reference-dependence across alternatives within attributes, in multiple alternative/multiple attribute search. If it is true that increasing the size of an information matrix promotes ALT search, then overloaded subjects may get "attached" to the first alternative they search.

My first result also provides an explanation for one of the previously unexplained main systematic deviations from optimality observed in GL's subjects' behavior (Sanjurjo, 2008); on 39% of all search actions subjects search conditionally too deeply within an alternative. This violation is part of subjects' "one time through" approach in which they over-search alternatives on the first pass through, then never return to search them further, rather than performing the WML-intensive exercise of switching back and forth between partially searched alternatives.

My second result follows directly from the characterization of the benchmark model of WML used for my first result. It follows that for any given search sequence uncontingent search WML dominates contingent search. This result demonstrates that to whatever extent WML is costly, limited, or increases the probability of decision error, there is a non-trivial tradeoff between searching contingently and uncontingently.

For my third result I enrich my model of WML to more fully capture the presence of systematicity in uncontingent search sequences. I demonstrate the general and highly intuitive notion that more systematic search rules require less WML than less systematic search rules (ones that look as if they're generated from a random process). It follows from this result that a "refined" (specific) version of ALT is the unique undominated uncontingent search sequence- so a refined version of my first result still holds under this enriched model of WML.

My third result provides a structural explanation for the two main systematic deviations from optimal search found in the GL dataset. Subjects' costly tendency to disproportionately favor adjacent alternatives when transitioning from one alternative to another increases the degree of systematicity in their search sequences. Similarly, "over-searching" in a particular alternative is part of GL subjects' "one time through" approach that makes search more systematic and thus reduces WML relative to a sequence in which the searcher instead switches back and forth between partially searched alternatives.[23] Further, this result also suggests an explanation for why ATT search is more common in GL subjects' behavior than otherwise lower WML search sequences (ATT is weakly WML dominated by all search sequences in the non-enriched model of WML)- because the maximum systematicity of the sequence itself is now acknowledged. This result is therefore also consistent with 98% of all search behavior in the GL dataset being either ALT or ATT, and 96% of all within alternative or attribute search transitions being adjacent.

In computing WML step-by-step through different search sequences, I create a diagnostic tool that can be applied to multiple alternative/multiple attribute information matrices of any type. I operationalize a fundamental component of information overload, complexity, processing errors, and thus the meta-decision to use heuristic processes, in general. My results apply to the types of information matrices found on information aggregating websites for virtually all consumption and investment goods including homes and automobiles, health care plans, 401k plans, and even to information tables in academic papers.

Some key questions remain for future research in this area. One such question is why people wish to avoid incurring large WML's; is it because WML is costly, limited, or associated with increases in the probability of decision error? Relatedly, whether the aggregate WML or the maximum WML of a search sequence has a more profound affect on decision making remains to be better understood. Further, throughout my analysis I assume optimal discard of no longer needed information from memory. A failure to

[23] As mentioned, this type of "over-searching" within alternatives is also consistent with the first main result, which is based on the intuition that minimum WML search sequences have the fewest possible number of alternatives open to further search at any point in the search sequence.

discard information optimally would clearly lead to higher WML's for given search sequences, and could even lead to reversals in my first result.

Possible extensions of this paper include both experimentation and further theoretical work. By forcing groups of subjects to search the same sized information matrix by ALT or ATT, respectively, one can learn more about the relationship between WML and errors in decision making. One can also directly manipulate the size of the information matrix, the costs of observing attribute values, making the information matrix "open" or "closed" in terms of the simultaneous observability of attribute values, provide different types of memory aids, present information verbally, written, or pictorially, pre-load subjects with other WM demands, and make integration of attribute values non-trivial. Manipulation of these experimental control variables will allow for a deeper understanding of what factors stand most in the way of real people from being perfect-reasoning optimal searchers and choosers.

With the characterization of necessary conditions for optimal multiple alternative/multiple attribute search in Sanjurjo (2008) it is possible to solve for optimal search policies in reasonably sized information matrices, where behavioral models that are optimal subject to WM limitations can be tested explicitly.

It is my hope that by further focusing on the very composition of human bounded rationality we can continue to find simple structural approaches that explain existing puzzles in human behavior, while also promoting further new insights.

Appendix- Proofs

The appendix contains proofs of all non-trivial Propositions presented in Sections II and III of this paper, along with general formulas for the maximum and aggregate

WML of ALT and ATT search sequences (compared). I begin by defining the sets of ALT and ATT search sequences along with the ALT (refined) and ATT (refined) subsets of these sets of sequences, which are used in my third result.

Definitions:

a_{ij} : value of attribute in alternative (row) i and attribute (column) j, where i $\in$ M and j $\in$ N.

$t(a_{ij} \mid s)$: the search ordering of a_{ij} under search sequence s, so $t \in \{1,2,3,...,\mathrm{MN}\}$. t can be thought of as a "counter" of the sequential steps in a search sequence. Thus, for each well-defined search sequence there is a one-to-one mapping from each unique value of t to a particular attribute in a particular alternative.

ALT (refined): ALT is short for "within alternative search" and (refined) refers to this being the most obvious and systematic version of within alternative search. In Section III I demonstrated that it is also the ALT (refined) (henceforth $\mathrm{ALT_R}$) uncontingent sequence that requires the minimum amount of WML to employ. There are four versions of ALT (refined) search. In one ($\mathrm{ALT_R}$ 1) search starts in the left-most attribute of the top alternative and proceeds according to adjacent left-to right transitions until the right-most attribute of the top alternative is searched. Search then continues on the left-most attribute of the 2nd from the top alternative, iterating, until the entire information matrix has been searched. The 4 total variations come from inverting the search sequence just described horizontally and/or vertically.

More formally, the four search strategies can be characterized by the unique orders in which they visit each attribute in the information matrix:

1) $t(a_{ij} \mid \mathrm{ALT_R}\,1) = (i-1)\mathrm{N} + j$
2) $t(a_{ij} \mid \mathrm{ALT_R}\,2) = (\mathrm{M}-i)\mathrm{N} + j$
3) $t(a_{ij} \mid \mathrm{ALT_R}\,3) = (i-1)\mathrm{N} + (\mathrm{N} - j + 1)$
4) $t(a_{ij} \mid \mathrm{ALT_R}\,4) = \mathrm{MN} + 1 - ((i-1)\mathrm{N} + j)$

These four $\mathrm{ALT_R}$ search strategies can be represented graphically by:

$$
\begin{array}{llll}
\rightarrow\rightarrow\ldots\rightarrow & \rightarrow\rightarrow\ldots\rightarrow & \leftarrow\ldots\leftarrow\leftarrow & \leftarrow\ldots\leftarrow\leftarrow \\
\downarrow & \uparrow & \hfill\downarrow & \hfill\uparrow \\
\rightarrow\rightarrow\ldots\rightarrow & \vdots & \leftarrow\ldots\leftarrow\leftarrow & \hfill\vdots \\
\downarrow & \uparrow & \hfill\downarrow & \hfill\uparrow \\
\vdots & \rightarrow\rightarrow\ldots\rightarrow & \hfill\vdots & \leftarrow\ldots\leftarrow\leftarrow \\
\downarrow & \uparrow & \hfill\downarrow & \hfill\uparrow \\
\rightarrow\rightarrow\ldots\rightarrow & \rightarrow\rightarrow\ldots\rightarrow & \leftarrow\ldots\leftarrow\leftarrow & \leftarrow\ldots\leftarrow\leftarrow \\
\mathrm{ALT_R}\,1 & \mathrm{ALT_R}\,2 & \mathrm{ALT_R}\,3 & \mathrm{ALT_R}\,4
\end{array}
$$

ALT: Like $\mathrm{ALT_R}$ but now the search sequence within any alternative can be rearranged in any order and the order in which each alternative is visited can also be rearranged in any order. Thus there are $\mathrm{N!^M M!}$ different versions of ALT.

ATT (refined): ATT is short for "within attribute search" and (refined) refers to this being the most obvious and systematic version of within attribute search. In Section II I show why it is also the ATT (refined) (henceforth $\mathrm{ATT_R}$) uncontingent sequence that requires the minimum amount of WML to employ within the class of all ATT sequences. There are four versions of $\mathrm{ATT_R}$ search. In one ($\mathrm{ATT_R}\,1$) search starts in the left-most attribute of the top alternative and proceeds according to adjacent top-to-bottom transitions across alternatives until the left-most attribute of the bottom alternative is searched. Search then continues to the 2nd from the left-most attribute in the top alternative, iterating, until the entire information matrix has been searched. The 4 total variations come from inverting the search sequence just described horizontally and/or vertically.

More formally, the four search strategies can be defined as:

1) $t(a_{ij} \mid \mathrm{ATT_R}\,1) = (j-1)\mathrm{M} + i$

2) $t(a_{ij} \mid \mathrm{ATT_R}\,2) = (j-1)\mathrm{M} + (\mathrm{M} - i + 1)$

3) $t(a_{ij} \mid \mathrm{ATT_R}\, 3) = (\mathrm{N} - j)\mathrm{M} + i$

4) $t(a_{ij} \mid \mathrm{ATT_R}\, 4) = \mathrm{MN} + 1 - ((j-1)\mathrm{M} + i$

These four $\mathrm{ATT_R}$ search strategies can be represented graphically by:

$$\begin{array}{cccc}
\begin{array}{ccccccc}
\downarrow & \rightarrow & \downarrow & \rightarrow & \cdots & \rightarrow & \downarrow \\
\downarrow & & \downarrow & & & & \downarrow \\
\vdots & & \vdots & & & & \vdots \\
\downarrow & & \downarrow & & & & \downarrow
\end{array}
&
\begin{array}{ccccccc}
\uparrow & & \uparrow & & & & \uparrow \\
\vdots & & \vdots & & & & \vdots \\
\uparrow & & \uparrow & & & & \uparrow \\
\uparrow & \rightarrow & \uparrow & \rightarrow & \cdots & \rightarrow & \uparrow
\end{array}
&
\begin{array}{ccccccc}
\downarrow & \leftarrow & \cdots & \leftarrow & \downarrow & \leftarrow & \downarrow \\
\downarrow & & & & \downarrow & & \downarrow \\
\vdots & & & & \vdots & & \vdots \\
\downarrow & & & & \downarrow & & \downarrow
\end{array}
&
\begin{array}{ccccccc}
\uparrow & & & & \uparrow & & \uparrow \\
\uparrow & & & & \uparrow & & \uparrow \\
\vdots & & & & \vdots & & \vdots \\
\uparrow & \leftarrow & \cdots & \leftarrow & \uparrow & \leftarrow & \uparrow
\end{array}
\\
\mathrm{ATT_R}\, 1 & \mathrm{ATT_R}\, 2 & \mathrm{ATT_R}\, 3 & \mathrm{ATT_R}\, 4
\end{array}$$

ATT: Like $\mathrm{ATT_R}$ but now the search sequence within any attribute (column) can be rearranged in any order and the order in which each attribute (column) is visited can also be rearranged in any order. This means there are $\mathrm{M!}^{\mathrm{N}}\mathrm{N!}$ different versions of ATT.

Result one:

WML of values (Uncontingent)

Proposition UV: ALT is the unique undominated $\mathrm{WML^v}$ search sequence for M,N ≥ 2.

Proof: An alternative can said to be 'open' if at least one attribute in it has been searched, but not all attributes in it have been searched. An alternative can be said to be 'closed' if all attributes in it have been searched.

$\mathrm{WML^v}$ = # of open alternatives + 1{if at least one alternative is closed} (by optimal disposal of unneeded information). It is possible to generate a set of search sequences such that, for each sequence, at most one alternative is ever open at a time. This property defines the set of sequences collectively referred to as the ALT(relaxed) sequence which yields a $\mathrm{WML^v}$-stream of 011…1|122….2|122…2|…|122…2 where the first digit represents the $\mathrm{WML^v}$ required for the first search in the ALT sequence, the second for the second, and so on. Vertical bars separate sequential groups of N searches.

Call a sequence that deviates from ALT ~ALT. ~ALT's first deviation from ALT must occur in one of the M sub-sequences of N searches. Once the first deviation occurs there must necessarily be two alternatives open simultaneously for at least one search. If this occurs at t_0, in the first N searches, then $WML^v(t_0+1|\sim ALT; M,N) = 2 > 1 = WML^v(t_0+1|ALT; M,N)$. If it occurs in any of the next M-1 groups of N searches then $WML^v(t_0+1|\sim ALT; M,N) = 3 > 2 = WML^v(t_0+1|ALT)$, thus it is impossible for any ~ALT sequence to be WML^v-preferred to ALT.

Further, suppose that ALT is not WML^v-preferred to all other sequences. Then it must be the case that some ~ALT is WML^v-preferred to ALT for some search sub-sequence. This can clearly not occur before ~ALT deviates from ALT. It also cannot occur between the t_0 where the first deviation from ALT occurs, and t_0+1. Thus it must occur for some sub-sequence of t_0+2 to MN. But if at least two alternatives remain open under ~ALT then WML^v is at least 2, which is the maximum WML^v of ALT. If only one alternative, or zero, is open, then ~ALT is now searching identically as ALT. Thus, in either case $WML^v(t|\sim ALT; M,N) \geq WML^v(t|ALT; M,N)$ for $t \in \{t_0 + 2, \ldots, MN\}$. □

<u>Proposition UVATT</u>: ATT is weakly WML^v dominated by all search sequences for M,N ≥ 2.

Proof: The WML^v of ATT is 0123…M-1|MM…M|…|MM…M|MM(M-1)(M-2)…2. The WML^v for any t in the search sequence is determined by the unique number of alternatives that have been searched at least once but not exhaustively, plus one that has been searched exhaustively (the one with the highest cumulative value), if there is at least one. Thus it is impossible that the WML^v be larger than ATT's for any of the first M searches, the second M searches, and so on until the last M searches. With one attribute left the WML^v must be 2. With two left it can be at most 3, with three left it can be at most 4, and so on. These are precisely the $WML^v(t|ATT; M,N)$ for the last M searches ($t \in \{(N-1)M, (N-1)M+1, \ldots, NM\}$) of the ATT sequence. □

It is easy to establish non-uniqueness by observing that any sequence with each of the first M searches in a different alternative, and each of the last M searches in a different alternative will yield an identical string of WML^{v}'s as ATT.

It follows trivially from Proposition UV that ALT WML^{v} dominates ATT (In fact, in the appendix I also prove that ATT is weakly WML^{v} dominated by all search sequences).

It then follows directly from this relationship that

$\max\limits_{\{t\}}\{\text{WML}^{\text{v}}(t \mid \text{ATT; M,N})\} > \max\limits_{\{t\}}\{\text{WML}^{\text{v}}(t \mid \text{ALT; M,N})\}$ and that

$\sum\limits_{t=1}^{\text{MN}} \text{WML}^{\text{v}}(t \mid \text{ATT; M,N}) > \sum\limits_{t=1}^{\text{MN}} \text{WML}^{\text{v}}(t \mid \text{ALT; M,N})$ for $\text{M,N} \geq 2$.

For any $\text{M, N} \geq 2$ $\max\limits_{\{t\}}\{\text{WML}^{\text{v}}(t \mid \text{ALT; M,N})\} = 2$, whereas

$\max\limits_{\{t\}}\{\text{WML}^{\text{v}}(t \mid \text{ATT; M,N})\} = \text{M}$, so clearly the difference in maximum WML^{v} between the ATT and ALT search sequences increases linearly in M- the number of alternatives.

With a bit of algebra it is easy to show that because

$\sum\limits_{t=1}^{\text{MN}} \text{WML}^{\text{v}}(t \mid \text{ATT; M,N}) = \text{M}^2(\text{N-1}) + (\text{M-1})$

and $\sum\limits_{t=1}^{\text{MN}} \text{WML}^{\text{v}}(t \mid \text{ALT; M,N}) = \text{2MN-N-M}$, then

$\sum\limits_{t=1}^{\text{MN}} \text{WML}^{\text{v}}(t \mid \text{ATT; M,N}) - \sum\limits_{t=1}^{\text{MN}} \text{WML}^{\text{v}}(t \mid \text{ALT; M,N}) = (\text{M-1})^2(\text{N-1})$. Thus, the difference in aggregate WML^{v} between ATT and ALT increases linearly in N and quadratically in M.

WML of Locations (Uncontingent):

Proposition UL: ALT is the unique undominated WML^{l} search sequence for $\text{M,N} \geq 2$.

Proof: WML^l = # of open alternatives + 1{if at least one alternative is closed} + 1{if at least two alternatives are closed and none are open} – 1. (by optimal disposal of unnecessary information). It is possible to generate a set of search sequences such that, for each sequence, at most one alternative is ever open at a time. This property defines the set of sequences collectively referred to as the ALT sequence which yields a WML^l-stream of 000…0|011….1|111…1|…|111…1 where the first digit represents the WML^l required for the first search in the ALT sequence, the second for the second, and so on. Vertical bars separate sequential groups of N searches.

Call a sequence that deviates from ALT ~ALT. ~ALT's first deviation from ALT must occur in one of the M sub-sequences of N searches. Once the first deviation occurs there must necessarily be two alternatives open simultaneously for at least one search. If this occurs at t_0, in the first N searches then WML^l (t_0+1|~ALT; M,N) = 1 > 0 = WML^l (t_0+1|ALT; M,N). If it occurs in any of the next M-1 groups of N searches then WML^l (t_0+1|~ALT; M,N) = 2 > 1 = WML^l (t_0+1|ALT; M,N), thus it is impossible for any ~ALT sequence to be WML^l-preferred to ALT.

Further, suppose that ALT is not WML^l-preferred to all other sequences. Then it must be the case that some ~ALT is WML^l-preferred to ALT for some search sub-sequence. This can clearly not occur before ~ALT deviates from ALT. It also cannot occur between the t_0 where the first deviation from ALT occurs, and t_0+1. Thus it must occur for some sub-sequence of t_0+2 to MN. But if at least two alternatives remain open under ~ALT then WML^l is at least 1, which is the maximum WML^l of ALT. If only one alternative, or zero, is open, then ~ALT is now searching identically as ALT. Thus, in either case WML^l (t|~ALT; M,N) ≥ WML^l (t|ALT; M,N) for t ε { t_0 + 2,…, MN} . □

<u>Proposition ULATT</u>: ATT is weakly WML^l dominated by all search sequences for M,N ≥ 2.

Proof: The WML^l of ATT locations is 0012…M-1|(M-1)(M-1)…(M-1)|…|(M-1)(M-1)…(M-1)|(M-1)(M-1)(M-2)(M-3)…1. The location WML^l at any point in the search sequence is determined by the unique number of alternatives that have been searched at least once but not exhaustively, plus one that has been searched exhaustively (the one with the highest cumulative value), if there is at least one, plus one if at least two alternatives are closed and none are open, minus one (with an uncontingent sequence, m values can be matched to m alternatives with only m-1 explicit mappings made). Thus it is impossible that the WML^{l} can be larger than ATT's for any of the first M searches, the second M searches, and so on until the last M searches. With one attribute left the WML^{l} must be 1. With two left it can be at most 2, with three left it can be at most 3, and so on. These are precisely the WML^{l} (t|ATT; M,N) for the last M searches (t ε {(N-1)M, (N-1)M +1, …, NM}) of the ATT sequence. □

It is easy to establish non-uniqueness by observing that any sequence with each of the first M searches in a different alternative, and each of the last M searches in a different alternative will yield an identical string of WML^{l}'s as ATT.

It follows directly from Proposition UL that ALT WML^{l} dominates ATT (In the appendix I also prove that ATT is weakly WML^{l} dominated by all search sequences). Subsequently,

$$\max_{\{t\}}\{\mathrm{WML}^{l}(t \mid \mathrm{ATT;\ M,N})\} > \max_{\{t\}}\{\mathrm{WML}^{l}(t \mid \mathrm{ALT;\ M,N})\} \text{ and}$$

$$\sum_{t=1}^{\mathrm{MN}}\mathrm{WML}^{l}(t \mid \mathrm{ATT;\ M,N}) > \sum_{t=1}^{\mathrm{MN}}\mathrm{WML}^{l}(t \mid \mathrm{ALT;\ M,N}) \text{ for M,N} \geq 2.$$

For any M, N ≥ 2, $\max_{\{t\}}\{\mathrm{WML}^{l}(t \mid \mathrm{ALT;\ M,N})\} = 1$, whereas $\max_{\{t\}}\{\mathrm{WML}^{l}(t \mid \mathrm{ATT;\ M,N})\} = \mathrm{M}-1$, so clearly the difference in maximum WML^{l} between the ATT and ALT search sequences increases linearly in M- the number of alternatives.

With a bit of algebra it is easy to show that because

$$\sum_{t=1}^{\mathrm{MN}}\mathrm{WML}^{l}(t \mid \mathrm{ATT;\ M,N}) = (\mathrm{N}-1)(\mathrm{M}-1)\mathrm{M}$$

and $\sum_{t=1}^{MN} \text{WML}^{l}(t \mid \text{ALT; M,N}) = (\text{M-1})\text{N-1}$, then

$\sum_{t=1}^{MN} \text{WML}^{l}(t \mid \text{ATT; M,N}) - \sum_{t=1}^{MN} \text{WML}^{l}(t \mid \text{ALT; M,N}) = (\text{M-1})^2(\text{N-1})\text{-M+2}$. This result shows that the difference in aggregate WML^{l} between ATT and ALT increases linearly in N and quadratically in M.

Total WML (Uncontingent):

<u>**Proposition 1UT**</u>: ALT is the unique undominated WML^{T} search sequence for M,N ≥ 2.

It follows directly from Propositions UV and UL that ALT is the unique undominated WML^{T} search sequence.

It follows directly from Propositions UV and UL that ALT is the unique undominated WML^{T} search sequence, and (less importantly) that ATT is weakly WML^{T} dominated by all other search sequences, so clearly, ALT WML^{T} dominates ATT.

For any M, N ≥ 2 $\max_{\{t\}} \{\text{WML}^{l}(t \mid \text{ALT; M,N})\} = 3$, whereas $\max_{\{t\}} \{\text{WML}^{l}(t \mid \text{ATT; M,N})\} = \text{2M-1}$, so clearly the difference in maximum WML^{T} between the ATT and ALT search sequences increases linearly in M- the number of alternatives, with a slope of two.

With a bit of algebra it is easy to show that

$\sum_{t=1}^{MN} \text{WML}^{T}(t \mid \text{ATT; M,N}) - \sum_{t=1}^{MN} \text{WML}^{T}(t \mid \text{ALT; M,N}) = 2(\text{M-1})^2(\text{N-1})\text{-M+2}$. This result shows that the difference in aggregate WML^{T} between ATT and ALT increases linearly in N and quadratically in M, twice as fast as for WML^{v} or WML^{l} alone.

WML of Values (Contingent):

<u>Proposition CV</u>: ALT is the unique undominated $\mathrm{WML}^{\mathrm{v}}$ search sequence for M, N $\geq$ 2.

Proof: The proof is identical to that in Proposition UV because the number of values to be stored only depends on the sequence of search, not on whether it was anticipated or not.

<u>Proposition CVATT</u>: ATT is weakly $\mathrm{WML}^{\mathrm{v}}$ dominated by all search sequences for M,N $\geq$ 2.

Proof: The proof is identical to that in Proposition UVATT because the number of values to be stored only depends on the sequence of search, not on whether it was anticipated or not.

Analogous to the analysis of uncontingent search sequences, it follows trivially from Proposition CV that ALT $\mathrm{WML}^{\mathrm{v}}$ dominates ATT. In the appendix I also prove that ATT is in fact weakly $\mathrm{WML}^{\mathrm{v}}$ dominated by all search sequences
Subsequently,

$$\max_{\{t\}} \{\mathrm{WML}^{\mathrm{v}}(t \mid \mathrm{ATT; M,N})\} > \max_{\{t\}} \{\mathrm{WML}^{\mathrm{v}}(t \mid \mathrm{ALT; M,N})\} \text{ and}$$

$$\sum_{t=1}^{\mathrm{MN}} \mathrm{WML}^{\mathrm{v}}(t \mid \mathrm{ATT; M,N}) > \sum_{t=1}^{\mathrm{MN}} \mathrm{WML}^{\mathrm{v}}(t \mid \mathrm{ALT; M,N}) \text{ for M, N} \geq 2.$$

For any M, N $\geq$ 2 $\max_{\{t\}} \{\mathrm{WML}^{\mathrm{v}}(t \mid \mathrm{ALT\ ; M,N})\} = 2$, whereas

$\max_{\{t\}} \{\mathrm{WML}^{\mathrm{v}}(t \mid \mathrm{ATT\ ; M,N})\} = \mathrm{M}$, so clearly the difference in $\mathrm{WML}^{\mathrm{v}}$, between the ATT and ALT search sequences increases linearly in M- the number of alternatives.

With a bit of algebra it is easy to show that

$$\sum_{t=1}^{\mathrm{MN}} \mathrm{WML}^{\mathrm{v}}(t \mid \mathrm{ATT_R\ ; M,N}) = \mathrm{M}^2(\mathrm{N}\text{-}1) + (\mathrm{M}\text{-}1)$$

and that $\sum_{t=1}^{\mathrm{MN}} \mathrm{WML}^{\mathrm{v}}(t \mid \mathrm{ALT_R\ ; M,N}) = 2\mathrm{MN} - \mathrm{N} - \mathrm{M}$, so that

$\sum_{t=1}^{MN} \mathrm{WML^v}(t \mid \mathrm{ATT_R}\,;\, \mathrm{M{,}N}) - \sum_{t=1}^{MN} \mathrm{WML^v}(t \mid \mathrm{ALT_R}\,;\, \mathrm{M{,}N}) = (\mathrm{M}\text{-}1)^2(\mathrm{N}\text{-}1)$. This result shows that the difference in aggregate $\mathrm{WML^v}$ between $\mathrm{ATT_R}$ and $\mathrm{ALT_R}$ increases linearly in N and quadratically in M.

Locations WML (Contingent):

Proposition CL: ALT is the unique minimum aggregate $\mathrm{WML^l}$ search sequence for M, N ≥ 2 (not proven).

Proof:

Thus, it follows from Proposition CL that

$\sum_{t=1}^{MN} \mathrm{WML}^{l}(t \mid \mathrm{ATT}\ ; \mathrm{M,N}) > \sum_{t=1}^{MN} \mathrm{WML}^{l}(t \mid \mathrm{ALT}\ ; \mathrm{M,N})$. It is also clearly true that

$\max_{\{t\}} \{\mathrm{WML}^{l}(t \mid \mathrm{ATT}\ ; \mathrm{M,N})\} > \max_{\{t\}} \{\mathrm{WML}^{l}(t \mid \mathrm{ALT}\ ; \mathrm{M,N})\}$ for $\mathrm{M, N} \geq 2$.

Notice that for contingent search, although ATT is a high WML search sequence, it is no longer weakly WML^{T} dominated by all search sequences. This is because in contingent search, ATT yields a minimum aggregate WML^{l2} (WML of searched attributes component), while it remains the maximum WML^{l1} (WML of the locations of value sums) sequence.

For any $\mathrm{M, N} \geq 2$ $\max_{\{t\}} \{\mathrm{WML}^{l}(t \mid \mathrm{ALT; M,N})\} = \mathrm{N+M-1}$, whereas

For any $\mathrm{M, N} \geq 3$ $\max_{\{t\}} \{\mathrm{WML}^{l}(t \mid \mathrm{ATT; M,N})\} = \mathrm{N+2M-4}$, so clearly the difference in WML^{l} between the ATT and ALT search sequences increases linearly in M- the number of alternatives, when $\mathrm{M, N} \geq 3$.

For $\mathrm{N} = 2, 2 \leq \mathrm{M} \leq 4$ and for $\mathrm{M} = 2, \mathrm{N} \geq 2$ $\max_{\{t\}} \{\mathrm{WML}^{l}(t \mid \mathrm{ATT; M,N})\} = \mathrm{N+M-1}$

For $\mathrm{N} = 2, \mathrm{M} \geq 5$ $\max_{\{t\}} \{\mathrm{WML}^{l}(t \mid \mathrm{ATT; M,N})\} = \mathrm{2M-3}$

In all cases the aggregate WML^{l} of ATT is equal to or larger than the WML^{l} of ALT. With a bit of algebra it is easy to show that

$$\sum_{t=1}^{MN} \mathrm{WML}^{l}(t \mid \mathrm{ATT; M,N}) = \left(\frac{\mathrm{(N\text{-}1)N}}{2}\right)\mathrm{M} + \left(\frac{\mathrm{(M\text{-}1)M}}{2}\right)\mathrm{N} + \mathrm{(N\text{-}1)(M\text{-}1)M} \text{ and that}$$

$$\sum_{t=1}^{MN} \mathrm{WML}^{l}(t \mid \mathrm{ALT; M,N}) = \left(\frac{\mathrm{(N\text{-}1)N}}{2}\right)\mathrm{M} + \left(\frac{\mathrm{(M\text{-}1)M}}{2}\right)\mathrm{N} + \mathrm{(M - 1)N - 1}, \text{ so that}$$

$$\sum_{t=1}^{MN} \mathrm{WML}^{l}(t \mid \mathrm{ATT; M,N}) - \sum_{t=1}^{MN} \mathrm{WML}^{l}(t \mid \mathrm{ALT; M,N}) = \mathrm{(M\text{-}1)^2 (N\text{-}1) - M + 2}, \text{ as in the}$$

uncontingent search case. This result shows that the difference in aggregate WML^{l} between ATT and ALT increases linearly in N and quadratically in M.

Total WML (Contingent):

<u>Proposition 1CT</u>: ALT is the unique minimum aggregate WML^{T} search sequence for M,N $\geq$ 2.

It follows directly from Propositions UV and CL that ALT is the minimum aggregate WML^{T} search sequence.

For any M, N $\geq$ 2 $\max_{\{t\}} \{\text{WML}^{\text{I}}(t \mid \text{ALT; M,N})\} = \text{N+M+1}$, whereas

For any M, N $\geq$ 3 $\max_{\{t\}} \{\text{WML}^{\text{I}}(t \mid \text{ATT; M,N})\} = \text{3M+N-4}$, so clearly the difference in WML^{I} between the ATT and ALT search sequences increases linearly in M (by a factor of two)- the number of alternatives, when M, N > 2.

For M = 2, N $\geq$ 2 $\quad \max_{\{t\}} \{\text{WML}^{\text{I}}(t \mid \text{ATT; M,N})\} = \text{N+2M-1}$

For N = 2, 2 $\leq$ M $\leq$ 4 $\quad \max_{\{t\}} \{\text{WML}^{\text{I}}(t \mid \text{ATT; M,N})\} = \text{2M+1}$

For N = 2, M $\geq$ 5 $\quad \max_{\{t\}} \{\text{WML}^{\text{I}}(t \mid \text{ATT; M,N})\} = \text{3M-4}$

In all cases the aggregate WML^{T} of ATT is equal to or larger than the WML^{T} of ALT.

With a bit of algebra it is easy to show that

$\sum_{t=1}^{\text{MN}} \text{WML}^{\text{T}}(t \mid \text{ATT; M,N}) - \sum_{t=1}^{\text{MN}} \text{WML}^{\text{T}}(t \mid \text{ALT; M,N}) = 2(\text{M-1})^2(\text{N-1}) - \text{M} + 2$, just as in the uncontingent search case. This result shows that the difference in aggregate WML^{T} between ATT and ALT increases linearly in N and quadratically in M, twice as fast as WML^{v} or WML^{I} alone.

Result three:

<u>Proposition USWML</u>: ALT (refined), ATT (refined), and diagonal analogues are together the unique minimum WML^{US} uncontingent search sequences for any information matrix larger than M=1 & N = 2 (or visa versa). They are the non-unique minimum WML^{US} uncontingent sequences for all smaller information matrices.[24]

Proof: First consider the cases where M, N ≥ 3. W.l.o.g. I focus on ALT1(strict). The uncontingent search sequence ALT(strict) can be fully represented by the search rule:
1) search the left-most attribute of the spatially highest unsearched alternative
2) proceed by transitioning one attribute to the right $\overrightarrow{(1,0)}$
3) when there are no more attributes to the right, repeat step 1

Suppose the starting point were not the left-most or right-most attribute in an alternative. Then at least three directions would be necessary just to search all of the attributes in that particular alternative. At least one additional direction would then be necessary to search the rest of the alternatives.

By symmetry, the same argument shows that a minimum WML^{US} search sequence must not start in an alternative that is not the spatially highest or lowest. Thus, search must start in a corner of the information matrix for a minimum WML^{US} search sequence.

Now suppose that search transitions are systematic but not adjacent. Then at least three directions (or equivalently one direction with two contingencies) are necessary in order to search all of the attributes in the first searched alternative (attribute, or diagonal). At least one direction would then be necessary to search the rest of the alternatives.

---

[24] Implicitly, this result shows that for very small (i.e. 1 x 2) information matrices the penalty in WML^{C} for following a non-systematic search sequence is zero, but as the dimensions become at least 2 x 2 the difference becomes non-zero (WM of 4 items vs. 3), and then this difference continues to grow along with the dimensions of the information matrix. Thus, the use of a systematic uncontingent search sequence becomes increasingly important with the size of the information matrix, if one wishes to keep down the WML required to implement it.

Suppose that a search transition is not vertical, horizontal, or diagonal. Then the search rule requires at least two more directions in order to search the rest of the attributes in the first alternative searched.

Thus any search sequence other than ALT(strict), ATT(Strict), or diagonal equivalents, requires strictly more directions.

Now consider the cases where M or N < 3.

If M = N = 1, then all uncontingent search sequences require just one direction.

If M = 1 and N = 2 (or visa versa) then ALT and non-systematic require the same number of directions.

If M = 1 and N > 2 (or visa versa) then ALT requires the unique minimum number of directions: two, using an argument similar to that used in the above proof.

If M = N = 2 ALT or ATT or diagonal require a starting point and two directions whereas any other uncontingent search sequence requires a starting point and at least three directions. The same is true for M = 2 and N > 2 (or visa versa). □

References

Baddeley, A. (1992): “Working memory,” *Science*, 255, 556-559.

Benabou, Roland, and Jean Tirole (2002): “Self-Confidence and Personal Motivation,” *Quarterly Journal of Economics*, 3, 871-915.

Bernheim, Douglas, and Thomadsen (2005): “Memory and Anticipation,” *Economic Journal*, 115, 271-304.

Chase, William, and Herbert Simon (1973): “The mind's eye in chess,” In W.G. Chase, editor, *Visual information processing*. New York: Academic Press. (pp. 215-281).

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.

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.

Hardie, Bruce, Eric Johnson, Peter Fader (1993): "Modeling Loss Aversion and Reference Dependence Effects on Brand Choice," Marketing Science, 12, 378-394.

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.

Mullainathan, Sendhil (2002): "A Memory-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 117, 735-774.

Newell, Allen, and Herbert Simon (1972): *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

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

Payne, John, James Bettman, and Eric Johnson (1988): "Adaptive strategy selection in Decision Making," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 534-552.

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

Sanjurjo, Adam (2008): "Search with Multiple Attributes and Alternatives: Theory & Empirics," Working paper.

Simonson, Itamar, and Amos Tversky (1993): "Context-Dependent Preferences," *Management Science*, 39, 1179-1189.

Stigler, George (1961): "The Economics of Information," *The Journal of Political Economy*, 69, 213-225.

Tversky, Amos (1972): "Elimination by Aspects: A Theory of Choice," *Psychological Review*, 79, 281-299.

Wilson, Andrea (2003): “Bounded Memory and Biases in Information Processing,” Working paper.

This chapter is currently being prepared for submission for publication of the material. The dissertation author was the primary investigator and author of this material.