Components of Effort for Repeated Tasks

ANDRES MUSALEM, UNIVERSITY OF CHILE
amusalem7@gmail.com

RICARDO MONTOYA, UNIVERSITY OF CHILE
ricmontoya@gmail.com

MARTIN MEIßNER, ZEPPELIN UNIVERSITY
mmeissner79@googlemail.com

JOEL HUBER, DUKE UNIVERSITY
joel.huber@duke.edu

Components of Attentional Effort for Repeated Tasks

Abstract

This paper identifies four attentional processes that increase efficiency and accuracy in repeated lexicographic tasks using an instructed strategy approach. We propose a framework to decompose attentional effort used to make a decision into four components: Orientation, Wrong Target, Duration, and Repetition. Orientation assesses attention to decision rules and the location of relevant information. Wrong Target measures wasted effort on unneeded information. Duration gauges time spent on each piece of needed information. Repetition measures the number of views on each relevant item. Greater Orientation is associated with lower effort in other components and increased accuracy. Repetition is most variable across individuals but generates the greatest improvement with practice. Duration is less affected by the other components and shows minimal improvement with experience. Finally, Wrong Target is similarly resistant to practice, but it is the only component strongly and positively associated with making errors.

Keywords: Attention Effort, Decision Analysis, Eye-tracking, Repeated Tasks

Total word count 9,817 (not including tables, figures, references and web appendices)

Components of Attentional Effort for Repeated Tasks

1 Introduction

Individuals often perform tasks repeatedly. For example, consider a financial analyst using balance sheet information to calculate the risk of a loan; a customer service representative processing a refund; a radiologist searching a chest image for cancerous nodules, a consumer choosing which products to purchase in every visit to a store, or a travel agent finding the flights that meet a traveler's requirements. For these tasks to be successful they require learning the rules and using acquired knowledge to find and process relevant information.

This paper uncovers four relevant attentional components critical to performing repeated tasks and proposes a conceptual framework to characterize them. We examine these components that combine to make up effort used in making a decision: Orientation, Wrong Target, Duration, and Repetition. Orientation is a measure of attention to the rules and the location of important information. Wrong Target quantifies the proportion of attention to unneeded information. Duration is the time spent accessing each piece of relevant information. Finally, Repetition assesses the extent to which relevant information is repeatedly accessed.

We demonstrate the applicability of this framework with an eye-tracking study where participants search for the best alternative using an instructed lexicographic rule (in line with other papers using an instructed strategy approach, see, for example, Fechner, Schooler, & Pachur, 2018). The lexicographic rule is particularly appropriate for three reasons. First, the simplicity of the task allows participants to perform it repeatedly in a single lab setting and identify the learning processes over time. Second, the defined rules for the lexicographic task make it possible to determine the extent to which the participants focused on the most valuable information. Finally, the lexicographic rule and the similar take-the-best rule

(Gigerenzer & Goldstein, 1996) have been found to be more effective than complex compensatory rules in a number of real decision contexts. For example, Graefe & Armstrong (2012) show election predictions are better assessed with a lexicographic rule than a compensatory rule, while Pachur & Marinello (2013) show that expert airport custom inspectors gain from a take-the-best rule, and Garcia-Retamero & Dhami (2009) demonstrate that experts predicting burglaries gain from a lexicographic rule. Generally, Gigerenzer & Gaissmaier (2011) propose that quick and frugal decision making characterized by take-the-best rule is most effective when there is high cue redundancy and high variability in cue weight, particularly in information intensive, cognitively demanding contexts.

Our goal is to understand the attentional processes that lead to greater efficiency and accuracy through the use of an instructional strategy approach. Specifically, the application of our proposed decomposition allows us to answer three research questions. First, how much do the four components vary across participants and change with experience? Second, does effort in one component alter the need for effort in the others? Finally, do the roles of the components shift with different performance incentives?

The next section describes our proposed conceptual framework, outlining the relevance of the effort components and the links to previous research that investigated these components.

2 A Conceptual Framework for Attention Effort

Our research builds on earlier work that investigated the sources of effort in rule-based lexicographic decisions. Bettman, Johnson, & Payne (1990) and Khader, Pachur, Weber, & Jost (2016) differ in their focus of the factors that generate effort, but share with us the interest in using measures for a better understanding of the underlying processes in repeated structured tasks.

Bettman et al. (1990) define elementary information processes called EIP's such as compare, eliminate, read or add that can be used to describe the operations needed when applying a decision strategy. Thus, a choice can be represented as a sequence of mental events in which the effort needed to implement a decision strategy can be quantified as a function of the elementary steps needed. The authors explore a number of tasks, including the use of a lexicographic rule, and show that addition and multiplication take substantially more time and perceived effort than comparisons or eliminations. They show that the EIP framework can be used to predict response times and subjective cognitive effort for defined choice strategies across tasks that differ in terms of the number of attributes and alternatives. Their results include but are not focused on the read, compare and eliminate EIPs that are central for the lexicographic rule. However, their idea of identifying elementary information processes provides an innovative way to characterize the effort implications across diverse decision processes.

Khader et al. (2013) ask participants to comply with a take-the-best lexicographic decision rule on binary choices. Processing of the decisions requires recall of relationships among attribute levels learned during an extensive training session. They track the time needed to complete each task and thereby estimate the effort from using information recalled from long-term memory. They also find a weaker temporal effect associated with attending to attributes that are otherwise irrelevant to the particular task. Importantly, using fMRI they are able to identify different neural locations associated with recalling location, facial or product information cues. The paper thus identifies the neurological basis for effort related to memory recall.

Both research approaches differ in the kind of cognitive effort associated with repeated decisions. Bettman et al. (1990) explores computational effort, while Khader et al. (2016) assesses memory retrieval. In contrast to both, our paper demonstrates that the attentional effort of a lexicographic task that does not require computation or long-term memory can provide insights about how attention evolves across four distinct components. We measure external attention to rules (Orientation), correct application of the rules (Wrong Target), attentional duplication (Repetition) and processing speed (Duration). Based on these measures we assess the extent to which each attentional component differentially decreases with time and impacts total attention. While these have been examined in a number of studies, ours may be the first that brings them together in one study. Because our research framework is new and the possible interaction effects of the components have not been studied, our research approach is exploratory in nature. Next, we review empirical findings regarding each of the four components.

2.1 Orientation

Orientation measures attention to rules and the location of important information to complete a decision task. Early researchers emphasized the importance of knowing the rules for proper performance (Langley & Simon, 1981). Russo & Leclerc (1994) suggested that the decision-making process consists of three consecutive stages: orientation, evaluation and verification. Their eye-tracking studies demonstrated that decision makers need time for orientation before evaluating stimuli in a task. Liechty, Pieters, & Wedel (2003) suggest that during exposure to complex scenes, attention switches between two latent states, which they labeled local and global. "The attention states themselves are unobservable but can be inferred from the patterns of eye movements that they give rise to. The idea is that people resolve the complex problem of attending to a natural scene by decomposing it into a set of simpler subproblems by attending to local regions in the scene over time" (p.130). Thus, global scanning serves to orient the participant to assist in local evaluations. Applied to a lexicographic task, our study permits a separation between global (orientation) focus, characterized by finding out where to look next, and a local (evaluation) stage that examines and processes task-relevant information.

2.2 Wrong Target

Wrong Target measures the extent to which the information from Orientation effort is incorrectly used. In the case of a lexicographic task, there is an optimal way to collect information that involves identifying the best alternatives in terms of the most important attribute and eliminating alternatives excluded by those important attributes. If more than one alternative remains, the process is repeated using the next most important attribute. Greater attention to Wrong Target results from violating the appropriate attribute order or from re-examining eliminated alternatives. Decision makers gain location knowledge. Orquin, Chrobot, & Grunert (2018) showed that predictable locations facilitate attention to relevant stimuli, either from examining the rules or from memory. Location knowledge has been shown to speed up visual search for simple visual targets (see e.g. Chun & Jiang, 1998). In line with Orquin & Mueller-Loose (2013), we expect that efficiency should increase through more fixations to task-relevant and fewer fixations to task-irrelevant information.

A number of eye-tracking studies measure attention to irrelevant information (Gegenfurtner, Lehtinen, & Säljö, 2011; Haider & Frensch, 1999). These studies demonstrate that inappropriate attention is strongly associated with low expertise across a number of fields, results that are consistent with the idea that Wrong Target provides a relevant measure of inefficient attention processing.

2.3 Fixation Duration

Fixation duration is a measure of the time required for each fixation. Shorter times are associated with scanning and automatic processes, whereas longer fixations are linked with deeper processing (Velichkovsky et al., 2002). Empirical studies found that increased cognitive load leads to increases in fixation duration (Rosch & Vogel-Walcutt, 2013). In line with this finding, Reutskaja, Nagel, Rangel, & Camerer (2011) provide evidence that average fixation duration decreases in more complex stimuli sets. Other studies confirmed that shorter average fixations of experts compared to novices enable them to more effectively interpret task-relevant information (Gegenfurtner et al., 2011).

2.4 Repetition

Repetition refers to the number of views on each relevant item. Decision-making research has used repeated fixations to help define stages of the decision-making process. Russo & Leclerc (1994) suggest that following orientation the first refixation indicates a transition to an evaluation stage. A final verification stage consists of consecutive fixations to a chosen option, using refixations to check for mistakes. Related research examining search among simulated store shelves by Van der Lans, Pieters, & Wedel (2008) provides evidence that repetition is associated with a greater ability to find a desired alternative. Glichrist & Harvey (2000) investigate refixation frequency and memory mechanisms in visual search. They find that participants frequently refixated on objects to compensate for limited functional memory. In the context of risky choice, research also indicates that Repetition decreases with task experience (Pachur, Schulte-Mecklenbeck, Murphy, & Hertwig, 2018) and increases with the difficulty of the task (Pachur, Hertwig, Gigerenzer, & Brandstätter, 2013). When considering multi-attribute judgments, Meißner, Musalem, & Huber (2016) show that decision makers are able to reduce effort and increase accuracy by repeatedly fixating on important attributes and attractive alternatives.

2.5 Incentives

In this paper, we assess the extent to which the roles of the components are stable with different incentive to be fast and accurate. This is important to measure the robustness of these roles in the presence and absence of incentives. Furthermore, it is helpful to investigate whether these roles become more prominent when participants have economic incentives to perform faster and more accurately. We were also interested in examining whether the findings of Ederer & Manso (2013), apply to our context. Indeed, Ederer & Manso (2013) show that managers with financial incentives in a management game write more elaborate notes and perform better by practicing on a subset of earlier trials that do not pay for performance. Rather than just relying on two experimental conditions (with and without economic incentives), we build on this research by defining three conditions. Participants in the control condition are simply asked to do as well as they could on the tasks. Participants in the full-incentive condition get an award if their answers are correct and if they are among the fastest 10% of participants. Participants in a paused-incentive condition can use the first (six) tasks for practice before the incentives become binding. The contrast between the control and incentive groups assesses the general impact of monetary incentives, while the contrast between the paused- and full-incentive groups assesses whether practice tasks encourage a greater learning and performance. Finding that the paused-incentive conditions improve performance would generalize the Ederer & Manso (2013) finding of the value of incentive-free practice.

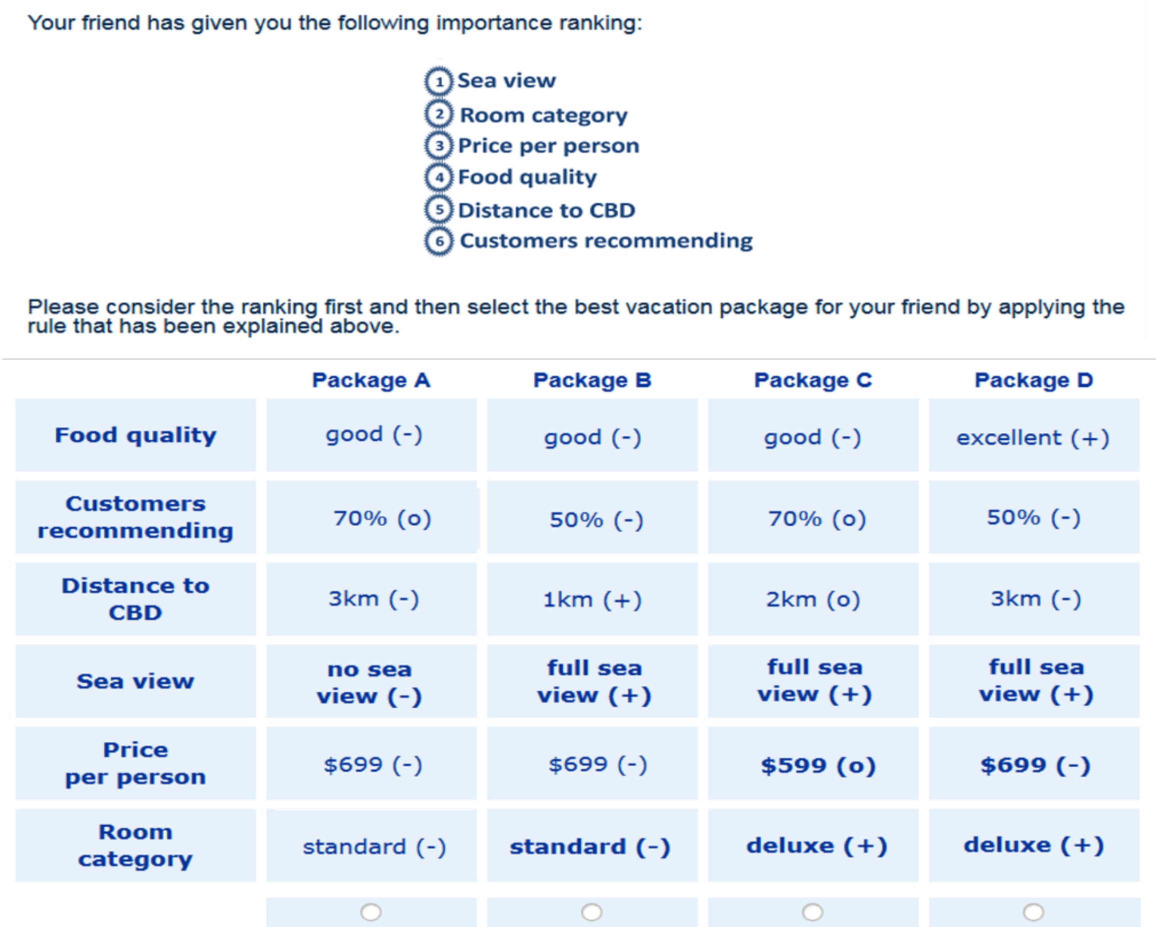
In summary, examining past research, there are studies of rule learning and acquisition from memory, studies of stimulus repetition and fixation duration, and studies demonstrating the efficient scanning of information leading to faster and more efficient choices. However, to the best of our knowledge, there are no studies that jointly measure the four components of Orientation, Wrong Target, Duration and Repetition together and their interaction with each other. Further, most research examining the constructs related to Orientation, Wrong Target, Repetition and Duration use somewhat different definitions than defined below for the lexicographic task. We will provide justification for our particular measures, and then later examine the robustness of the results to different definitions and methods of analysis.

3 Method

3.1 Stimuli

To introduce and develop the proposed effort components of the current study, it is useful to understand how they are measured in our study. Figure 1 provides a typical example of the lexicographic tasks presented to participants. The top of the figure displays the lexicographic rule while the left-hand column provides the locations of the six attributes that characterize four vacation choices.

Figure 1. Example of a task in which participants apply a lexicographic rule


Your friend has given you the following importance ranking:

1. Sea view
2. Room category
3. Price per person
4. Food quality
5. Distance to CBD
6. Customers recommending

Please consider the ranking first and then select the best vacation package for your friend by applying the rule that has been explained above.

	Package A	Package B	Package C	Package D
Food quality	good (-)	good (-)	good (-)	excellent (+)
Customers recommending	70% (o)	50% (-)	70% (o)	50% (-)
Distance to CBD	3km (-)	1km (+)	2km (o)	3km (-)
Sea view	no sea view (-)	full sea view (+)	full sea view (+)	full sea view (+)
Price per person	$699 (-)	$699 (-)	$599 (o)	$699 (-)
Room category	standard (-)	standard (-)	deluxe (+)	deluxe (+)

Note: Participants identified a package that follows the assigned lexicographic rules. In the example, Packages B, C and D have the best Sea View; C and D are both excellent on Room

Category; Package C is cheaper than Package D, leaving C as the best alternative. The minimum information needed for a correct answer are reflected in cells in **bold** above.

We asked participants to imagine that a friend plans to go on vacation at a particular city. The participant's task was to help a friend by selecting a vacation package for her. These vacation packages differ with respect to the six attributes and levels shown in Table A1 in Web Appendix A. Further details on the stimuli, procedure, apparatus and how we analyzed the eye-tracking data are available in Web Appendix A.

To evaluate the vacation packages, the participant is asked to follow a lexicographic rule that reflects the friend's preferences across attributes. The task was designed to be easy but tedious. It requires attention to the rules at the top of Figure 1 and the horizontal location from the attribute labels on the left before making the decision in the grid.

3.2 Procedure

Participants were told that the goal of the experiment was to monitor their attention as they solved several related tasks. A lab assistant greeted and then directed each participant to sit in front of a computer monitor that would present all stimuli. Adjusting the seat height and the remote eye-tracker assured an optimal recording quality of the eye tracker. Participants were also asked to sit relatively still and to solve all tasks without interruption, giving all answers solely with the computer mouse. Then participants provided informed consent. Each participant was randomly assigned to one of the three incentive conditions described in Section 2.5.

The experiment started with a detailed explanation of the attributes and their corresponding levels. The participant's task was to carefully read the information about the attributes and their levels. Then, the participants received information on a friend's priorities in terms of the importance of attributes and their levels, and learned that their task would be to determine the best option for their friend. Two practice tasks then followed. First, the computer showed the participant four alternative vacation packages where only one alternative was best on the most important attribute. The next practice task required a second attribute to break a tie. The process rotated back to the start of the practice tasks until both tasks were successful. Before participants started answering the block of 12 tasks the lab assistant rechecked and adjusted the instrument calibration. Then, the participant learned the incentive structure that would apply in their case.

Within subjects the attribute order and their locations in the grid were unchanged across the 12 tasks. Each subject saw 2 tasks requiring one attribute, 2 tasks requiring two attributes, and so on up to six attributes. The idea is to provide each participant with a full range of effort levels across the tasks. The order of the 12 tasks was however, randomized across participants.

3.3 Apparatus

Eye movements were recorded using the Tobii T120 remote eye-tracking system with a sampling rate of 120 Hz (Tobii Software, 2016). This system is calibrated to have a deviation under 0.4 degrees of visual angle between true and measured gaze direction. The infrared sensors built into a 17’’ TFT monitor have a resolution of 1280 x 1024 pixels. The system adjusts to changes in the seating positions of participants. Accordingly, participants could make moderate movements in front of the computer monitor and a chin rest was not needed.

For this analysis we treated all fixations characterizing the attribute order or their labels as defining one area of interest (AOI). Then the information about each alternative on each attribute is captured by fixations within the 24 cells within the grid, each treated as separate AOIs. We refer readers to Web appendix A for more details regarding the apparatus and analysis of eye-tracking data.

3.4 Participants

In all, one-hundred and ninety-four engineering students (57% male) successfully completed the task. Three incentive conditions were available, with 69, 65 and 60 participants in the control, full- and paused-incentive conditions, respectively. Since all participants were students, we did not ask for their age. 56.6% of students participating were male. The three incentive subsamples did not differ significantly regarding gender ($\chi^2 = 2.31, p = .31$) or the percentage of right-eye dominant participants ($\chi^2 = .11, p = .95$).

3.5 Operational Definitions of the Components of Attention Effort

Below we define four components and two control variables affecting Attention Effort. Attention Effort is defined by the sum of fixation times to all areas of interest in a task rather than time measured by the computer clock. Clock time also includes time for fixations less than 60 milliseconds, saccades, blinks, and attention away from the tracked areas of interest. For our data, clock time is about 8% greater than Attention Effort but across tasks and participants has a .99 correlation with it. Attention Effort shows consistently stronger relationships than clock time with variables of interest like complexity and learning. If our task had involved processes like addition, multiplication, or calling on long-term memory, then attention outside of the screen might be more informative, but with our task there is little need for attention outside the measured AOIs.

We next turn our attention to the operational definitions of the components that make up Attention Effort. While each of these components could be measured using different metrics, the chosen operationalizations allow attention time to be determined as the product of these metrics.

Orientation refers to the effort to understand the rules and guidelines of the task and identify the location of the relevant pieces of information. For our study, we measure Orientation by total attention time on all AOIs divided by the time in the 6-by-4 grid. Greater values of this ratio indicate that a larger fraction of time is spent attending to instructions and the labels of attributes and alternatives.

Wrong Target is effort wasted from examining unneeded information. It is assessed as the ratio between attention time spent in the grid divided by the time on the subset of cells needed to make a correct lexicographic decision. Hence, greater Wrong Target implies less selective information processing evidenced by a greater fraction of time spent on irrelevant information. Rather than focusing on needed cells, it is possible to define Wrong Target as the number of attributes that are accessed in the wrong lexicographic order (Khader et al., 2016). Our measure of Wrong Target is preferred for our task because it not only penalizes attention to attributes in the wrong order, but also the acquisition of information about alternatives that should have been excluded.

Duration is the average time spent attending each piece of relevant information, measured as the ratio of the time on the needed cells over the number of fixations on those cells. Duration could have been measured by average fixation duration of all fixations in a task, rather than just those required for the correct decision. The decision to base Duration on needed cells follows from work suggesting that effort on needed cells decreases with experience (Gegenfurtner et al., 2011; Orquin & Mueller Loose, 2013).

Repetition, the fourth component, is defined as the number of repeated fixations per needed cell and it is hence measured as the ratio between the number of fixations to needed cells and the number of needed cells. Focusing on these cells emphasizes the functional aspect of repetition in this task reinforcing the location of important attributes and can be used to check the accuracy of the choice. Our focus on Repetition for needed cells is therefore in line with earlier research by Glichrist & Harvey (2000) who investigated search task in which only relevant stimuli were provided.

Complexity is a critical control variable characterizing effort required within each of the 12 tasks. It is operationalized as the minimum number of cells out of 24 (6 attributes x 4 alternatives) needed to identify the best alternative. In the Figure 1 example, Complexity is 4+3+2 =9. Complexity varies randomly between 4 and 21 cells across tasks. We note that one could operationalize complexity considering both the number of needed attributes and the number of needed alternatives per attribute. This alternative definition yielded very similar insights.

By design, the four components and Complexity mathematically determine Attention Effort, such that:

Attention Effort = Orientation *Wrong Target *Duration *Repetition *Complexity (1)

where

Orientation = Attention Effort / Time on grid

Wrong Target = Time on grid / Time on needed cells

Duration = Time on needed cells / Fixations on needed cells

Repetition = Fixations on needed cells / Number of needed cells

Complexity = Number of needed cells

There are several advantages to building components which when multiplied together with Complexity perfectly predict Attention Effort. First, having multiplicative components accounts for all the variance in Attention Effort. Further, the structure permits a separation of *direct* and *net* effect on Attention Effort due to a change in each component. The *direct* effect is simply the percent change in Attention Effort from a change in one component, assuming that a change in any component has no effect on the others. If instead we allow components to alter each other within the task, we can estimate a *net* effect that incorporates cross effects between components. We will show that allowing for a correlation among components leads to both better prediction of the components of effort and better insight about their interrelationships.

Accuracy is not directly part of the model since over 96% of the tasks were done correctly by participants. Our high accuracy rates reflect similar levels found in Bettman et al. (1990) or Khader et al. (2016). However, we will measure accuracy to see if accuracy shifts with experience or is significantly related to the components of Attention Effort.

4 Modeling Individual Participant Effort Components

We formulate a statistical model which has three main features. First, it is a joint model of the four attentional components. Second, it allows for heterogeneity among participants, and finally it controls for learning and task difficulty. We elaborate and justify each of these below.

Joint Model. It is important to note the four components are defined as interleaved ratios where the numerators and denominators of one component are linked to the next component. This formulation, which produces a multiplicative decomposition of Total Attention Effort, may induce artefactual correlations among the components. It is, however, an empirical question whether these components are correlated. Our formulation addresses this issue by jointly modeling all four components allowing them to be correlated with each other.

Task Experience and Complexity. A goal of our research is to determine the extent to which the four components vary with practice or the difficulty of a task. Previous studies (e.g., Meißner et al., 2016) suggest that participants facing repeated tasks decrease task time with experience. We measure the rates of change on each of the four effort components that participants achieve as they accumulate more experience.

Finally, our design varies the difficulty of the tasks faced by each participant. In some tasks, the participant may find the best alternative by inspecting only a few cells, while in others the same participant may be required to access most of the cells. These different requirements may affect the effort of the participant along the four components. We control for these differences in task difficulty by using the Complexity measure defined earlier. Since Complexity varies randomly across tasks and participants, including it as a covariate substantially reduces the error term.

Throughout, the joint model is formulated with all variables transformed to natural logarithms. Working in a log space has four benefits. First, since the relationships between the components are multiplicative, then the logs additively decompose the log of Attention Effort. Second, that additivity generates estimates that can scale across different levels of analysis, so that the coefficients for a number of pooled analyses will be identical to those averaged across individuals. Third, there is substantial evidence that learning effects are robustly accounted for with a power model of logged experience (Anderson & Schooler, 1991; Ritter & Schooler, 2001). Finally, with the current data, the distributions of the four component measures across decision makers are positively skewed. Taking logs substantially reduces that skewness for most components.[1] Shapiro-Wilk tests indicate that the distributions of all logged variables, except orientation, do not significantly differ from normality.

[1] Attention Effort and four of its five components become more normally distributed when logged. The only exception is Orientation which is slightly more normal in raw compared with its logged form.

Consistent with these features of the model, we define y_{ift} as the effort allocated to attention component factor f by participant i at task t, with f = {o, w, d, r}, where each of these elements refers to Orientation, Wrong Target, Duration and Repetition. We also control for the complexity (C_{it}) of task t for participant i to adjust for the effort allocated as a function of the minimum number of cells needed to identify the best alternative. Thus:

$$\ln(y_{ift}) = \beta_{if} + \gamma_{if}\ln(t) + \delta_f \ln(C_{it}) + \epsilon_{ift}, \qquad (1)$$

where: $\epsilon_{it} \equiv (\epsilon_{iot}, \epsilon_{iwt}, \epsilon_{idt}, \epsilon_{irt})'$ and $\epsilon_{it} \sim N(0, \Omega)$.

The intercept β_{if} represents the baseline allocation to attention factor f by participant i. The logarithms of learning ($\ln(t)$) and Complexity ($\ln(C_{it})$) are zero centered to facilitate interpretation of the model parameters. Thus, the intercept β_{if} reflects the effort level for component f of individual i at the geometric means of task number and Complexity. The parameter γ_{if} measures the degree of learning of subject i with respect to component f, since it is an estimate of changes in effort associated with a change in task experience. The coefficient δ_f controls for the impact of Complexity on component f, while ϵ_{ift} represents fluctuations in component f after controlling for individual differences, learning and Complexity.

These fluctuations measure whether a participant places a stronger (or weaker) emphasis on a particular component for a specific task. We use a general covariance matrix for these fluctuations (Ω) that accounts for the fact that components reflect different ratios of fundamental variables which, in itself may generate correlations among components. The Seemingly Unrelated Regression (SUR) framework allows the concatenation of different effort components across participants. This approach of combining several equations into one model to improve estimation efficiency was proposed by Zellner (1962) and has been cited as one of the most successful and lasting innovations in econometrics (Griffiths, 2003).

Our model also allows participants to be heterogeneous in their mean effort levels (β_{if}) for each component. Rather than using a fixed effects approach for the baseline (β_{if}) and learning parameters (γ_{if}), a random-coefficients model simultaneously estimates heterogeneity across participants in terms of the baseline (β_{if}) and learning parameters (γ_{if}). Accordingly, we denote by $\beta_i \equiv (\beta_{io}, \beta_{iw}, \beta_{id}, \beta_{ir})$ and $\gamma_i \equiv (\gamma_{io}, \gamma_{iw}, \gamma_{id}, \gamma_{ir})$ and let $\beta_i \sim MVN(\theta_\beta, V_\beta)$ and $\gamma_i \sim MVN\left(\theta_\gamma, V_\gamma\right)$; where V_β is a full variance-covariance matrix and V_γ is a diagonal variance matrix.[2]

Web Appendix B details the Bayesian estimation. We use a Bayesian procedure implemented in Stan (Carpenter et al., 2017) to estimate the parameters. The Bayesian estimation approach has been used in other studies using eye tracking. For instance, Wedel, Yan, Siegel, & Li (2016) investigate the extent to which eye movements reveal effective strategies in physician search for lung nodules in x-rays.

The SUR analysis relies on a joint model of the four components and allows participants to be heterogeneous in their mean effort allocated to each component and in their learning curves. To test the incremental value of the SUR analysis, we estimated a simplified version of our main model where each component is modeled independently, thus eliminating the SUR error structure. The test of the original model generated a very strong support for our SUR model with a 694 Bayes factor in favor of our model over one that separated the component analyses.

The estimated model parameters can assess the direct and net effects of each component factor *f* on total attention effort. The direct effect measures the extent by which total effort changes when one of the components is increased, assuming all other components

[2] Using a full variance-covariance specification for $\boldsymbol{V_\gamma}$ increases the number of parameters to be estimated, increasing the complexity of the model, but leads to similar results.

remain unchanged. When determining this direct effect, a reasonable change in a component can be obtained from its standard deviation across participants:

$$Direct_f = Stdev(\beta_{if}) = \sqrt{V_{\beta_{ff}}}. \tag{2}$$

In contrast with the direct effect, the net effect considers the associations across components. For example, a participant who allocates more effort to Orientation might avoid task-irrelevant information and hence demonstrate a lower value of Wrong Target. To account for these associations, the net effect measures how a change in one component affects total attention effort not only via changes in that component but also through other components.

Further, the net effect in an additive model takes a particularly easy form that depends on the direct effects weighted by the correlation between each pair of components. Web Appendix B derives this result.

$$Net_f = Direct_f + \sum_{f' \neq f}^{F} Direct_{f'} * Corr(f', f) \tag{3}$$

Finally, our model estimates also give us insights in terms of learning along each of the attention components. If γ_f is the coefficient of task time in log scale, the percent change in factor f over the 12 tasks can be measured as:

$$\%\text{ change in } f = \gamma_f * \ln 12 \tag{4}$$

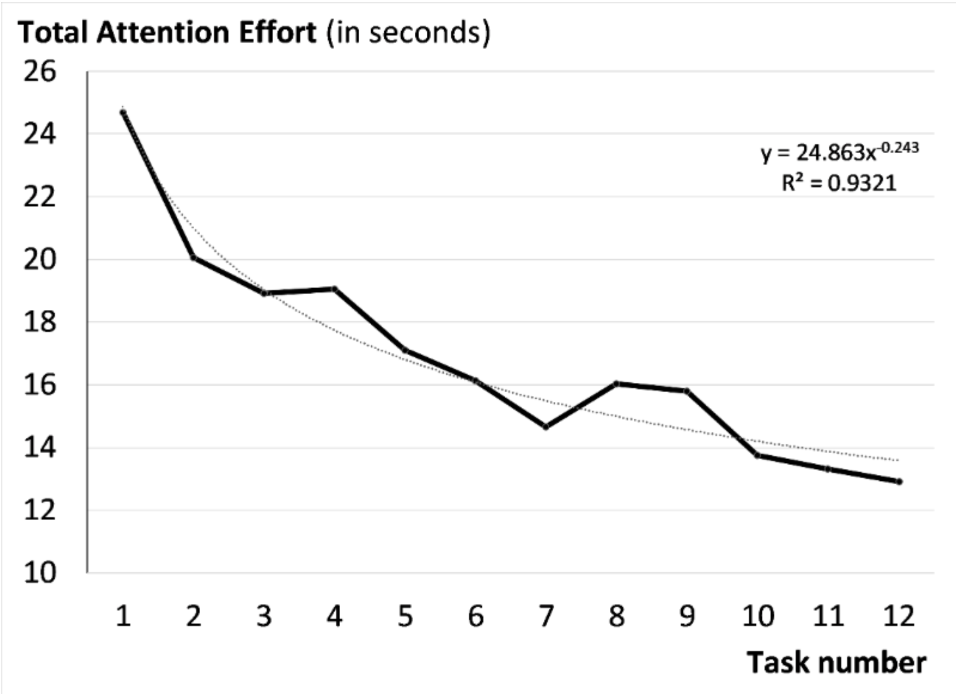
5 Results from the Joint Model

5.1 Total Attention Effort

Figure 2 provides the time path of average Attention Effort across 194 participants, measured as the sum of fixation times to all AOIs in a task. The fitted power function closely approximates the actual data and generates an R^2 =.93, substantially better fit than the R^2 =.85

of a linear relationship. The constant term indicates that Attention Effort for the first task averages nearly 25 seconds while the fitted power function shows that the average drops to around 13 seconds by task 12, consistent with participants learning how to complete each task with less effort. That -.243 coefficient generates a 55% drop in effort across 12 tasks.

Figure 2. Attention Effort across 194 Participants



5.2 Effort Components

Detailed estimation results are provided in Web Appendix C. Table 1 gives the correlation matrix of the four effort components across participants derived from the estimated variance-covariance matrix, V_β.

Table 1. Derived Correlations among Effort Component Intercepts

	Orientation	Wrong Target	Duration	Repetition
Orientation	1.00			
Wrong Target	-0.54*	1.00		
Duration	-0.36*	0.15	1.00	
Repetition	-0.17	0.00	-0.03	1.00

*p<.05, n=194

Table 1 shows that the only significant correlations involve Orientation. It is significantly and negatively associated with Wrong Target and Duration and marginally with Repetition, implying that participants who pay attention to the rules and labels have fewer fixations on uninformative cells, spend less time per fixation and have (marginally) fewer repetitions on needed cells. Table 2 shows the means, rates of change, along with the direct and net effects of each component.

Table 2. Means, Confidence Intervals, Changes, Net and Direct Effects of Effort Components from Bayesian Analysis

Components	**Mean**	**2.5%**	**97.5%**	**% Change from task 1 to task 12**	**Direct Effect**	**Net Effect**
Orientation	2.06	1.45	2.91	-17%	19%	1%*
Wrong Target	1.34	1.03	1.74	-4%	14%	6%*
Duration	0.32	0.23	0.44	-6%	18%	12%*
Repetition	2.02	1.32	3.10	-28%	24%	20%

Note. Mean estimates are approximated as the exponential of the mean of the intercepts for each component. Direct and Net Effects are obtained using equations 2 and 3, respectively and then expressed as a fraction of their corresponding means. Confidence intervals give estimates of the heterogeneity among participants for each component. * indicates a significant difference between net and direct effects at p<0.05.

The means and confidence intervals across subjects are transformed back from log form to be consistent with their original definitions. Orientation is the ratio of total time over grid time. A mean of around 2, indicates that about as much time is spent on instructions and labels than on information within the 6 x 4 grid, with the low end of the range spending about 50% and the upper end at 200% of grid time. Wrong Target relates to the proportion of time on unneeded information. A score of 1.34 implies that 34% of fixations are not needed, with 95%, ranging between 3% to 74%. Duration reflects the number of seconds per fixation on needed cells. The average fixation takes around 1/3 of a second with a 95% range across participants between approximately between 1/5 and 4/10 of a second. Finally, Repetition is measured by the total number of fixations on needed cells divided by the number of needed cells. A value of 2.02 indicates that each needed cell is accessed twice on average. There is high variance about that estimate shown by the fact that almost 95% of participants averaging between 1.32 and 3.10 fixations per needed cell.

The next column uses Equation 4 to estimate the change in effort for the average respondent for each component across the 12 tasks. Repetition drops by 28% and orientation drops by 17%. It makes sense that Repetition and Orientation would decrease with practice. Once a participant understands the meaning of the information in the grid then there is less need to review the rules or repeat access to the same information. By contrast, both Wrong Target and Duration show smaller improvements of 6% and 4% respectively, suggesting that there is either less motivation or ability to reduce effort for Wrong Target and Duration. Because of the multiplicative relationship among the components, the percent changes can be combined yielding a 55% drop in total Attention Effort.

Now consider direct and net effects in the last two columns of Table 2. The direct effect of a standard deviation change in each component follows from Equation 2 transformed back from log form. The greatest direct effect of 24% is from Repetition.

Orientation and Duration follow with 19% and 18% respectively, followed by Wrong Target at 14%.

These direct effects suggest that examining Repetition would be the best way to identify participants with the fastest task time. However, the direct effects assume components do not impact each other. Equation 3 estimates the net effect by weighting each component by its correlation with the others. For example, the net effect of a standard deviation shift in Orientation from the correlations in Table 1 and the direct effects in Table 2. Table A4 of Web Appendix C provides detail.

Net Effect (Orientation) = Direct Effect + through {Wrong Target, Duration, Repetition}

Net Effect (Orientation) = .18 + .13*-.54 + .17 *-.36 + .22 *-.17

= .01

The role of Orientation is clarified by the contrast between its direct effect of 19% and its net effect of 1%. Subjects with higher levels of Orientation display lower levels of the remaining three components, dropping the net effect of standard deviation shift in Orientation to a negligible change on Attention Effort. A similar pattern is observed for Wrong Target, where a direct increase of 14% is reduced to a non-significant net effect of 6%. For the other two components (Duration and Repetition), a similar but less extreme result is obtained. Their direct effects (+18% and +24%, respectively) are only partially compensated by their indirect effects, and their resulting net effects remain significantly positive (+12% and +20%, respectively).

The joint model shows that the correlation among components generally decreases the net time of other components as they substitute each other. To assess the generality of these results, it is useful to see if the processes associated with the components differ across the three incentive conditions.

5.3 Impact of Performance Incentives

The previous analysis of 194 participants pools data across three incentive assignments. By running the joint models within each incentive condition, it is possible to assess the degree to which the results generalize. Recall that the control condition provides no incentive for speed or accuracy, the Full-Incentive condition rewards relative speed for those with 100% accuracy across tasks, and the Paused-Incentive condition allows the first six tasks to be treated as practice that could help prepare a participant for later trials.

Table 3 displays the mean, direct effect, net effect and change from tasks 1 to 12, for each condition using the same analysis as used in the pooled analysis. Generally, the results across conditions are very similar. An asterisk after the control condition indicates a significant difference between control and the average across the two incentivized conditions. An asterisk after the Paused-Incentive condition indicates a ($p<0.05$) significant difference from the Full-Incentive condition. Given the 32 comparisons, even with no effect at least one ($p<.05$) is expected by chance.

Table 3 suggests that the components are for the most part robust to the incentive conditions. The most remarkable exception corresponds to the decrease in mean repetition levels as economic incentives are provided to participants. To study the evolution of the attention components under the different incentive conditions, it is useful to examine their time path across the 12 tasks. The estimates from the graphs below derive from a two-step process. The Bayesian analyses in log space predicted each component for each participant with centered task number and complexity. Then we adjusted each independent variable to reflect its value assuming that its complexity was at its average level. The graphs then show the complexity-adjusted average scores across participants in each condition. Across the components, the figures suggest intriguing differences. Figure 3 provides a graph for

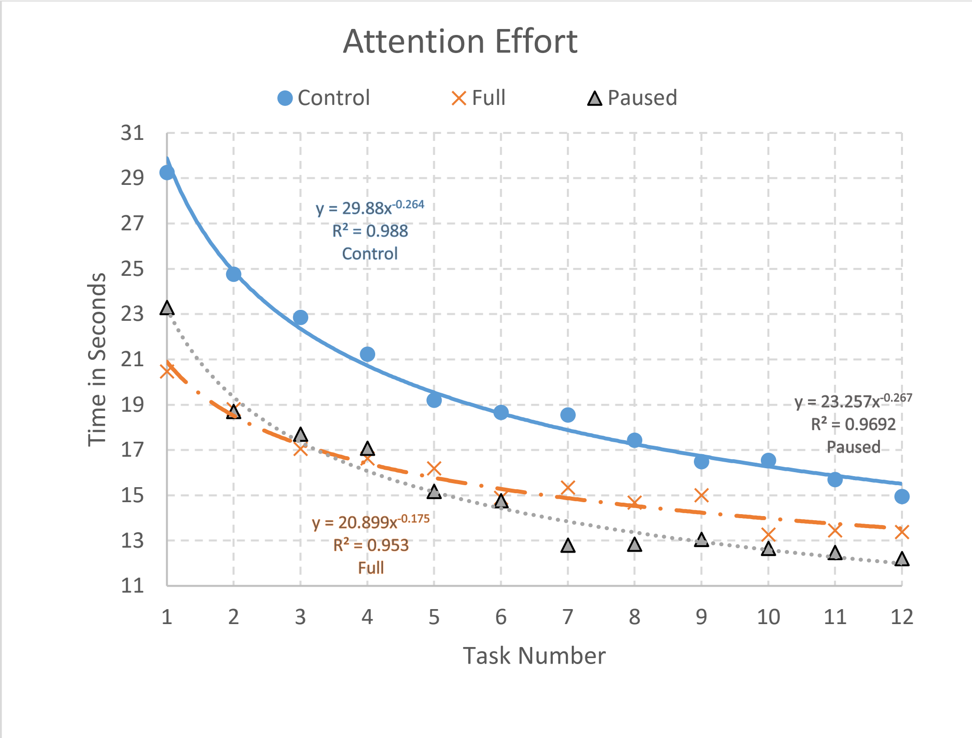
Attention Effort. Confidence intervals for the values in Figures 3-7 are available in Web Appendix F.

Table 3. Effort Components across Incentive Conditions

	Orientation	**Wrong Target**	**Duration**	**Repetition**
Mean				
Control	2.08	1.36	0.32	**2.27***
Full-Incentive	2.04	1.35	0.32	1.86
Paused-Incentive	2.05	1.31	0.31	1.94
Direct Impact on Total				
Control	18%	16%	20%	26%
Full-Incentive	23%	16%	20%	33%
Paused-Incentive	19%	10%*	18%	28%
Net Effect on Total				
Control	-9%	15%*	21%*	22%
Full-Incentive	0%	-5%	12%	25%
Paused-Incentive	6%	5%	-1%	18%
Change from Task 1 to 12				
Control	-20%	-4%	-4%	-32%*
Full-Incentive	-14%	2%	-6%	-25%
Paused-Incentive	-16%	-11%*	-8%	-25%

Notes: * indicates a non-Bonferroni corrected significance at p <.05, while bold indicates significance after applying a Bonferroni correction for multiple comparisons. Asterisks above Control indicate that it is significantly different from the average incentive condition, asterisks above Paused-Incentive indicate that Paused- is significantly different from Full-Incentive condition. Mean values are approximated as the exponential of the mean of the intercepts for each component. Direct and Net Effects are obtained using equations 2 and 3, respectively and then expressed as a fraction of their corresponding means. As an alternative to the Bonferroni correction, using the output of the Bayesian estimation, we verify that all tests marked with a * (except those for Net effect totals) are jointly significant with more than 95% posterior probability. Full model results are available from the authors upon request.

Figure 3. Impact on Attention Effort from Incentives and Task Experience



The incented conditions take significantly less total time compared with control. The lines provide the fitted power function, where the exponents reflect the percent change in response to a percent change in experience. All conditions exhibit a high fit of attention effort as a function of task experience. The difference in change scores demonstrates that the average time in the first task is almost 33% (29.2 vs. 21.9 seconds) higher in the control condition than in the incentive conditions, but that difference drops to 16% (14.9 vs.12.8 seconds). Put differently, the participants without incentives initially spend more time than those with incentives, but are able to significantly lessen that gap with practice (t=2.937, p<.01).

Additionally, those in the Paused-Incentive condition (compared to the Full-Incentive condition) have greater Attention Effort in the first task (t=2.252, p=.026) followed by lower attention in tasks 7 to 12 resulting in a greater change in the Paused- over Full-Incentive condition (t= 4.405, p<.01). This pattern is consistent with an exploration-exploitation strategy (Peterson, Hammond, & Summers, 1965; Lauretro, Stefano. Canessa, & Zollo, 2015). Exploitation in our context corresponds to the use of the knowledge gained in the six practice tasks to enable better performance in six later tasks. Evidence of exploration-exploitation in the paused incentive condition take different forms as we explore the time paths of the four components of Attention Effort, beginning with Repetition.

Figure 4. Repetition across Incentives and Task Experience

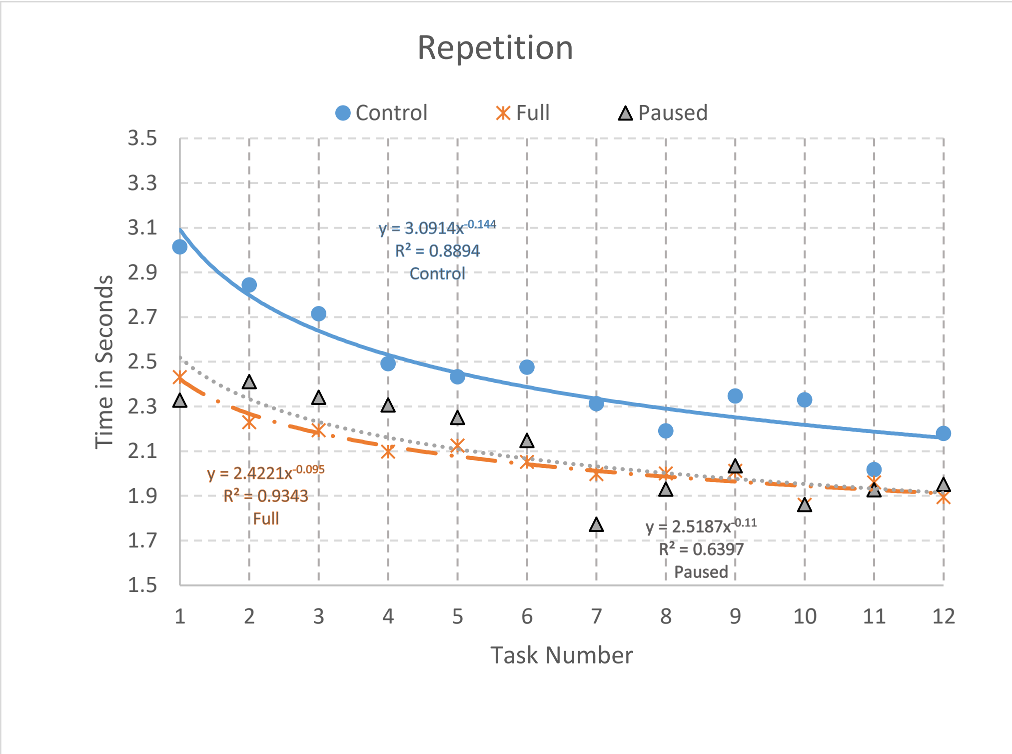


Figure 4 graphs Repetition across incentives and task experience. The control and full-incentive conditions exhibit a high fit of repetition as a function of task experience, while the paused condition shows a moderate fit. The Paused-Incentive condition demonstrates slightly greater change across the 12 tasks compared to the Full-Incentive condition, similar to the pattern for Attention Effort, but the difference is not significant (t=1.204, p=.229). Note, however that in five out of the first six tasks average Repetition for the Paused-Incentive is higher than the Full-Incentive condition, but the reverse happens in four out of the last six tasks. Thus, a simple sign test provides partial evidence for the hypothesis that practice encouraged more exploration earlier and more exploitation later once incentives kicked in.

Figure 5. Duration across Incentives and Task Experience

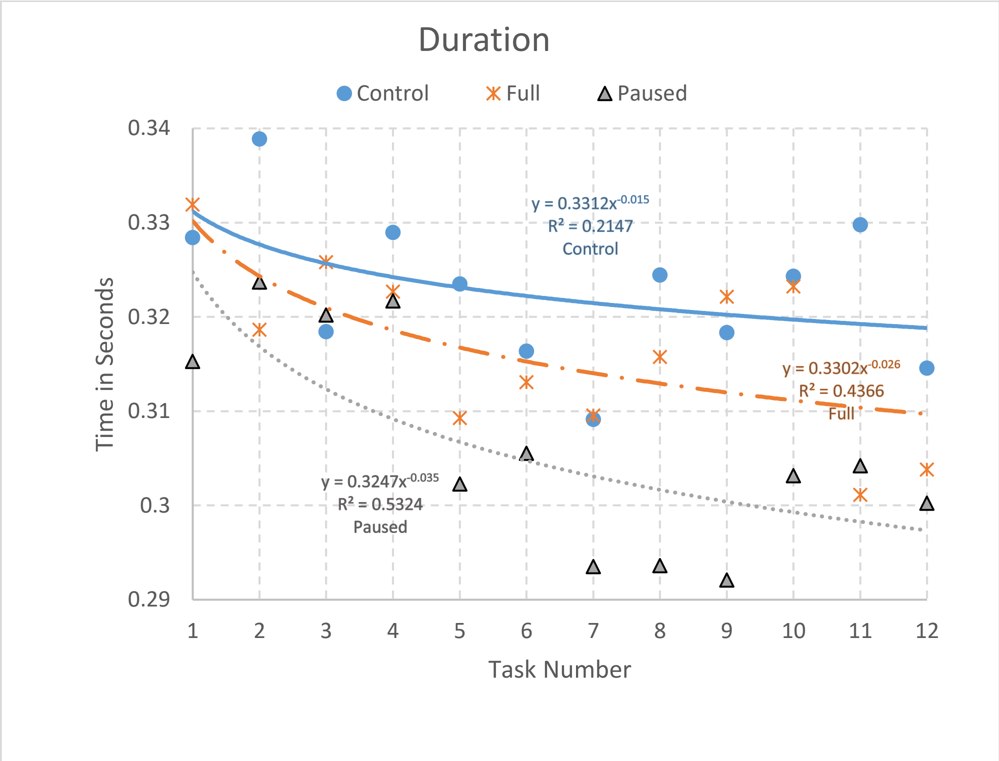


Figure 5 provides the pattern for the Duration effort component. All conditions exhibit a lower fit as a function of task experience, when compared to those obtained for Repetition. This implies weaker improvements in Duration with practice. All three conditions generate similar times per fixation in the first four tasks, but for seven of the next eight paused-incentives fixations take less time than either the control or full incentives. Participants in the Paused-Incentive condition may have been able to learn better how to process the acquired information in the early rounds when speed was not incentivized and that enabled them to increase speed later. However, a general test of difference in learning across 12 tasks between conditions is not significant (Paused vs. Full: t=.75, p=.45; Paused vs. Control: t=1.76, p=.08), reflecting substantial variances within each condition.

Figure 6. Wrong Target across Incentives and Task Experience

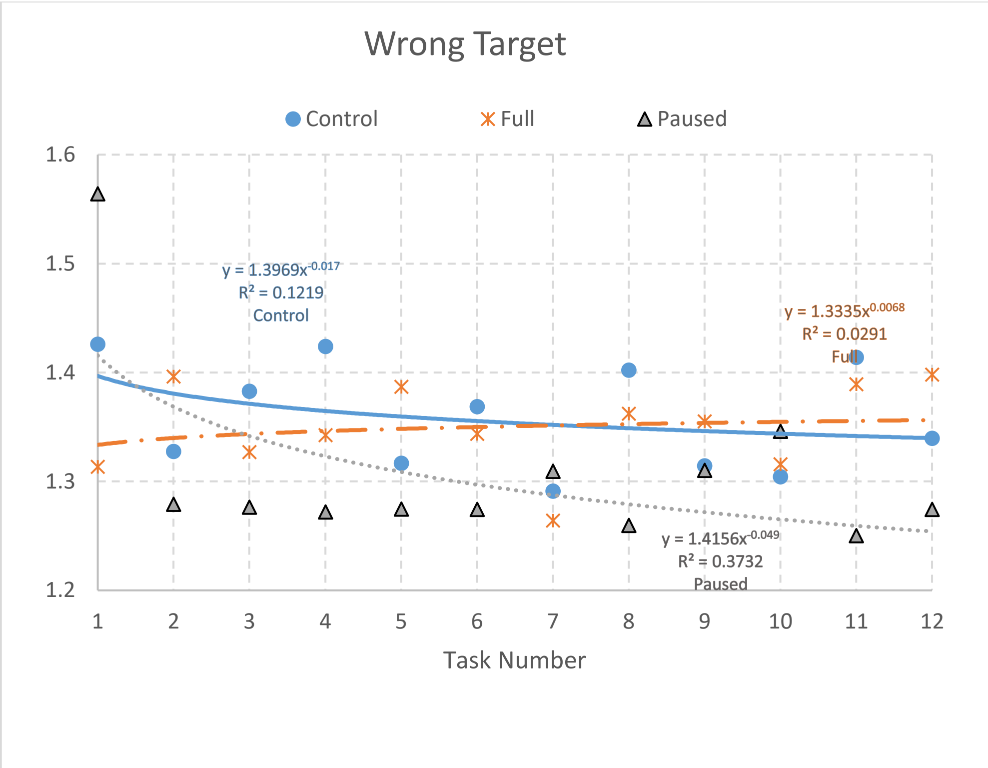


Figure 6 shows a unique pattern for Wrong Target. This figure shows little evidence of improvement with experience, as reflected by the low fit of each of the three models. The analysis indicates that those in the Paused-Incentive condition had greater improvement compared with those in the Full-Incentive condition. Much of that result comes from a high Wrong Target performance in the very first task and consistent low values thereafter. Focusing on the grid in that task may have helped participants to explore different relationships among the data points, enabling them to learn about the labels and locations of important attributes. This behavior may be similar to exploratory behavior (Hardy et al., 2014) which generally increases learning.

Figure 7. Orientation across Incentives and Task Experience

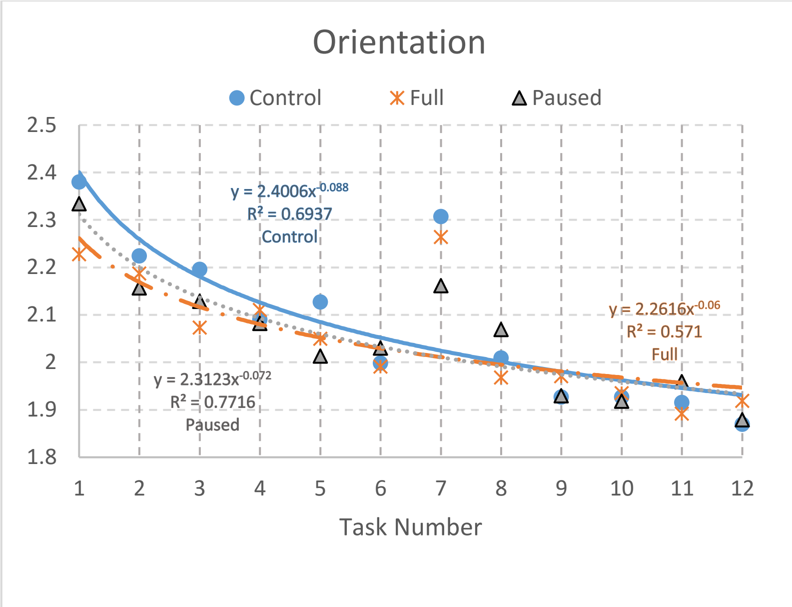


Figure 7 gives the pattern for Orientation across incentives. Orientation shows improvements with experience. These improvements are consistent with the relatively higher model fits obtained for this component. There are no statistical differences between the three conditions with respect to mean levels or changes across 12 tasks. There is, however, a consistent increase at task 7 that was generated by an introduction of the next tasks after task 6. That pause encouraged a review of the rules and attribute positions that increased Orientation. The impact of that information is weaker for those in the Paused-Incentive condition because they had been told to expect a change after task 6. However, that difference in task 7 is not statistically significant (Paused vs. Full: t=0.79, p=.43).

The major purpose of manipulating incentives was to establish the extent to which the effort component results apply across performance incentives. Overall, we find surprising consistency in the distributions of effort for the components across incentive conditions. As shown on Table 3, correlations among the components and the estimations of direct and net effects on total effort are very similar.

However, graphs provide promising evidence that participants in the Paused-Incentive condition increased effort during the practice tasks and reduced effort later. An alternative mechanism is that the pressure to perform may have interfered with learning for those in the Full-Incentive condition. If so, practice without performance incentives may provide for thinking smarter that later generates thinking faster (Einhorn & Hogarth, 1986; Tversky & Kahneman, 1986).

5.4 Impact of Incentives on Accuracy

As mentioned earlier, errors, defined as identifying the incorrect alternative, occur approximately at a rate of one in 20 tasks. It is not appreciably improved with practice across 12 tasks. The correlation between task number and error is -.04. By contrast the correlation of task number with Attention Effort is -.34. These results are consistent with research showing that decreasing attention time is relatively easy and common (Kool, McGuire, Rosen, & Botvinick, 2010; Shah & Oppenheimer, 2008). Additionally, Bettman et al. (1990) showed that decision makers attend more to effort reduction than accuracy minimization. Todd & Benbasat (1992) suggest a reasonable processing mechanism accounting for the primacy of effort reduction over error reduction: "Effort may be weighed more heavily than accuracy because feedback on effort expenditure is relatively immediate, while feedback on accuracy is subject to both delay and ambiguity" (p.375).

Looking at error across individuals, the average error level is 4.9% in the Full-Incentive condition but is 2.5% in the Paused-Incentive condition. That difference is marginally significant at $p<0.10$ level. Moreover, a supportive pattern occurs with the association of error levels and the different components of effort. Table 4 displays the correlations of error for each participant with their estimated intercepts for Orientation, Wrong Target, Duration, and Repetition.

Table 4. Correlation of Effort Components with Error

Incentive condition	Orientation	Wrong Target	Duration	Repetition
All pooled (n=194)	**-0.33***	**0.59***	0.16*	0.01
Control (n=69)	**-0.39***	**0.61***	**0.38***	0.28*
Full-Incentive (n=65)	-0.31*	**0.64***	-0.05	-0.14
Paused-Incentive (n=60)	-0.31*	**0.44***	0.23	-0.04

Notes. * Indicates non-Bonferroni corrected significance at $p < .05$, while bold indicates significance after applying a Bonferroni correction for multiple comparisons.

Both Duration and Repetition are not consistently related to accuracy. When pooling all conditions, participants with higher levels of Orientation have significantly lower levels of error ($r=-.33$), while Wrong Target is positively correlated with error ($r=.59$). Simply put, those with greater Orientation make fewer errors, while those with high levels of Wrong Target make more errors.

It is important to be cautious about generalizing these results from error levels. Errors occur for less than 6% of all tasks, and more than 72% of the participants have no errors at all. Thus, reliable results about the drivers of accuracy await studies that are more inherently error-prone and studies that separate incentives for accuracy and speed, rather than combining them together.

5.5 Functional Roles of the Effort Components

The pattern of results demonstrates that the four effort components have different functional roles. Briefly, Orientation is unique in enabling the other components to do their job more efficiently. Wrong Target, by contrast, provides a robust measure of dysfunctional processing that is difficult to alter with practice. Duration differs strongly across people, but like Wrong Target, is difficult to change with practice. Finally, Repetition varies most strongly across people, but it is associated with the greatest net impact on Attention Effort, and the one that declines the most with experience. Below we provide more detailed accounts of these generalizations.

5.5.1 Orientation

Orientation is central because it is the only component that is significantly associated with a reduction in error and effort. As shown in Table 1, its negative correlation with Wrong Target (r=-.54) implies that greater effort in Orientation is associated with more effective positioning of fixations on relevant pieces of information. Its negative correlation with Duration (r =-.36) means that each fixation takes less time, and its marginally significant negative correlation with Repetition (r=-.17) implies that those with strong Orientation make fewer fixations on each needed cell. Consequently, participants one standard deviation greater on Orientation have a direct 19% increase in Attention Effort, which is fully compensated by the effort reduction in the three other components, yielding a negligible net change on Attention Effort.

5.5.2 Wrong Target

Wrong Target measures inefficiency coming from a tendency to focus attention on irrelevant information. Participants in our study did not differ substantially on this component. The 14% direct effect on Attention Effort from a one standard deviation shift in Wrong Target is significantly lower than the direct effects for the other three components (p = 0.02). Wrong Target has the strongest negative correlation with Orientation (r=-0.54). With practice, Wrong Target does not change much, dropping a non-significant 4% across 12 tasks for the average participant. Its lack of variance across participants and relatively slow response to experience suggest that it is hard to overcome poor attention strategy, either within or between decision makers. Like multiple-cue weighting strategies, it is hard to alter learned strategies when they are wrong (Peterson et al., 1965). In sum, Wrong Target has a unique role among the components as a consistent measure of misguided but persistent Attention Effort.

There is more than just an effort cost associated with Wrong Target. Decision makers with high wrong target scores are also more likely to make errors. The correlation across decision makers between number of errors and Wrong Target is 0.59. Across the different incentive conditions, focusing on unneeded information is a strong predictor of making mistakes.

Our measure of Wrong Target shares findings from a number of eye-tracking studies that assess attention to irrelevant information (Gegenfurtner et al., 2011; Haider & Frensch, 1999). These studies demonstrate that inappropriate attention is strongly associated with low expertise across a number of fields, results that are consistent with the idea that Wrong Target provides a durable measure of inefficient attention processing.

5.5.3 Duration

The time taken for a fixation can vary substantially across decision makers. More than 95% of the sample had average fixation times between a one-fifth and two-fifths of a second. However, the direct effect on Attention Effort from a one standard deviation shift in Duration is 18% and its net effect is a still significant 12%. Duration is also stable over time, its decline across the 12 tasks is not significant for the average participant. While Duration is significantly lower for those with higher Orientation, its correlations with Wrong Target and Repetition are not significant. Thus, the analogy of Duration with clock speed is fitting in this context. Participants may differ by a factor of two in their clock speed, but compared with other components a standard deviation shift in Duration has less impact on Attention Effort and it is less amenable to improvements over time.

5.5.4 Repetition

We were surprised by the magnitude of the 2.02 mean Repetition score showing that needed cells are viewed on average more than twice. Only a few studies document a measure of Repetition. However, Reutskaja et al. (2011) show that in a choice of product images refixations occurred in about 25% of the trials. Repetition is useful for the participants when applying the lexicographic rule in two ways. First, it is helpful in verifying a relationship between alternatives. Second, it facilitates checking the correctness of information processed earlier (Russo & Leclerc, 1994).

Participants one standard deviation higher in Repetition have an average 24% increase in Attention Effort. That increase is significantly greater than the corresponding changes in the other components. In addition, Repetition is only marginally associated with Orientation and not reliably related to Wrong Target or Duration. Participants are generally good at learning to avoid Repetition. It has the greatest decline with experience, dropping 28% across 12 tasks for the average participant, demonstrating that decision makers have less need to revisit information after they come to understand its implications. In sum, while Repetition may be the variable that has had the least empirical investigation, it is the largest driver of effort in our study, and generates the greatest improvement in response to practice.

For more substantial decisions Repetition can be both apparent and quite aversive. Consider what happens if one has to go over the same information many times or if one gets lost and has to retrace steps. For our task it appears that Repetition does not reach high levels of conscious awareness. It may be one of many tasks the brain effectively manages automatically. We find it surprising that that greater improvement in decision speed comes less from attention to the rules and format, or avoidance of unneeded information, or shorter duration of fixations, but instead largely flows from decreases in repeatedly accessing the same information.

6 General Discussion

This paper proposes a framework for studying and decomposing attention effort of individuals performing repeated tasks. Our results best apply to structured repeated tasks such as identifying a medical tumor from a photographic image, selecting a candidate from resumes, or choosing the best route on a map. These tasks involve following rules to make a series of decisions requiring attention to and processing of specific information. Within that domain we believe our results provide the following conclusions about the components of effort that lead to greater efficiency.

1. *Knowing the rules and where to find important information are critical for performance.* There is general agreement that learning has a central role in effectiveness (Langley & Simon, 1981). In our study participants with high Orientation levels are more likely to attend to relevant information and answer correctly. People who pay attention to the rules take less time in each single fixation and repeat fixations slightly less often. The manipulation of incentives generated intriguing effects on all components except for Orientation. Given its central role on performance it would be valuable to directly increase either the incentive to attend to the rules or the ease of getting that information. In our study all participants had to successfully make two simple trial runs (see Web Appendix A), but manipulating the number and complexity of those runs might confirm the benefits of greater effort in Orientation.
2. *Attention paid to less useful information is a strong predictor of poor performance.* Wrong Target is a measure of a tendency to focus on less relevant information and is negatively associated with accuracy. It has relatively low variability across participants

and on average shows minimal improvement with practice. Generally, Wrong Target differs little across incentive conditions, with one surprising exception. For those in the Paused-Incentive condition, the first task has greater focus on Wrong Target and then that decreases effort in later tasks. It is thus possible that the exploratory effort in that first task enabled more efficient exploitation later.

3. *Repeatedly viewing the same information is a dominant source of attention effort in early tasks.* Repetition accounts for the greatest variation in total effort across people, even after accounting for its complementary impact on other components. It also shows the greatest drop with practice, suggesting that early repetition increases speed of recognition and understanding later. Both initial Repetition and its reduction are greatest for participants in the control condition. Among those with incentives to perform quickly and accurately, having tasks for practice increases early Repetition as an investment in learning, and decreases it later when performance matters.
4. *Fixation duration is independent of the other components and shows little consistent shift with practice.* Average fixation duration varies from 200 to 400 milliseconds across participants. However, its change across the 12 tasks tends to be less than 10% compared with changes twice as large for Repetition or Orientation. Further, its change is not significant for the average participant. In all, like clock speed in computers, Duration once set is relatively difficult to change.

These results suggest important questions for future extensions. This research shares three methodological characteristics that are rarely aligned in one study. First, each task has a unique correct response; second, it uses eye-tracking to provide detailed measures of attention, and finally the proposed framework structures the components multiplicatively. It is useful to consider each of these features individually.

Having a unique correct response for each task makes it possible to define the optimal path to identify the correct answer. That property permits detailed assessment of biases in behavior (Creyer, Bettman, & Payne, 1990; Huber, Ariely, & Fischer, 2002). Knowing the optimal response is crucial to be able to derive metrics such as Wrong Target. In addition, an agent task enables an assessment of accuracy. In our case the simplicity of the agent task limits the relevance to the general effort-accuracy tradeoff. Our research is similar to earlier empirical research that instructed participants to use specific choice strategies (Jiang, Potters, & Funaki, 2016; Schoemann, Schulte-Mecklenbeck, Renkewitz, & Scherbaum, 2019). We believe that future studies could derive more insight from decomposing effort components for more complex and directly relevant tasks, such as selecting candidates for a job or finding items within a complex image. In those tasks, variation in error level could provide deeper insights into how attention can be modified to increase both speed and accuracy.

Second, eye-tracking is needed to assess processing for this task. For our study, objective measures of the four effort components would be very difficult to observe without detailed measures of attention to instructions and accessed cells. Participants may have had a sense of how efficiently and confidently they performed, but most have very little idea about how much time they took or how accurate they were (Fennema & Kleinmuntz, 1995). Eye-tracking enables an assessment of objective components of effort, and thus is most relevant to cases where it is possible to precisely determine attention (Lohse & Johnson, 1996). Fortunately, conducting large-scale eye-tracking studies has become much easier with cheaper and more reliable equipment.

In our study, defining multiplicative components that together equal Attention Effort had important advantages, as discussed earlier in Section 3.5. The multiplicative framework can also be applied to other decision tasks (e.g., elimination by aspects) and Web Appendix E provides a discussion of the requirements for its application in other contexts. Alternatively, one may formulate a non-multiplicative model, particularly when these requirements are not met. To test the usefulness of a non-multiplicative model we redefined the variables for our study. We measured Wrong Target as the fixation time per unneeded cells, and further broke the multiplicative model by defining repetition and duration on all, rather than on needed cells in the grid.

With those changes Wrong Target became a stronger driver of Attention Effort, but the general relationships among the components had very similar means and growth rates compared with our original model. Thus, Repetition still has the greatest variance across participants initially and the greatest drop over time. Duration continues to have little change with experience and has minimal relationships with the other components. Orientation again shows its value in reducing effort in the other components, while Wrong Target remains as the measure that best identifies poor search strategy and implementation. While it is possible to explore different metrics and analysis measures, we are generally pleased by the robustness of our initial findings.

It is also important to acknowledge that the relationships among the components reflect their distributions in a population, rather than their causal relationships. Given that people are poorly aware of the attentional processes, a difficult but important goal is to test the impact of altering the individual components of effort on both total effort and accuracy. Some important possibilities could be tested. For example, increasing practice sessions could be shown to reduce all the other effort components, while implementing time pressure could be found to decrease repetition but have minimal effect of duration.

Ultimately, it is valuable to understand the components that lead to faster and more accurate activities. This study has the advantage of revealing attention processes that are reasonable and, from our perspective, surprising. We are less sure about the replicability of particular results to totally different tasks that involve greater learning and task time. Still, the real benefit of the study is to provide a framework for attention-focused tasks, and to suggest roles for a number of new constructs that have promise to apply more generally.

References

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*(6), 396–408. https://doi.org/10.1111%2Fj.1467-9280.1991.tb00174.x

Bettman, J. R., Johnson, E. J., & Payne, J. W. (1990). A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes*, *45*(1), 111–139. https://doi.org/10.1016/0749-5978(90)90007-V

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Creyer, E. H., Bettman, J. R., & Payne, J. W. (1990). The impact of accuracy and effort feedback and goals on adaptive decision behavior. *Journal of Behavioral Decision Making*, *3*(1), 1–6. https://doi.org/10.1002/bdm.3960030102

Ederer, F., & Manso, G. (2013). Is pay for performance detrimental to innovation?. *Management Science*, *59*(7), 1496–1513. https://doi.org/10.1287/mnsc.1120.1683

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99*(1), 3–19. https://doi.org/10.1037//0033-2909.99.1.3

Fechner, H. B., Schooler, L. J., & Pachur, T. (2018). Cognitive costs of decision-making strategies: A resource demand decomposition analysis with a cognitive architecture. *Cognition, 170*, 102–122. https://doi.org/10.1016/j.cognition.2017.09.003

Fennema, M. G., & Kleinmuntz, D. N. (1995). Anticipations of effort and accuracy in multiattribute choice. *Organizational Behavior and Human Decision Processes, 63*(1), 21–32. https://doi.org/10.1006/obhd.1995.1058

Garcia-Retamero, R., & Dhami, D. K. (2009). Take-the-best in expert-novice decision strategies for residential burglary. *Psychonomic Bulletin & Review, 16*(1), 163–169. https://doi.org/10.3758/PBR.16.1.163

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*(4), 523–552. https://doi.org/10.1007/s10648-011-9174-7

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–482. https://doi.org /10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, *103*(4), 650–669. https://doi.org/10.1037/0033-295X.103.4.650

Gilchrist, I. D., & Harvey, M. (2000). Refixation frequency and memory mechanisms in visual search. *Current Biology*, *10*(19), 1209–1212. https://doi.org/10.1016/S0960-9822(00)00729-6

Graefe, A., & Armstrong, J. S. (2012). Predicting elections from the most important issue: A test of the take-the-best heuristic. *Journal of Behavioral Decision Making*, *25*(1), 41–48. https://doi.org/10.1002/bdm.710

Griffiths, W. (2003). Bayesian inference in the seemingly unrelated regressions model. Gilles DEA, ed. *Computer-Aided Econometrics* (Marcel Dekker, Inc., New York, Basel), 228–255.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *5*(1), 172–190. https://doi.org/10.1037/0278-7393.25.1.172

Hardy, J. H., Day, E. A., Hughes, M. G., Wang, X., & Schuelke, M. J. (2014). Exploratory behavior in active learning: A between- and within-person examination. *Organizational Behavior and Human Decision Processes*, *125*, 98–11. https://doi.org/10.1016/j.obhdp.2014.06.005

Huber, J., Ariely, D., & Fischer, G. (2002). Expressing preferences in a principal-agent task: A comparison of choice, rating and matching. *Organizational Behavior and Human Decision Processes*, *87*(1), 66–90. https://doi.org/10.1006/obhd.2001.2955

Jiang, T., Potters, J., & Funaki, Y. (2016) Eye-tracking social preferences. *Journal of Behavioral Decision Making*, *29*(2-3), 157–168. https://doi.org/10.1002/bdm.1899

Khader, P. H., Pachur, T., & Jost, K. (2013). Automatic activation of attribute knowledge in heuristic inference from memory. *Psychonomic Bulletin & Review*, *20*(2), 372–377. https://doi.org/10.3758/s13423-012-0334-7

Khader, P. H., Pachur, T., Weber, L. A., & Jost, K. (2016). Neural signatures of controlled and automatic retrieval processes in memory-based decision-making. *Journal of Cognitive Neuroscience*, *28*(1), 69–83. https://doi.org/10.1162/jocn_a_00882

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665–683. https://psycnet.apa.org/doi/10.1037/a0020198

Langley, P., & Simon, H. A. (1981). The central role of learning in cognition. Anderson JR, ed. *Cognitive Skills and Their Acquisition* (Lawrence Erlbaum Associates, Hillsdale, New Jersey), 361–380.

Liechty, J., Pieters, R., & Wedel, M. (2003). Global and local covert visual attention: Evidence from a Bayesian hidden Markov model. *Psychometrika*, *68*(4), 519–541. https://doi.org/10.1007/BF02295608

Lohse, L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Processing Behavior*, *68*(1), 28–43. https://doi.org/10.1006/obhd.1996.0087

Lauretro, D., Stefano, B., Canessa, N., & Zollo, M. (2015). Understanding the exploration-exploitation dilemma: An fMRI study of attention control and decision performance. *Strategic Management Journal*, *36*, 319–338.

Meißner, M., Musalem, A., & Huber, J. (2016). Eye tracking reveals processes that enable conjoint choices to become increasingly efficient with practice. *Journal of Marketing Research*, *53*(1), 1–17. https://doi.org/10.1509%2Fjmr.13.0467

Orquin, J. L., Chrobot, N., & Grunert, K. G. (2018). Guiding decision makers' eye movements with (un) predictable object locations. *Journal of Behavioral Decision Making*, *31*(3), 341–354. https://doi.org/10.1002/bdm.2060

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206. https://doi.org/10.1016/j.actpsy.2013.06.003

Pachur, T., Hertwig, R., Gigerenzer, G., & Brandstätter, E. (2013). Testing process predictions of models of risky choice: A quantitative model comparison approach. Frontiers in Psychology, 4, 646. https://doi.org/10.3389/fpsyg.2013.00646

Pachur, T., & Marinello, G. (2013). Expert intuitions: How to model the decision strategies of airport customs officers. *Acta Psychologica*, *144*(1), 97–103. https://doi.org/10.1016/j.actpsy.2013.05.003

Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O., & Hertwig, R. (2018). Prospect theory reflects selective allocation of attention. *Journal of Experimental Psychology: General*, *147*(2), 147–169. https://doi.org/10.1037/xge0000406

Peterson, C. R., Hammond, K. R., & Summers, D. A. (1965). Multiple probability-learning with shifting weights of cues. *The American Journal of Psychology*, *78*(4), 660–663. https://doi.org/10.2307/1420932

Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review*, *101*(2), 900–926. https://doi.org/10.1257/aer.101.2.900

Ritter, F. E., & Schooler, L. J. (2001). The learning curve. *International Encyclopedia of the Social and Behavioral Sciences*, *13*, 8602–8605.

Rosch, J. L., & Vogel-Walcutt, J. J. (2013). A review of eye-tracking applications as tools for training. *Cognition, Technology & Work*, *15*(3), 313–327. https://doi.org/10.1007/s10111-012-0234-7

Russo, J. E., & Leclerc, F. (1994). An eye-fixation analysis of choice processes for consumer nondurables. *Journal of Consumer Research*, *21*(2), 274–290. https://doi.org/10.1086/209397

Schoemann, M., Schulte-Mecklenbeck, M., Renkewitz, F., & Scherbaum, S. (2019). Forward inference in risky choice: Mapping gaze and decision processes. *Journal of Behavioral Decision Making*, *32*(5), 521–535. https://doi.org/10.1002/bdm.2129

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics Made Easy: An Effort-Reduction Framework. *Psychological Bulletin, 134*(2), 207–222. https://psycnet.apa.org/doi/10.1037/0033-2909.134.2.207

Tobii Software (2016). Tobii studio 3.4.5 user manual. (Retrieved February 2020) https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-studio-user-manual.pdf.

Todd, P., & Benbasat, I. (1992). The use of information in decision making: An experimental investigation of the impact of computer-based decision aids. *MIS Quarterly*, *16*(3), 373–393. https://doi.org/10.2307/249534

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, *59*(4), S251–S278.

Van der Lans, R., Pieters, R., & Wedel, M. (2008). Eye-movement analysis of search effectiveness. *Journal of the American Statistical Association*, *103*(482), 452–461. https://doi.org/10.1198/016214507000000437

Velichkovsky, B. M., Rothert, A., Kopf, M., Dornhöfer, S. M., & Joos, M. (2002). Towards an express diagnostics for level of processing and hazard perception. *Transportation*

Research Part F: Traffic Psychology and Behaviour, *5*(2), 145–156. https://doi.org/10.1016/S1369-8478(02)00013-X

Wedel, M., Yan, J., Siegel, E. L., & Li, H. (2016). Nodule detection with eye movements. *Journal of Behavioral Decision Making*, *29*(2-3), 254–270. https://doi.org/10.1002/bdm.1935

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, *57*(298), 348–368. https://doi.org/10.2307/2281644

Web Appendix A: Additional information about the eye-tracking experiment

This web appendix provides additional details about the experiment regarding the stimuli, procedure, apparatus and analysis of the eye-tracking data.

Stimuli: attribute levels, lexicographic importance and ordering

As shown in Table A1, we use text only to describe the features, attributes and alternatives in all tasks to limit salience differences from non-textual images that could affect our results (Milosavljevic, Navalpakkam, Koch, & Rangel, 2012).

Table A1. Attributes and levels used to describe the vacation package

Attributes	Levels		
Food quality	Poor (-)	Average (o)	Good (+)
Customer recommending	50% (-)	70% (o)	90% (+)
Distance to CBD	3km (-)	2km (o)	1km (+)
Sea view	No view (-)	Side view (o)	Full view (+)
Price per person	\$899 (-)	\$799 (o)	\$699 (+)
Room category	Standard (-)	Superior (o)	Deluxe (+)

The attributes were shown at the top of the screen in order of decreasing importance. That order of the attributes (lexicographic rule) varied across subjects, while the order of the attributes in the grid was unchanged across subjects and across tasks. We expected that the task would be easier for those whose attribute importance order corresponded with their order in the grid. Accordingly, six versions of the attribute importance order shown in Table A2 were tested to see if additional effort was generated by that lack of correspondence. The correlation varies between r=.82 where the most important attributes are near the top, and r=-.66 where they are near the bottom. We tested the impact of the different components across these six conditions and found very little difference in the allocation of effort components or their relationships with each other.

Table A2. Attribute importance order given to participants

	Order 1	Order 2	Order 3	Order 4	Order 5	Order 6
Attribute 1	1	4	3	5	2	6
Attribute 2	4	6	2	1	5	3
Attribute 3	2	5	4	6	3	1
Attribute 4	3	1	5	4	6	2
Attribute 5	5	3	6	2	1	4
Attribute 6	6	2	1	3	4	5
Correlation (order of the attributes in the decision matrix, attribute importance order given to participants in the question text)						
	0.83	**-0.66**	**0.09**	**-0.26**	**0.03**	**-0.03**

Finally, it is important to control for the order in which the attributes needed to identify the best alternative are presented to participants. To do that the order of the 12 tasks was randomized across subjects. Of course, for some participants the least complex tasks came first while for others the most complex tasks came first. Both effects could alter the average effort components and their learning across 12 tasks. We control for complexity, learning and mean effort in one equation to resolve this issue.

Procedure: dominant eye test

The experiment ended with a standard dominant eye test. The interviewer asked the participants to point to a far object with an outstretched arm using both eyes. While still pointing, the participants closed one eye at a time. The eye-tracking the finger pointing at the target became our measure of dominance, with 61% of the participants assessed as right-eye dominant. In the analyses we use the average of both eyes for all participants. However, our results remain almost identical if we use the dominant eye instead.

Apparatus: setting and software

The eye-tracking experiment took place in a windowless room illuminated with artificial light. To avoid distraction only the lab assistant and the participant were present in the course of an experiment. A heavy chair discouraged large changes in seating position or distance from the eye-tracking monitor. The assistant made sure that the distance from the participants' eyes to the monitor was between 50 and 80 cm (ideally 60 cm). The seating height was adjusted so that the participant's eyes were on the same level as the center of the monitor. The software used the standard 9-point calibration routine (with a gray background color) included in the Tobii software (Tobii Software, 2016). The presence of the assistant guaranteed that the participants completed the experiment in one sitting without pauses between tasks. We programmed our own software to present the stimuli in an online survey. By running the survey several times on the computer used before the experiment started, we made sure that all stimuli appeared immediately after clicking "next". There was no fixation cross before seeing the stimuli, but our results were validated by finding almost identical results from deleting the first fixation. Thus, there were no breaks between tasks, except for an encouraging reminder at the end of task 6. Participants could take as much time as they wanted to make a decision in every task. Participants indicated they had normal or corrected to normal vision and were allowed to wear glasses or contact lenses. They had to remove mascara if they had used it. None of the participants had droopy eyelids so that we had no problems with the system not finding the participants' pupils.

Analysis of the eye-tracking data: areas of interest, fixations and eye-tracking quality

The areas of interest (AOI) were defined as non-overlapping cells. All AOIs in the 6x4 information grid have the same size (168 x 73 pixels). Moreover, in order to calculate the extent of orientation we defined three additional types of AOIs: question text (900 x 300 pixels), attribute descriptions (199 x 73 pixels) and alternative labels (168 x 24 pixels). The AOIs including the features and attribute descriptions were substantially bigger than the descriptions, as can be seen from Figure 1. Different definitions of the AOIs have been tested in line with Orquin, Ashby, and Clarke (2016) to assure that the definition of the AOIs do not substantially influence the results. We emphasize that we assume that the center of the fixation, i.e. the x-y-coordinate, indicates the identity of the fixated AOI. In line with several other researchers (e.g., Yang, Toubia, & de Jong, 2015; Shi, Wedel, & Pieters, 2013) we expect that neighboring AOIs have been processed and we cannot exclude information from parafoveal viewing.

Fixations were defined as continuous gazes within each area of interest. The Tobii studio software preprocessed the eye-tracking data and the standard Tobii fixation filter (I-VT filter) identified fixations and saccades. We used the default parameters and, thus, fixations that were shorter than 60ms were discarded.

The assistant who ran the experiment in the lab watched the generated eye-tracking videos to validate their eye-tracking quality. The exclusion of participants is thus based on the eye-tracking experience of the lab manager, but is not based on an objective quality threshold. Out of 209 participants who took part in the experiment, the data of 13 participants were incomplete or showed an unusual horizontal or vertical drift. For the remaining 196 participants, the eye-tracking quality was rated as being sufficiently high in all tasks. Two other participants were later excluded because of incomplete data. Thus, we exclude 7.2% of participants from the analysis. Given that participants had to have high eye-tracking quality in all tasks, we consider that percentage of lost data to be fair.

Web Appendix B: Hierarchical SUR Model and Bayesian Estimation

We define y_{ift} as the time allocated to attention component factor f by participant i at task t, where $f \in \{\mathrm{o}, \mathrm{w}, \mathrm{d}, \mathrm{r}\}$. We control for the complexity of the task and the number of completed tasks. The model is formulated as follows:

$$\ln(y_{ift}) = \beta_{if} + \gamma_{if} \ln(t) + \delta_f \ln(C_{it}) + \varepsilon_{ift}, \qquad \text{(B1)}$$

where:

$$\varepsilon_i \equiv (\varepsilon_{iot}, \varepsilon_{iwt}, \varepsilon_{idt}, \varepsilon_{irt})' \text{ and } \varepsilon_i \sim N(0, \Omega);$$

$\mathrm{Ln}(t)$ and $\ln(C_{it})$ are the natural logarithms of the task number and complexity, respectively. To facilitate the interpretation of the model parameters, these two variables are mean centered. Hence, the intercept β_{if} represents the expected (log) allocation to attention component f by participant i given (geometric) mean levels of experience and complexity.

We use a random coefficients specification to model heterogeneity across participants in terms of the baseline (β_{if}) and learning parameters (γ_{if}). Accordingly, denote by $\beta_i \equiv (\beta_{io}, \beta_{iw}, \beta_{id}, \beta_{ir})$ and $\gamma_i \equiv (\gamma_{io}, \gamma_{iw}, \gamma_{id}, \gamma_{ir})$. We then let $\beta_i \sim MVN(\theta_\beta, V_\beta)$ and $\gamma_i \sim MVN(\theta_\gamma, V_\gamma)$; where V_β is a full variance-covariance matrix and V_γ is a diagonal variance matrix.[1]

We use a Bayesian estimation approach. Bayesian methods facilitate the estimation of models with random effects circumventing the need for high dimensional integration. They also provide a numerically convenient way to obtain confidence intervals (i.e., posterior probability intervals) for the parameter estimates without assuming our data set is infinitely large.

[1] Using a full variance-covariance specification for $\boldsymbol{V_\gamma}$ leads to similar results.

We specify the following hyperprior distributions: $\theta_\beta \sim N(0, 2^2 I_4)$, $\theta_\gamma \sim N(0, 2^2 I_4)$, $\delta_f \sim N(0, 2^2)$, where I_4 is an identity matrix with 4 rows and columns. Denote each of the diagonal elements of V_γ by $\rho^2_{\gamma,f}$. Following the recommendations in the Stan manual (Stan Development Team 2017, p.39, p.145), we specify $V_\beta = (\rho_\beta I_4) L_\beta L_\beta{}' (\rho_\beta I_4)'$. Note that L_β is the lower triangular decomposition of the correlation matrix based on V_β, while ρ_β is a vector of standard deviations for each component of β and thus equal to the square root of the diagonal elements of V_β. A prior on V_β can be specified by defining priors on L_β and ρ_β, as follows: $L_\beta \sim LKJ(1)$ and $\rho_{\beta,f} \sim Cauchy(0, 2.5)$. The same approach is used to specify a prior distribution on Ω as follows: $\Omega = (\rho_\varepsilon I_4) L_\varepsilon L_\varepsilon{}' (\rho_\varepsilon I_4)'$, where L_ε is the lower triangular decomposition of the correlation matrix based on Ω, while ρ_ε is a vector of standard deviations for each component of ε. We then specify the following priors: $L_\varepsilon \sim LKJ(1)$ and $\rho_{\varepsilon,f} \sim Cauchy(0, 2.5)$.

We then use a Hamiltonian Monte Carlo (HMC) approach to estimate the model parameters, where the estimation is conducted in Stan based on its implementation in R (Carpenter et al., 2017). The output of this estimation is a set of draws of the model parameters from their posterior distributions.

Direct and net effects

The calculation of direct and net effects relies on the assumption that the distribution of the parameters follows a multivariate normal distribution across respondents. A fundamental property of the Multivariate Normal distribution (Tong, 2012) establishes that:

If $\binom{x}{y} \sim MVN\left(\binom{\mu_x}{\mu_y}, \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}\right)$ then

$$E(x|y = y_0) = \mu_x + (y_0 - \mu_y)\frac{\sigma_{xy}}{\sigma_y^2} \qquad \text{(B2)}$$

Now, let $y_0 = \mu_y + \sigma_y$, then

$$\begin{aligned} E\big(x\big|y = \mu_y + \sigma_y\big) &= \mu_x + (\mu_y + \sigma_y - \mu_y)\frac{\sigma_{xy}}{\sigma_y^2} \\ &= \mu_x + \sigma_y\frac{\sigma_{xy}}{\sigma_y^2} \\ &= \mu_x + \frac{\sigma_{xy}}{\sigma_y} \qquad \text{(B3)} \\ &= \mu_x + \frac{\sigma_{xy}}{\sigma_y}.\frac{\sigma_x}{\sigma_x} \\ &= \mu_x + Corr(x,y)\sigma_x \end{aligned}$$

Where the correlation between x and y is given by $\frac{\sigma_{xy}}{\sigma_x\sigma_y}$.

Hence:

$$E\big(x\big|y = \mu_y + \sigma_y\big) - E\big(x\big|y = \mu_y\big) = E\big(x\big|y = \mu_y + \sigma_y\big) - \mu_x = Corr(x,y)\sigma_x \qquad \text{(B4)}$$

Therefore, when component f changes from μ_f to $\mu_f + \sigma_f$, the indirect effect of component f on component f' is given by

$$E\big(f'\big|f = \mu_f + \sigma_f\big) - \mu_{f'} = Corr(f,f')\sigma_{f'}, \qquad \text{(B5)}$$

Given this result, and using each of the HMC draws (after warmup), the calculation of direct and net effects is done as follows:

$$Direct_f = \frac{1}{R}\sum_{r=1}^{R}\sqrt{V_{ff}^{(r)}} \qquad \text{(B6)}$$

$$Net_f = Direct_f + \frac{1}{R}\sum_{r=1}^{R}\sum_{f'\neq f} corr\big(\beta_{if},\beta_{if'}\big)^{(r)}\sqrt{V_{ff}^{(r)}} \qquad \text{(B7)}$$

where *r* denotes one of the *R* draws from the Bayesian estimation, $V_{ff}^{(r)}$ corresponds to the variance of component *f* baselines and $corr\left(\beta_{if}, \beta_{if'}\right)^{(r)}$ corresponds to the correlation between the baselines of components *f* and *f'*, as estimated in draw *r* of the HMC procedure.

Web Appendix C: Estimation Results

Table A3. Parameter estimates for the random and common coefficients of the SUR model: posterior means, posterior standard deviations, 95% posterior probability intervals.*

			mean	sd	2.5%	97.5%
Random	Orientation	Intercept	0.72	0.01	0.70	0.75
	Orientation	Learning	-0.07	0.01	-0.09	-0.06
	Wrong	Intercept	0.29	0.01	0.27	0.31
	Wrong	Learning	-0.02	0.01	-0.03	0.00
	Duration	Intercept	5.75	0.01	5.73	5.78
	Duration	Learning	-0.02	0.01	-0.04	-0.01
	Repetition	Intercept	0.70	0.02	0.67	0.74
	Repetition	Learning	-0.13	0.01	-0.15	-0.11
Common	Orientation	Complexity	-0.06	0.01	-0.08	-0.04
	Wrong	Complexity	-0.27	0.01	-0.29	-0.25
	Duration	Complexity	0.01	0.01	-0.01	0.02
	Repetition	Complexity	-0.12	0.01	-0.14	-0.09

* Posterior standard deviations (sd) and 95% intervals (2.5% and 97.5%) are measures of parameter estimate uncertainty.

Table A4. Standard deviation of intercepts and learning coefficients across participants: posterior means, posterior standard deviations, 95% posterior probability intervals.*

		StdDev	sd	2.5%	97.5%
Intercepts	Orientation	0.18	0.01	0.16	0.20
	Wrong target	0.13	0.01	0.11	0.16
	Duration	0.17	0.01	0.15	0.19
	Repetition	0.22	0.02	0.19	0.25
Learning	Orientation	0.05	0.01	0.04	0.07
	Wrong target	0.03	0.01	0.00	0.05
	Duration	0.04	0.01	0.01	0.05
	Repetition	0.05	0.02	0.01	0.07

* Posterior standard deviations (sd) and 95% intervals (2.5% and 97.5%) are measures of parameter estimate uncertainty.

Web Appendix D: Model Comparison

Our analysis relies on a joint model of the four components and it allows respondents to be heterogeneous in their mean effort allocated to each component and in their learning curves. One could alternatively model each component independently or assume that all respondents are homogeneous. However, each of these model simplifications severely damages the ability of our formulation to explain the attentional data. Specifically, we estimated two variations of our main model: i) Independent-Heterogeneous: each component is modeled independently (i.e., eliminating the SUR error structure), but respondents are heterogeneous; and ii) Joint-Homogeneous: all respondents are assumed to be homogeneous while the components are allowed to be correlated with each other. We then computed the log marginal likelihood of each model (see Table A5), which was then used to calculate Bayes factors. Each Bayes factor compares the fit of our model (i.e., Joint-Heterogeneous) to each of these two alternative formulations. The Bayes factor, a standard tool for model selection in Bayesian Analysis (Kass & Raftery, 1995), compares the evidence in favor of a model against an alternative formulation, given the observed data.

In terms of the first alternative model, the logarithm of the Bayes factor comparing our model against the Independent-Heterogeneous is equal to 1418-724=694, providing very strong evidence in favor of the joint modeling of the four components. Similarly, comparing our model to one assuming all respondents are homogeneous, we obtain an even greater logarithm of the Bayes Factor (2,361), once again providing very strong support for our heterogeneous formulation. Consequently, both tests justify the specification of a joint model of the four components and the use of a heterogeneous formulation. In addition, one can also compare the models in terms of prediction. We computed the mean absolute deviation (MAD) for the main model and its two variations discussed above (also displayed in Table

A5). Overall, we find that the main model provides a substantial improvement in terms of in-sample fit.

Table A5. Model Comparison in terms of MAD and Log Marginal Likelihood.

			MAD				
Model	SUR	Hetero.	ln O	ln W	ln D	ln R	LogML
Main	y	y	0.12	0.14	0.12	0.18	1418.0
Sur model	n	y	0.15	0.16	0.13	0.21	724.0
No participant heterogeneity	y	n	0.17	0.17	0.18	0.25	-943.1

Web Appendix E: Requirements for a similar multiplicative decomposition for other choice tasks.

Multiplicative decompositions similar to this study have to meet certain requirements. First, the repeated task needs to be defined using rules accessed in a separate area. Second, the information on the alternatives needs to be placed in separate cells so that it is possible to track the access to each piece of information. Third, each cell needs to be categorized as either relevant or irrelevant for determining the best alternative. This classification enables an assessment of efficient processing as a function of the proportion of needed cells accessed in a task and allows both Repetition and Duration to be defined with respect to needed cells.

There are many ways these requirements can be satisfied, as illustrated by the following examples. The rules can specify the order of the attributes as in a lexicographic rule, or in the case of a conjunctive rule it could specify the acceptable values for each attribute. A fixed elimination-by-aspects could be tested by specifying the features that must be acceptable across hierarchically ordered features. It is also possible to have the grid remain constant across tasks, but change the rules. Then orientation depends on how often the new rules or altered attribute labels need to be accessed.

For compensatory rules, the application of our multiplicative model requires some modifications. On the one hand, since all cells are relevant for the decision maker, the wrong target metric becomes constant and hence the estimation of a joint model should exclude that component. On the other hand, one could consider a more general classification of cells into needed and not needed. In particular, it may be possible to assess the value of each cell, given the information learned so far. That measure could be reasonably estimated as a function of the importance of each attribute combined with the current expected value of each alternative, thus requiring a non-multiplicative structure. This assessment could be used to classify cells into those that are more and less crucial for identifying the best option.

Finally, if one were to consider instead a non-multiplicative model of attention, direct and net effects of components must be measured in different ways. The direct effect can be estimated through a multiple regression with total attention as the dependent variable and using all components as explanatory variables in order to partial out the effects of component interactions. The net effect is then reflected in a simple regression of each component on the dependent variable.

Web Appendix F: Confidence intervals for Figures 3-7.

The following table provides the values of the observations in Figures 3-7 with their respective asymptotic 95% confidence intervals.

Table A6. Means and confidence intervals for the values in Figures 3-7.

	O adj			W adj			D adj			R adj			Total adj		
							Condition 1								
task	mean	2.5%	97.5%	mean	2.5%	97.5%	mean	2.5%	97.5%	mean	2.5%	97.5%	mean	2.5%	97.5%
1	2.38	2.24	2.52	1.43	1.35	1.51	0.33	0.31	0.35	3.01	2.77	3.28	29.2	26.5	32.3
2	2.22	2.11	2.35	1.33	1.26	1.40	0.34	0.32	0.36	2.84	2.63	3.08	24.8	22.9	26.8
3	2.20	2.06	2.34	1.38	1.29	1.48	0.32	0.30	0.34	2.72	2.52	2.92	22.8	20.9	25.0
4	2.09	1.97	2.21	1.42	1.32	1.53	0.33	0.31	0.35	2.49	2.33	2.67	21.2	19.5	23.1
5	2.13	2.00	2.26	1.32	1.23	1.41	0.32	0.31	0.34	2.43	2.25	2.63	19.2	17.5	21.0
6	2.00	1.88	2.12	1.37	1.28	1.47	0.32	0.30	0.34	2.48	2.31	2.65	18.7	16.9	20.6
7	2.31	2.18	2.45	1.29	1.23	1.36	0.31	0.29	0.33	2.31	2.15	2.49	18.5	17.0	20.2
8	2.01	1.89	2.13	1.40	1.31	1.50	0.32	0.31	0.34	2.19	2.06	2.33	17.4	16.2	18.7
9	1.93	1.81	2.05	1.31	1.24	1.39	0.32	0.30	0.34	2.35	2.21	2.49	16.5	15.2	17.8
10	1.93	1.82	2.04	1.30	1.24	1.37	0.32	0.31	0.34	2.33	2.15	2.52	16.5	15.2	18.0
11	1.92	1.80	2.04	1.41	1.31	1.52	0.33	0.31	0.35	2.02	1.88	2.17	15.7	14.3	17.2
12	1.87	1.76	1.99	1.34	1.26	1.43	0.31	0.30	0.33	2.18	2.03	2.34	14.9	13.8	16.2
							Condition 2								
task	mean	2.5%	97.5%	mean	2.5%	97.5%	mean	2.5%	97.5%	mean	2.5%	97.5%	mean	2.5%	97.5%
1	2.23	2.08	2.39	1.31	1.24	1.39	0.33	0.31	0.35	2.43	2.26	2.62	20.5	18.9	22.1
2	2.19	2.03	2.36	1.40	1.29	1.51	0.32	0.30	0.34	2.23	2.05	2.42	18.8	17.2	20.5
3	2.07	1.94	2.21	1.33	1.23	1.43	0.33	0.31	0.35	2.19	2.03	2.37	17.0	15.6	18.6
4	2.11	1.98	2.25	1.34	1.27	1.42	0.32	0.30	0.34	2.10	1.93	2.28	16.6	15.2	18.1
5	2.05	1.91	2.20	1.39	1.27	1.52	0.31	0.29	0.33	2.13	1.98	2.28	16.2	15.0	17.5
6	1.99	1.85	2.14	1.34	1.26	1.43	0.31	0.29	0.33	2.05	1.90	2.22	14.9	13.7	16.2
7	2.26	2.07	2.48	1.26	1.18	1.36	0.31	0.29	0.33	2.00	1.83	2.19	15.3	13.9	16.9
8	1.97	1.85	2.09	1.36	1.27	1.46	0.32	0.30	0.33	2.00	1.85	2.16	14.7	13.4	16.0
9	1.97	1.85	2.10	1.36	1.27	1.44	0.32	0.30	0.34	2.01	1.88	2.15	15.0	13.8	16.3
10	1.93	1.79	2.09	1.32	1.20	1.44	0.32	0.30	0.34	1.86	1.71	2.03	13.3	12.3	14.3
11	1.89	1.76	2.04	1.39	1.29	1.50	0.30	0.28	0.32	1.96	1.81	2.12	13.4	12.4	14.6
12	1.92	1.77	2.08	1.40	1.30	1.50	0.30	0.29	0.32	1.89	1.76	2.04	13.4	12.3	14.6
							Condition 3								
task	mean	2.5%	97.5%	mean	2.5%	97.5%	mean	2.5%	97.5%	mean	2.5%	97.5%	mean	2.5%	97.5%
1	2.33	2.16	2.53	1.56	1.38	1.77	0.32	0.30	0.34	2.33	2.13	2.55	23.3	21.5	25.3
2	2.16	2.03	2.29	1.28	1.20	1.36	0.32	0.30	0.35	2.41	2.21	2.63	18.7	17.0	20.5
3	2.13	2.01	2.25	1.28	1.21	1.34	0.32	0.30	0.34	2.34	2.14	2.56	17.7	16.2	19.3
4	2.08	1.95	2.22	1.27	1.21	1.34	0.32	0.30	0.34	2.31	2.14	2.49	17.1	15.9	18.3
5	2.01	1.89	2.15	1.27	1.22	1.33	0.30	0.28	0.32	2.25	2.09	2.43	15.2	13.9	16.5
6	2.03	1.91	2.16	1.27	1.21	1.34	0.31	0.29	0.32	2.15	2.01	2.30	14.8	13.7	15.9
7	2.16	2.02	2.31	1.31	1.22	1.41	0.29	0.28	0.31	1.77	1.67	1.89	12.8	12.0	13.7
8	2.07	1.94	2.21	1.26	1.20	1.32	0.29	0.28	0.31	1.93	1.81	2.06	12.8	12.1	13.7
9	1.93	1.80	2.06	1.31	1.24	1.38	0.29	0.28	0.31	2.03	1.90	2.18	13.1	12.1	14.1
10	1.92	1.80	2.04	1.35	1.24	1.46	0.30	0.29	0.32	1.86	1.74	2.00	12.7	11.9	13.4
11	1.96	1.82	2.11	1.25	1.18	1.32	0.30	0.29	0.32	1.93	1.81	2.05	12.5	11.7	13.3
12	1.88	1.76	2.01	1.27	1.20	1.36	0.30	0.28	0.32	1.95	1.81	2.11	12.2	11.2	13.2

References in the web appendices:

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Milosavljevic, M., Navalpakkam, V., Koch, C., & Rangel, A. (2012). Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology, 22*(1), 67–74. https://doi.org/10.1016/j.jcps.2011.10.002

Orquin, J. L., Ashby, N. J., & Clarke, A. D. (2016). Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making, 29*(2–3), 103–115. https://doi.org/10.1002/bdm.1867

Shi, S. W., Wedel, M., & Pieters, F. G. M. (2013). Information acquisition during online decision making: A model-based exploration using eye-tracking data. *Management Science, 59*(5), 1009–1026. https://doi.org/10.1287/mnsc.1120.1625

Stan Development Team (2017). Stan modeling language user's guide and reference manual, version 2.18. (Retrieved February 2020 from) https://mc-stan.org/docs/2_18/reference-manual/index.html.

Tong, Y. L. (2012). *The multivariate normal distribution* (Springer et al., New York et al.).

Yang, L., Toubia, O., De Jong, M. G. (2015). A bounded rationality model of information search and choice in preference measurement. *Journal of Marketing Research, 52*(2), 166–183. https://doi.org/10.1509%2Fjmr.13.0288