

Identifying decision strategies in a consumer choice situation

Nils Reisen^{*1,2}, Ulrich Hoffrage¹, and Fred W. Mast²

¹ Faculty of Business and Economics, University of Lausanne

² Institute of Psychology, University of Lausanne

Abstract

In two studies on mobile phone purchase decisions, we investigated consumers' decision strategies with a newly developed process tracing tool called *InterActive Process Tracing* (IAPT). This tool is a combination of several process tracing techniques (Active Information Search, Mouselab, and retrospective verbal protocol). After repeatedly choosing one of four mobile phones, participants formalized their strategy so that it could be used to make choices for them. The choices made by the identified strategies correctly predicted the observed choices in 73% (Experiment 1) and 67% (Experiment 2) of the cases. Moreover, in Experiment 2 we directly compared Mouselab and eye tracking with respect to their impact on information search and strategy description. We found only minor differences between these two methods. We conclude that IAPT is a useful research tool to identify choice strategies, and that using eye tracking technology did not increase its validity beyond that gained with Mouselab.

Keywords: decision strategies, process tracing, verbal protocols, decision making, eye tracking, Mouselab.

1 Introduction

Identifying the processes that underlie judgment and decision making has been of great interest to researchers for several decades already. In this context, two major paradigms have been used: structural modeling and process tracing (Abelson & Levi, 1985; Billings & Marcus, 1983; Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Ford, Schmitt, Schlechtman, Hults, & Doherty, 1989; Harte & Koele, 1995; Payne, 1976; Svenson, 1979). Structural modeling aims to uncover psychological processes by relating the provided information to the decisions or judgments, typically via multiple linear regression analysis. The parameters in these models are thought to represent important features of participants' decision strategies, for instance, if a particular attribute receives a high weight in a regression equation it is interpreted as being very important for the decision maker. Despite its popularity, this approach has been criticized for ignoring the pre-decisional phase, that is, the processes that take place be-

tween stimulus presentation and final decision. For example, Svenson (1979) concluded that it is "gradually becoming clear that human decision making cannot be understood simply by studying final decisions" (p. 86) and, similarly, Payne, Braunstein, and Carroll (1978) argued that the "input-output analyses that have been used in most decision research are not fully adequate to develop and test process models of decision behavior" (p. 19). As a response to these and other objections against structural modeling (for an overview, see Bröder, 2000), Payne (1976) and others developed the process tracing approach by adapting methods from research on human problem solving (Newell & Simon, 1972). As opposed to structural modeling, the aim of process tracing is to directly describe the processes taking place during the pre-decisional phase. To achieve this, the participants' information search and integration is closely observed while they work on the decision task. Frequently used methods within this paradigm are information boards (e.g., Payne, 1976; Payne, Bettman, & Johnson, 1993), verbal protocols (e.g., Ericsson & Simon, 1984), the recording of eye movements (e.g., Lohse & Johnson, 1996; Russo & Leclerc, 1994; Russo & Rosen, 1975), and the method of Active Information Search (AIS; Huber, Wider, & Huber, 1997).

In the following, we briefly describe these process tracing methodologies and discuss their strengths and weaknesses. We then present a new tool called *InterActive Process Tracing* (IAPT), which we developed to identify the decision processes underlying preferential choice.

*We would like to thank the following people for their help with the planning and execution of the experiments: Gregory Affolter, Richard Ciapala, Julien Finci, Gabriella Sinicco, Huseyin Cumhur Tekin, Eren Vardarli, and Vasko Vitanov. A special thanks goes to Giovanni Rivera Diaz, Ada Lezama Lugo and Lucas Sinclair who programmed the software for the experiments. We also thank Dario Bombari for his technical assistance with the eye tracker, Felix Reisen and Chris M. White for their help with the simulations, and Chris M. White, Jan K. Woike, John Antonakis, Jonathan Baron and two anonymous reviewers for their helpful comments on previous versions of this manuscript. Finally, we are grateful for financial support provided by the Schweizer Nationalfonds (Grant numbers 105511–111621/1 and 100011–116111/1).

Table 1: Strengths and weaknesses of four process tracing techniques.

Strengths	Weaknesses
<i>Mouselab</i>	
<ul style="list-style-type: none"> + Convenient to use. + A large amount of data: which and how much information is retrieved and the sequence of the information acquisition. 	<ul style="list-style-type: none"> – Overly structured: participant may be influenced as to what information to use or to consider important. – Only data concerning the search for information, but no data concerning information integration.
<i>Eye Tracking</i>	
<ul style="list-style-type: none"> + A large amount of data: which and how much information is retrieved and the sequence of the information acquisition. + Very fast and effortless information acquisition. + Mostly nonreactive: behavior cannot easily be censored by the participants. + Better suited than Mouselab to problems with more complex information displays. 	<ul style="list-style-type: none"> – Expensive equipment. – A reliable calibration cannot be achieved for all participants. – Overly structured: participant may be influenced as to what information to use or to consider important. – Only data concerning the search for information, but no data concerning information integration.
<i>Active Information Search (AIS)</i>	
<ul style="list-style-type: none"> + Enhanced realism: participants are less affected by the experimental setup. 	<ul style="list-style-type: none"> – Less exact monitoring of the information acquisition process than with Mouselab. – Only data concerning the search for information, but no data concerning information integration.
<i>Retrospective Verbal Protocol</i>	
<ul style="list-style-type: none"> + Rich and detailed information: information search and integration. + No interference with decision making when participants work on the task. 	<ul style="list-style-type: none"> – Doubts that people can introspectively access their cognitive processes. – Reactivity: forgetting and fabrication. – Extremely time-consuming analysis.

IAPT uses various elements of the process tracing measures mentioned above to combine their strengths and simultaneously overcome some of their weaknesses. We subsequently describe two experiments in which we successfully applied IAPT to identify participants' decision strategies. Finally, we conclude with a discussion of our findings and outline avenues for future research.

1.1 Process tracing techniques

1.1.1 Information search: Mouselab, eye tracking, and the method of Active Information Search

A range of techniques has been developed within the process tracing paradigm, each of them having both strengths and weaknesses (see Table 1). A popular method is Mouselab (Payne et al., 1993), the computerized version of the information board (Payne, 1976). In a typical Mouselab-based study, participants have the opportunity to acquire information about the choice alterna-

tives by using the computer mouse to click on, or move a pointer over, the cells of an attributes-by-alternatives matrix. Mouselab provides data concerning the information acquisition phase, such as which cells are looked up, in which order, and how much time was spent looking at each cell. Besides being relatively easy to use for experimenters, this method is also quite convenient for participants because they are confronted with a relatively well-structured decision situation in which all the available information is clearly arranged.

Another, and in this context very similar, way to trace the participants' information search is to record their eye movements. Instead of using a computer mouse to obtain information, here participants simply have to look at a screen where the information is displayed. The eye tracking equipment records which information is fixated and so produces data that are similar to Mouselab's. However, for eye tracking, the process of information acquisition resembles more a natural situation (simple reading)

as compared to Mouselab (opening cells).

Similar to Mouselab and eye tracking, the method of Active Information Search (Huber, Wider, & Huber, 1977) is aimed at discovering the information that is actually requested by the decision maker. In contrast to studies using Mouselab, however, the decision task in a typical AIS study is presented with as little structure as possible. In this manner, participants can build up a cognitive representation of the task that is virtually unaffected by the experimental setup (Brucks, 1988; Huber et al., 1977). Specifically, the participants receive a minimal description of the decision situation and have to query the experimenter for any further information.

A major weakness of the information search techniques is that they provide no direct data about how participants integrate the obtained information (for other reactive effects of information boards, see Arch, Bettman, & Pakkar, 1978). Although it is commonly assumed that characteristics of the evaluation process can be deduced from the way in which participants search for information (e.g., Harte & Koele, 2001), it is not entirely clear exactly how information search and information integration are related to each other (for a critical position, see Bröder, 2000; Rieskamp & Hoffrage, 2008).

1.1.2 Information integration: Retrospective verbal protocol

One way to gain more explicit insight into the processing of the obtained information is to collect verbal protocols, which can be done in two different ways. Concurrent verbal protocols are collected while the participant works on the task, whereas retrospective verbal protocols are collected only after task completion. In both variants, the participants are asked to “think aloud,” that is, to tell the experimenter everything that comes or came to their minds when working on the task. Typically, these verbalizations are recorded and subsequently coded by the experimenter.

Although intuitively appealing, serious concerns have been raised regarding the use of verbal protocols in general and retrospective protocols in particular. In a classic paper, Nisbett and Wilson (1977) questioned the assumption that people have introspective access to their cognitive processes and concluded that people’s ability to observe and report upon higher order mental operations is often small or even not existent. Ericsson and Simon (1984) challenged this conclusion and claimed that “better methods for probing for that awareness (concurrent or immediate retrospective reports) would yield considerable insight into the cognitive processes occurring in *most* of the studies discussed by Nisbett and Wilson” (p. 29, italics in the original). However, they point out that retrospective verbal protocols should be collected

immediately after task completion and that the general instruction should be “to report everything you can remember about your thoughts during the last problem” (p. 19). When these conditions are met then retrospective verbal reports can be powerful means for studying cognitive processes. In contrast, Russo, Johnson, and Stephens (1989) have a more negative view on verbal protocols. They argue that in concurrent protocols the instruction to think aloud may interfere with the task the participant is working on, which can alter the accuracy of the response. Even worse, these authors found significant reactivity when collecting verbal protocols retrospectively. This reactivity was manifested in errors of omission (forgetting), that is, the participants could not recall the processes they used, and errors of commission (fabrication), that is, they reported processes that did not actually happen. Russo et al. (1989) conclude that retrospective protocols should be dismissed as nonveridical.

In our view, the position taken by Russo et al. (1989) is overly pessimistic, especially given that the problems associated with retrospective protocols are not without remedies. First, the problem of forgetting can be effectively diminished when cues are provided that facilitate the participants’ recall during the collection of the retrospective protocol.¹ Such a procedure has been shown to increase the completeness of the verbal protocol (see Gog, Paas, Merriënboer, & Witte, 2005, for an overview). Second, to verify whether fabrication really occurred and whether the verbal protocols do or do not accurately describe participants’ decision processes, one can compare the protocols to some behavioral data. If, for example, the protocol data are used to formulate an algorithm that can replicate the decisions made by the participants then this provides considerable evidence for the validity of such protocols.

1.1.3 InterActive Process Tracing (IAPT)

Given that each of the four process tracing techniques described above has weaknesses and limitations, we developed a new method that uses and combines features of these methods, thereby overcoming some of their downsides. As pointed out by various authors, multimethod approaches are a particularly useful way to trace decision behavior (e.g., Einhorn et al., 1979; Harte & Koele, 2001; Payne, 1976; Payne et al., 1978; Riedl et al., 2008; Russo, 1978).

A major feature of our method is that an attempt is made to detect the cognitive processes interactively with the participant, which is why we call it *InterActive* Process Tracing. In the experiments, participants first se-

¹Interestingly, Russo et al. (1989) were one of the first to use this method but they did not observe the positive effects found in other studies.

lected the attributes they considered important (AIS), then they made a series of choices (Mouselab in Experiments 1 and 2, eye tracking in Experiment 2), and finally, they were interviewed about their choice strategies. Note, however, that the last phase of our method deviates from the conditions specified by Ericsson and Simon (1984) in that participants were not asked to report a stream of thought but rather to construct, in retrospect, a precise process model that resembles their own decision strategy as closely as possible. We are aware that these changes in the procedure might reduce the validity of the verbal protocols. However, the described strategies can be used to retrospectively predict² the choices actually made by the participants. The degree of correspondence between the actual choices and the predictions of the described strategies can then be used as a measure of the validity of the described strategies.

1.1.4 Approaches similar to IAPT

Similar procedures have been used by other authors in various contexts (e.g., Bettman, 1970; Einhorn et al., 1979; Larcker & Lessig, 1983; Li, Shue, & Shiue, 2000). Bettman (1970), for example, obtained concurrent verbal protocols from five housewives who were encouraged to think aloud while shopping. Based on these protocols, he then developed a computational model and subsequently tested whether this model could replicate the decisions made by the participants reasonably well. He found that the predictions were highly accurate. In another study, Larcker and Lessig (1983) asked participants to evaluate the stocks of 50 actual companies with respect to possible purchase. Immediately after the evaluation, participants provided a verbal report of their procedure and developed diagrammatic representations of the manner in which they made their judgment (with the assistance of the researcher). In addition, a linear model was estimated. The retrospective process tracing models predicted the participants' actual choices correctly in 84.4% of the cases (chance was 50%), which was even higher than the percentage of correct predictions made by the linear model (73%). Finally, Einhorn et al. (1979) and Li et al. (2000) used concurrent verbal protocols to construct a model that was subsequently validated by comparing its predictions to the decisions made by the participants. Again, the models predicted the decisions quite well.

In the two experiments described below, we used our new method of IAPT to address the question of whether people are indeed able to gain introspective access to their cognitive processes, and ultimately, to what extent those

²Note that because the strategies were calculated only after the participants' choices, the correct term in this context would be postdiction rather than prediction. However, in the following we still use prediction because it is the more standard terminology.

verbal protocol data are instrumental in constructing process models that can accurately predict their choices. In addition, we were interested in the convergent validity of the information search techniques and the verbal protocol.

2 Experiment 1

2.1 Method

2.1.1 Participants

Participants were 37 students (8 female and 29 male) of the Swiss Federal Institute of Technology of Lausanne (EPFL) with a mean age of 23.8 years ($SD = 2.6$ years).

2.1.2 Task

In each of 30 choice trials, participants selected one of four mobile phones for hypothetical purchase. The stimuli were mobile phones because university students generally have both interest in and some knowledge about this product category. The phones were real phones sold in the USA in January 2006 and were drawn randomly from a pool of 50 in each trial, with the only restriction being that no phone appeared twice in the same trial. Each participant received exactly the same set of stimuli. To avoid biases due to previously established preferences and to force participants to collect relevant information from the information board rather than from their own memory, phone brand and model name were not displayed.

2.1.3 Design

Participants were randomly assigned to one of two groups. In the *without-list* condition, participants were asked to select the attributes on which they wanted information, without any further help from the experimenter. This was meant to enhance the realism of the decision situation. In the *with-list* condition, participants also first freely selected attributes but were then presented with a list containing all of the 33 available attributes. From this list they could choose any number of further attributes that had not occurred to them spontaneously.

2.1.4 Procedure

The experiment consisted of three phases: an attribute selection phase, an information acquisition and choice phase, and finally, a strategy identification phase. The participants completed the first two phases in a total of approximately 30 minutes and the last in approximately 25 minutes.



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Task # 3

To show the information of a specific cell, click twice. After making a choice, select the round button below.



	Product 1	Product 2	Product 3	Product 4
Price	330 Euro	356 Euro	124 Euro	141 Euro
Integrated Digital Cam...	yes	1.3 Mpixel	0.3 Mpixel	1152 x 864 pixel
Dimensions		107.5x49x24.5 mm	82x43x23 mm	
MP3 player		yes	yes	

Select a Product:

Figure 1: Screen-shot of the computer-based process-tracing measure used in Experiment 1 (after 12 cells had been clicked on).

Phase 1: Selection of Attributes. Participants were asked to state the attributes they were interested in and the experimenter entered them into the computer program. If participants had a clear idea of what they wanted but did not know the exact name of the attribute then the experimenter provided some assistance while trying not to influence the participant in any way regarding the selection of attributes. Whenever an attribute did not exist as specified by the participants (e.g., the attribute “usability”, which was not in the set of available attributes due to its high degree of subjectivity), they were informed that this information was not available.

After the participants in both conditions had completed the selection of the attributes — their final set of attributes is henceforth referred to as the *selected attributes* — they ranked these attributes with respect to their importance. They were informed that in the next phase the attribute they considered most important would appear on the top and the one they considered least important on the bottom of the information board. Moreover, participants in both conditions were informed that, once this ranking was complete, they could not access any information other than that concerning the selected attributes.

Phase 2: Information Acquisition and Choices. In this phase, the information on the selected attributes was presented in an attributes-by-alternatives matrix (see Figure 1), similar to the display used in the Mouselab procedure. The information could be obtained by using the computer mouse to click in the appropriate cells. Once a cell had been clicked on, the information contained within it

remained visible throughout the remainder of the trial.³ There were no constraints regarding the amount of or the order in which the information was considered. Participants could make a choice at any time during a given trial and could proceed to the next trial only after having selected one of the options. They could not go back to earlier trials.

Phase 3: Strategy Identification. In Phase 3, the participant and the experimenter interacted closely to gain an exact description of the participant’s strategy. Specifically, the participants were asked to explain and formalize their strategy in an exact enough manner so that it was possible to create an algorithm which could stand in for the decision maker in future choice situations.⁴ For instance, when participants wanted to eliminate “too expensive” alternatives the experimenter asked them to define precise cut-offs. Similarly, when the strategy required decisions based on subjective attributes such as design, the participants were asked to assign values to the alternatives for these attributes. Finally, when the strategy demanded the calculation of ratios or overall values, participants were asked to assign weights to the attributes. To reduce biases due to forgetting, we presented screen-shots of the information board of five of the trials. These screen-shots were taken when the participants had made

³Note that this is different from the standard form of Mouselab, where the cells close as soon as the mouse is moved away. We think that this form is easier to use for participants and, for the current purpose, we found no reason to adhere to the standard procedure.

⁴Because Levene’s test for the equality of variances proved significant, we adjusted the degrees of freedom accordingly.

Table 2: Participants' Strategies: Three Examples. Participant 5 used a purely additive strategy, the strategy of participant 37 was exclusively based on elimination, and participant 32 combined the two features.

<i>Participant 5:</i>	1) Look at the following attributes: Video clip playback with sound, FM stereo, Speech recording, Integrated speakerphone, VibraCall, Voice command, MMS, SMS, and Email support. Take the phone that possesses the greatest number of these attributes. 2) If there is a tie, choose one of the tied phones at random.
<i>Participant 32:</i>	1) Eliminate all phones that do not have SMS and whose standby time is less than 300 hours. 2) If the standby time of all phones is less than 300, choose the phone with the highest standby time. 3) Otherwise, assign the following attribute weights: VibraCall = 3, GPRS = 2, and Bluetooth = 1. For each attribute that the phone possesses, assign a value of 4. Multiply attribute value with attribute weight and choose the phone that has the highest score. 4) If there is a tie, choose the phone with the highest standby time.
<i>Participant 37:</i>	1) Eliminate all phones that do not have SMS and VibraCall. Select the cheapest phone. 2) If two or more products are equal in price, choose the smallest phone.

a choice (a procedure known as cued retrospective reporting; Gog et al., 2005). We selected these cuing trials, that were different for each participant, by first dividing the 30 trials into five equal segments and then randomly selecting one trial in each segment, excluding the very first trial. While proceeding through these cuing trials, the participants had to specify for some attributes how the values of the alternatives map onto specific values that could be used more easily within his or her strategy. To give an example, for the color attribute, the value "blue" might be assigned a value of 10, the value "black" a value of 5 and so on, depending on the participant's preferences. The experimenter was careful not to influence the participant in any way when assisting with the formulation of the strategy. This phase was completed once a strategy had been (a) described by the participant, (b) formalized and written down by the experimenter, and (c) verified by the participant. The outcome of this procedure will henceforth be referred to as a participant's *described strategy*.

2.1.5 Payment

To enhance participants' motivation to carefully describe and formalize the strategies they used (cf. Hertwig & Ortmann, 2001), they were informed that their remuneration depended on the number of times their strategies correctly predicted their choices. They received 1 Swiss Franc (1 SFR = approximately 0.78 USD at the time the study was conducted) for each correct prediction, with a minimum guaranteed amount of 10 SFR. This procedure resulted in an average payment of 22 SFR ($SD = 4$). Note that, while working on Phases 1 and 2, participants were not aware that they would be asked to formalize their strategy in Phase 3, or how their payment would be determined.

2.2 Results

Due to incomplete or faulty transcription of their strategies, six participants were excluded from the analyses, leaving 16 participants in the without-list condition and 15 in the with-list condition. Overall, participants included 5.7 (17%) of the 33 available attributes in the information board. The attributes that were selected most often were price (68%), digital camera (55%), size (52%), and mp3 player (39%) (details regarding the selected attributes available from the authors upon request). There was no difference between the two conditions (5.13 and 6.33 of 33, respectively; $t(19) = 1.49, p = .15$).

2.2.1 Described strategies

The strategies were classified according to several dimensions. In general, two types of strategies could be identified: elimination strategies and additive strategies. The former eliminate alternatives from the consideration set based on attribute values, for instance, when a particular attribute value does not reach the acceptance threshold specified by the participant (for an example, see Table 2, participant 37). Thus, they follow a logic similar to that of lexicographic strategies like the Elimination-By-Aspects strategy (Tversky, 1972), or that of the take-the-best heuristic (TTB, Gigerenzer & Goldstein, 1996; Gigerenzer, Hoffrage, & Kleinbölting, 1991). The number of attributes used for elimination varied between one and nine ($M = 3.03, Mdn = 3$). About a third of the participants (10 of 31) used *Just-Noticeable-Differences* when eliminating alternatives (see the Prediction accuracy section for further details). Strategies of the second type add the values (either weighted or not) of all or some attributes for each alternative to determine an overall score for the alternatives (e.g., Table 2, participant 5).

Of the 31 participants, almost all (30) used elimination

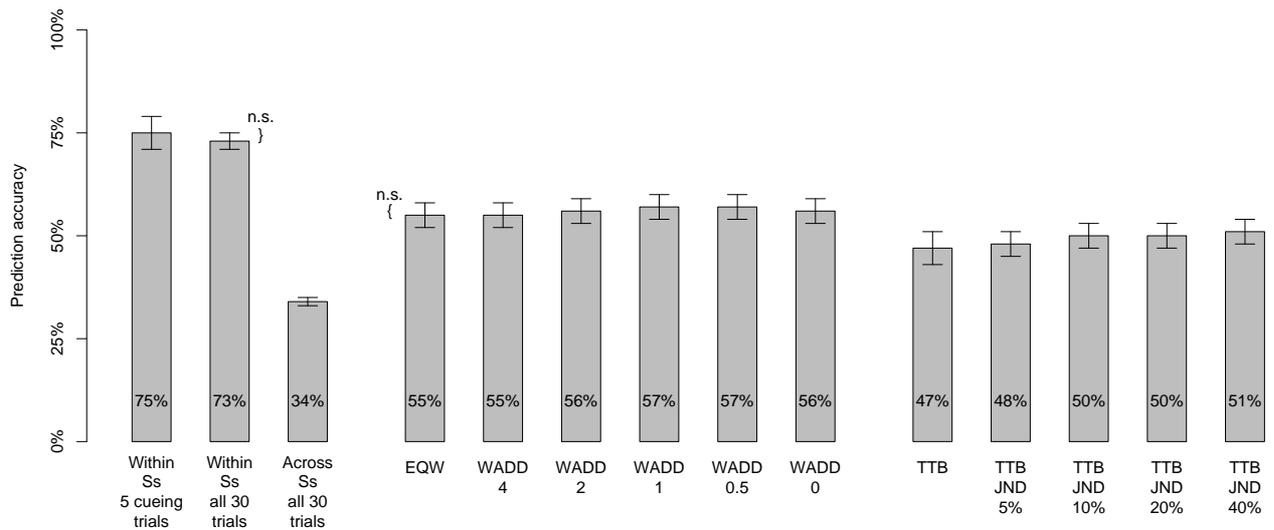


Figure 2: Percentage of choices correctly predicted by various decision strategies in Experiment 1, with standard errors. EQW = Equal Weighting, WADD = Weighted ADDitive, TTB = Take-The-Best, JND = Just Noticeable Difference.

and 23 (74%) added up attribute values in a linear fashion. Of those 23 participants, 17 (74%) assigned weights to the attributes according to their subjective importance (e.g., participant 32). Finally, 22 of all 31 participants (71%) combined the two types of strategy (e.g., participant 32).

2.3 Prediction accuracy

We calculated the degree to which the strategies described by the participants could predict their own choices. The averaged percentage of correct predictions across all 30 trials was 73% (Figure 2, second bar). Within the subset of the five cuing trials, the averaged prediction accuracy was virtually the same (75%, first bar). Note that these percentages are far greater than the 25% that would be obtained when choosing randomly. This indicates that the described strategies had reasonable predictive power.

Chance, however, may not be a good standard of comparison, because for a certain number of trials some mobile phones may be favored over others independently of the strategy used, especially when a phone dominated the others on that trial. Thus, a high number of correct predictions does not necessarily imply that participants were able to accurately describe their strategies. Therefore, we determined, as another benchmark against which the 73% correct predictions could be compared, the percentage of correct predictions that resulted from using a certain participant’s strategy to predict the choices of all other participants. Across all participants, this resulted in 34% correct predictions (Figure 2, third bar) — much closer to

chance level than to the percentage of correct predictions that resulted when using the participants’ own strategies to predict their choices. This result gives further evidence for the uniqueness of the participants’ strategies and indicates that they cannot be replaced easily by each other.

As a third benchmark, we determined the fit when modeling the observed choices with two established strategies from the literature.⁵ Specifically, we used six variants of the Weighted ADDitive (WADD) strategy, which is computationally demanding, and five variants of the take-the-best heuristic (Gigerenzer & Goldstein, 1996), a lexicographic strategy that applies one-reason decision making and that is hence quite easy to execute (see Figure 2). Each of the six variants of WADD calculated a score for each alternative by adding up the weighted values of each attribute and then choosing the alternative with the highest overall score.⁶ The variants differed with respect to the skewness of these weights. At one extreme, we used Equal Weights (EQW). At the other extreme, we used a set of noncompensatory weights, that is, the weight of the attribute that was ranked highest by a participant was bigger than the sum of the weights of all the lower-ranked attributes, the weight of

⁵Because our main goal was to test whether IAPT provides valid descriptions of strategies (rather than to model discrete choices with sophisticated statistical tools such as conjoint analysis), we used benchmark strategies that could be formulated and executed without fitting them to the participants’ choices.

⁶To be able to compare the values of the different attributes to each other, we first standardized these values by performing z-transformations and subsequently used these z values when multiplying by the weights of the attributes. In addition, the attributes weight, dimensions, and price were multiplied by -1, because lower values on these attributes are generally perceived to be better.

the attribute that was ranked second highest was bigger than the sum of all following weights, and so on (WADD₀).⁷

Take-the-best was originally formulated for inferential tasks in which two alternatives had to be compared to each other on a given criterion. Rieskamp and Hoffrage (1999; 2008) generalized TTB from two-alternative to multi-alternative choice tasks. For the preferential choice task used in the present experiment, this heuristic works as follows. It looks up the values on the most important attribute (as specified by the participant) and chooses the alternative with the best value. If two or more alternatives have this best value, then take-the-best eliminates all other alternatives from further consideration and compares the remaining alternatives on the second most important attribute, and so on (for another way of generalizing TTB, see Rieskamp & Hoffrage, 1999). TTB is a fast and frugal heuristic: it is easy to execute (once the cue ordering has been determined), and generally requires only a small amount of information.

However, it does not seem psychologically plausible to assume that information search is stopped in each and every case in which alternatives differ on the most important attribute. To capture an insight derived from early research on psychophysics, we created versions of TTB that operated with various Just-Noticeable-Differences (JND). A JND is the difference between the attribute values on two alternatives that is sufficiently small to treat the values as psychologically equal. We used five levels of JNDs that we applied to all selected attributes, namely 0%, 5%, 10%, 20%, and 40%. The five corresponding strategies are referred to as TTB₀, TTB₅, TTB₁₀, TTB₂₀ and TTB₄₀, respectively. For calculating these JNDs, the standard of reference was the alternative with the most attractive attribute value (in the respective trial). For instance, if the most important attribute of a particular participant was price, and the cheapest phone in a given trial cost 100 SFR, TTB₂₀ would have eliminated all phones that were more expensive than 120 SFR.

We predicted the 30 choices of each participant separately using each of these variants of WADD and TTB. The only difference in each strategy between participants was the ranking of the selected attributes, which was determined by the participants' responses in Phase

⁷Specifically, for this noncompensatory variant, the weight of a given attribute was $1/2^{(r-1)}$, where r is the rank of the selected attribute in the attribute hierarchy established by the participant. The other four variants (WADD₄, WADD₂, WADD₁, and WADD_{0.5}) were obtained by adding a constant (4, 2, 1, or 0.5, respectively) to each attribute weight in the noncompensatory set of weights. It is obvious that adding nothing to the attribute weights in the noncompensatory set of weights will maintain the noncompensatory structure, whereas adding a constant will reduce the relative differences between the attribute weights. As the constant approaches infinity, the relative differences approach zero, thereby ultimately turning the set of noncompensatory weights into a set of equal weights (i.e., EQW = WADD_∞).

1. As can be seen in Figure 2, the fit of the variants of WADD ranged between 55% and 57% correct predictions, suggesting that (consistent with Dawes, 1979) different weighting schemes did not make a big difference ($F(1, 44) < 1, p = .87$).⁸ The fit of the variants of TTB (averaged across all participants) ranged between 47% and 51%. Overall, the factor JND turned out to be significant ($F(2, 55) = 3.27, p = .049, MS_e = 6.309$). The most important result, however, is that for each of these established strategies the fit is much lower than for the described strategies (all t 's (30) > 5.7, all p 's < .001).⁹ Even when we selected the best fitting model for each participant, be it linear or lexicographic, the fit of the best-fitting model (66%) was still lower than the fit achieved when applying IAPT ($t(30) = 2.85, p = .007$).

2.3.1 Information search

Given that we used two different procedures (i.e., Mouse-lab and retrospective verbal protocol), we can verify whether the way in which participants searched for information is in agreement with the strategies they described. We focused on three main questions. First, are the described strategies reflected in the direction of the participants' search for information? Second, did they stop acquiring further information for a specific alternative once this alternative should be eliminated according to their described strategies? And third, does the frequency with which they accessed information on the selected attributes reflect the attributes' ranking that they had established in the first phase of the experiment?

Direction of information search. To examine the direction of the participants' information search we used the *Payne Index* (PI, Payne, 1976), which indicates whether the information search tends to proceed within or across attributes (alternative-wise vs. attribute-wise). An alternative-wise search pattern is associated with compensatory strategies whereas attribute-wise search is indicative of noncompensatory strategies. A score of 1.0 represents a fully alternative-based search whereas a score of -1.0 represents a fully attribute-based search. However, for asymmetrical matrices (i.e., when the number of attributes is not equal to the number of alternatives), the expected PI score for a random information search is not zero.¹⁰ Therefore, instead of taking zero

⁸The assumption of sphericity was violated, so the Greenhouse-Geisser correction was used in this analysis and the following analysis concerning TTB.

⁹We obtained the same result when we measured contingency between IAPT's predictions and observed choices using Goodman and Kruskal's (1954) λ .

¹⁰Another criticism is that the value of the PI varies as a function of the number of transitions in a particular trial. Therefore, it can lead to inaccurate conclusions and the values of the index observed under different combinations of attributes and alternatives or even different numbers of transitions are not directly comparable (Böckenholt & Hy-

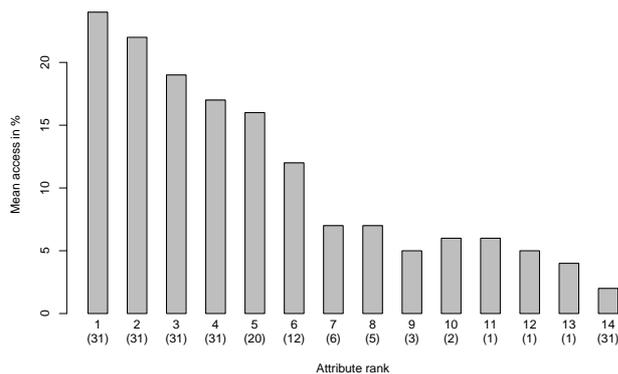


Figure 3: Mean percentage of accesses per attribute rank in Experiment 1. The numbers in parentheses below an attribute rank indicate how many participants used the corresponding number of attributes or more.

as a reference point to distinguish alternative-wise from attribute-wise search, we used the expected value of a random search in a particular matrix. To obtain these chance PIs, we first simulated 10,000 random sequences of information search for each participant and each trial, with the number of boxes opened by the simulation being equal to the number of boxes opened by the participant in the respective trial. We then calculated the PI for each sequence and, finally, the mean of these PIs, which served as the values for our chance PIs. It turned out that participants' chance PIs ranged between -0.03 and 0.62 . Twenty-two (71%) participants had an observed PI that differed significantly from their chance PI and that indicated an attribute-wise search, and 5 (16%) of the participants had an observed PI that indicated an alternative-wise search (the remaining 4 participants could not be classified). This finding is in line with other process tracing studies where it has been found that attribute-wise search patterns prevail (Ford et al., 1989).

Two other search measures also indicate the use of noncompensatory decision strategies: the depth and the variability of search (Ford et al., 1989). Participants accessed on average 76% of the information (range: 47% to 100%, $SD = 17%$) and accessed equal amounts of information on each alternative in only 35% of the trials.

Eliminations and information search. We also tested whether the elimination of alternatives as described by the participants' strategies was reflected in their information search. We assumed that, as soon as an alternative was eliminated because its value on one of the attributes failed to reach the threshold, the participant should not have acquired any more information about that alternative. Indeed, participants stopped search on a particular alternative after its elimination in one third (33%) of the

nan, 1994). Moreover, extreme PI values have a higher probability of occurrence than do intermediate values (see also Footnote 17).

trials. However, in the remaining two thirds (67%) at least one piece of information was acquired on an alternative even though it was already doomed to elimination.

Frequency of access. To test whether there is a relation between an attribute's rank and the frequency with which information about this attribute was accessed, we tested whether information about attributes that were reported to be more important was acquired more frequently than information about less important attributes. Generally, we found that the more important an attribute was rated on average, the more often it was accessed by the participants (see Figure 3).¹¹ However, it should be noted that attribute importance was confounded with the vertical position on the screen, which may have artificially enhanced this effect.

2.4 Discussion

Our main finding is that people facing a consumer choice situation are able to verbally formalize the strategy they used to make their decisions. The strategies identified with our method correctly predicted the observed choices in 73% of the cases, which is far greater than chance. Moreover, the identified strategies were able to predict the actual choices much better than several variants of linear and lexicographic strategies. Thus, our findings do not lend support to Nisbett and Wilson's (1977) claim that people's ability to observe and report upon higher order mental operations is underdeveloped — if existent at all. On the other hand, in 27% of the cases the decisions made by the described strategies did not correspond to the actual choices.

One simple reason for these prediction errors could be that at least some participants changed their strategy (including parameters of their strategy such as elimination thresholds) while proceeding through the choice phase. Such changes over time could not be considered in the analysis because in Phase 3 the participants were asked to formalize only *one* strategy. Although this explanation might potentially account for some misclassifications, the interviews did not provide much evidence for

¹¹To determine whether the number of accesses per rank was significantly different from each other, we conducted a within-participant one-way ANOVA with attribute rank as an independent variable. We used only the first four ranks for the analysis because this was the minimum number of attributes selected by all individual participants. The linear trend was highly significant ($F(1, 30) = 18.9, p = .001$). We then calculated the correlation between attribute rank and frequency of access separately for each participant, standardized these correlations by means of a Fisher transformation and calculated the mean over all participants. This standardized mean was re-transformed and resulted in a correlation of -0.83 . Moreover, we calculated the correlation between (1) each participant's correlation between attribute rank and frequency of access and (2) the number of attributes this participant accessed. This correlation was $.025$, indicating that participants who used only a small number of attributes did not spuriously inflate the former correlation.

such changes over time. Moreover, there was virtually no difference in the prediction accuracy between the first and the second half of the trials (72.5% and 73.6%, respectively; $t(30) = -0.43$, $p = .67$), which does not support the hypothesis that their strategies differed over time.

Another reason for the wrong predictions could be execution errors and unreliable choices. From the literature on bootstrapping, for instance, it is well known that laypeople and experts are often unable to execute a strategy reliably and without errors. This is also the major explanation why, in almost all studies on this issue, linear models outperform the people on whom these models are based (for a review, see Dawes, Faust, & Meehl, 1989). Moreover, in the second experiment we describe below, participants repeated half of the trials but made identical choices in both trials in only 73% of the cases. Future research could both check for participants' re-test reliability (see Experiment 2) and also confront them with those cases in which the strategy they had formulated in Phase 3 deviated from their own previous choices. It would be interesting to know whether they would change the formulation of the strategy or whether they would prefer to choose differently.

Finally, the mismatch between described strategies and observed processes could be due to the fact that the participants' strategy description resulted from an inductive inference, that is, from an attempt to characterize the conditions under which a specific alternative is chosen. This description should not be confused with the strategy the participants used when making the choices — maybe such strategies did not even exist in the first place and the descriptions were just constructed post-hoc, after the experimenter asked the participants to do so. Likewise, we cannot exclude the possibility that participants used configural strategies (Garcia-Retamero, Hoffrage, & Dieckmann, 2007) in Phase 2 but did not report this in Phase 3 as such strategies are complex and thus hard to describe.

2.4.1 Information search vs. described strategies

Many of the described strategies are in line with previous research stating that people often start with a non-compensatory strategy to reduce the number of alternatives in the choice set, and then switch to a compensatory strategy to make a decision between the remaining options (Billings & Marcus, 1983; Ford et al., 1989; Payne, 1976; see, however, Glöckner & Betsch, 2008). Such two-step strategies pose a challenge for any attempt to contrast the described strategies and the choices they predict with the information acquisition data. And in fact, our findings are mixed.

First, the information search measures generally indicated that participants engaged in more noncompensatory search, which is consistent with the finding that most of

the described strategies contained noncompensatory elements. However, beyond such indication for noncompensatory processing, no correspondence could be found between the described strategies and other information search measures (depth and variability of search). Second, participants' search for information reflects, by and large, their ranking of the attributes. Third, however, participants very frequently (i.e., in 66% of the trials) looked up information for alternatives that they should have already eliminated according to the strategy they described.

Given that the protocol and information search data converge only to a certain degree, the question arises as to what extent a given strategy actually directs the search for information, and, ultimately, how valid and specific the conclusions are that can be drawn from information search data (for a critique on information search techniques see Bröder, 2000). A possible explanation for the discrepancy between people's actual search behavior and the search behavior that is expected given their strategies is that the acquisition of information serves the purpose of giving a general overview of the choice options rather than providing only the information that is needed for the execution of a decision strategy. It may be that the particular strategy is generated and executed only after having obtained a certain amount of information. Considering the fact that strategy choice is often adaptive (cf. Payne et al., 1993; Bröder & Newell, 2008), it is reasonable to assume that a decision maker first acquires a certain amount of information and then decides on a strategy (or just certain parameters of it such as thresholds).

Overall, the first test of IAPT yielded reasonably satisfactory results. In Experiment 2, we sought to further develop and eventually improve it by integrating eye tracking technology.

3 Experiment 2

One of the fastest and most natural ways for humans and many other species to acquire information about something is to simply look at it. Eye movements are very fast, accurate, and, due to their spontaneity, they cannot easily be censored by the participants. Consequently, the recording of eye movements is expected to yield very reliable and complete data about information search. The researchers' optimism concerning this technology is reflected in a large number of studies that used eye tracking in a variety of disciplines, such as neuroscience, psychology, and marketing, to name just a few (see Duchowski, 2002; Rayner, 1998, and Wedel & Pieters, 2007 for reviews). Of interest for our purposes is that the information about the choice alternatives can be presented in virtually the same matrix as in the Mouselab setup (with its cells uncovered), which makes direct comparisons be-

tween the two methods feasible. Despite the evident similarity between these two information search techniques and the possible advantages of eye tracking (see Table 1), very few studies directly compared them. In a study by van Raaij (1977), 20 housewives chose among thirteen alternative brands of coffee, each described on four attributes. In a first session, they examined actual product packages and their eye movements were recorded. Four months later, the same participants now made their choices using an information board. Although choices were faster with eye tracking, participants acquired more information in this condition (more than half of the available information) than in the information board condition (about a third of the available information). More than half of the searched information was accessed twice or more with eye tracking, but no reacquisitions were observed in the information board condition.

Lohse and Johnson (1996) compared Mouselab with eye tracking using apartment selection tasks and gambles. As predicted, they found meaningful differences between the two methods. With eye tracking, participants were faster, had more fixations, and more reacquisitions but examined a smaller percentage of the total information and their information search showed a more variable pattern. Moreover, participants tended to search more attribute-wise with eye tracking than with Mouselab. The authors concluded that the recording of eye movements has several advantages: it is faster and less demanding for the participants, it leads to more accurate task performance in choices between gambles (especially when processing demands are increased), and it is better suited for larger problems (i.e., more alternatives and/or attributes). Similarly, in his comparison of several process tracing methods Russo (1978) also came to the conclusion that eye tracking has advantages not offered by other methods. Moreover, he argues for a simultaneous use of eye tracking and verbal protocols.

Bearing on these results, in Experiment 2 we used eye tracking in addition to Mouselab within IAPT to test for possible influences of the research method on the participants' cognitive processes and behavior, and, ultimately, whether the use of eye tracking increased the proportion of observed choices that were correctly predicted by the strategies revealed by our method. A higher percentage of correct predictions and a higher convergence between the described strategies and the information search data would be indicative of such an improvement. A further, minor point of interest in Experiment 2 was the phenomenon of choice deferral, that is, the decision *not* to select any of the presented options. As opposed to the forced choice paradigm used in most of the studies on preferential choice (and also in our first experiment), we explicitly wanted to give our participants the possibility to defer choice in any given set. We think that this is es-

sential for the type of choice situation examined in our experiments because in real life, people frequently (e.g., more than 95% of the time, Sismeiro & Bucklin, 2004) decide not to buy any of the options available in a certain (online) store.

3.1 Method

3.1.1 Participants

Participants were 27 students (5 female and 22 male) of the Swiss Federal Institute of Technology of Lausanne (EPFL) and the University of Lausanne with a mean age of 24.6 years ($SD = 3$).

3.1.2 Task and stimuli

As in Experiment 1, the task was to select a mobile phone for purchase out of a set of four. The four phones presented in each trial were drawn randomly¹² from the pool of phones used in the first experiment (except for one which disappeared from the market in the meantime).

3.1.3 Apparatus

For Phases 1 and 2 of IAPT we used a computer-based process tracing measure very similar to the one in Experiment 1. It was synchronized with the eye tracker so that stimuli presentation in both conditions could be done with the same program.

We used the iView XTMHi-Speed eye tracker, manufactured by SensoMotoric Instruments (SMI; Teltow, Germany), which works at a sampling rate of 1250 Hz, a spatial resolution of 0.01° and a gaze position accuracy of 0.25°. Only one eye was recorded and the gaze position was determined using the pupil and corneal reflection method. The system has a chin rest to avoid head movements. We used a 17-inch screen for stimulus presentation and the distance between the participants' eyes and the screen was about 50 cm. The illumination of the screen was kept constant and room lighting did not interfere with the recording capabilities of the eye tracker.

3.1.4 Design and procedure

Each participant experienced both of the two conditions, Mouselab (ML) and eye tracking (ET) (in Phase 2 of

¹²This random process had the following constraints: (a) Any set of four phones consisted of four distinct phones, that is, no phone appeared more than once in a given set. (b) Half of the trials (randomly determined) used the same phones in both conditions, but in a different, random order. Here, we excluded the order that was the exact reverse of the original order as well as all the orders where two phones were next to each other in the same order as in the first condition. In addition, the first trial of the first condition was never repeated and the last trial of the first condition was never repeated as the first trial in the second condition.

IAPT), with the order counterbalanced (13 participants began with ML and 14 with ET). Each condition consisted of 12 trials. Half of the trials of the first condition were repeated in the second condition, but with a different, random ordering of the alternatives (see Footnote 11 for details). Participants completed the first two phases in approximately 30 minutes and the last one in approximately 25 minutes. In addition, between five and ten minutes were needed for the calibration of the eye tracker.

Except for the changes related to the new research questions and some minor modifications, the general procedure was identical to the one of the first experiment. The changes were as follows. First, because there were no differences between the with-list condition and the without-list condition in Experiment 1, the list of attributes was now shown to all participants. Second, given that many participants in Experiment 1 requested information about phone brand and name we now replaced the image of the phone with this information. Third, to open a cell it was sufficient to move the mouse over it (instead of clicking as in Experiment 1). The cell closed again when the mouse was moved away. This modification was necessary to be able to compare the data from Mouselab and eye tracking. Fourth, we increased the size of the cells so that in Phase 2 participants could not read the information contained in the cells neighboring the fixated cell. Due to size limitations of the screen, the maximum number of attributes that could be selected was ten. The cell size was kept constant irrespective of the number of selected attributes, with each cell being 60 mm wide and 33 mm high (visual angles of 6.8° and 3.8°, respectively). Because our aim was to keep the situation as natural and realistic as possible, we informed the participants about this limit only when the number of attributes they selected exceeded this number. Apart from the fact that the cells were initially covered in the Mouselab condition, the interface was completely identical in both conditions. Fifth, in Phase 2 participants were now given the possibility to choose none of the four alternatives. To defer choice, participants had to click a button labeled “Choose none of these.” After that, they had to indicate why they deferred by selecting one of two reasons: “Because none of them is good enough” or “Because I am not sure which is the best.” There was no cost to deferring choice, and participants could do so as often as they wished. Sixth and finally, instead of presenting screen-shots of the information board (i.e., cuing trials), in Phase 3 we tried to enhance recall by letting the participants repeat one of the trials of the first condition of Phase 2. This repeated trial was randomly selected from the set of 12 (with the exception of the first trial). After that, participants were presented with an empty matrix so that the values shown in the repeated trial did not influence the participant when describing his or her strategy.

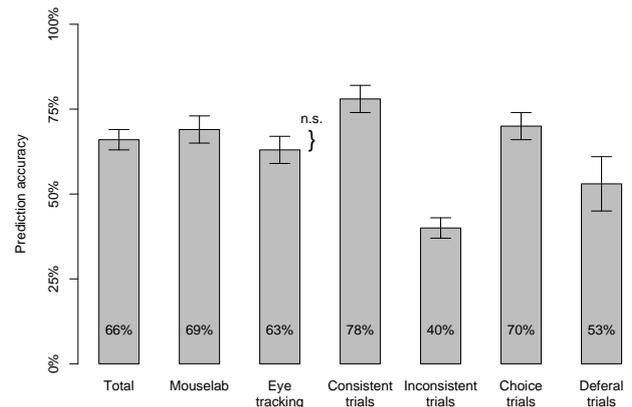


Figure 4: Percentage of choices correctly predicted by the participants’ decision strategies in Experiment 2. The vertical bars denote standard errors.

3.1.5 Payment

In the introduction to Phase 3, participants were informed that they will receive 1.50 Swiss Francs (1 SFR = approximately 0.82 USD at the time the study was conducted) for each correct prediction of their described strategy, with a minimum guaranteed payment of 10 SFR. The average payment was 25 SFR ($SD = 6$).

3.2 Results

On average, participants selected 22% of the available attributes. The attributes that were determined most often were price (96%), size (85%), stand-by time (59%), and digital camera (56%) (further details available from the authors upon request). All analyses regarding differences between the conditions were done using a mixed design ANOVA including the within-participant variable of condition and the between-participants variable of order.

3.2.1 Deferrals

In 31% of the trials of the ML condition and in 30% of the trials of the ET condition, participants did not choose any of the phones presented. This is in line with most of the literature on choice deferral (e.g., Dhar, 1997; White & Hoffrage, in press). The deferral option was used by all but two participants (93%). For most of the deferrals (86%, across conditions) participants indicated that none of the available options was good enough and for 14% they indicated that they were not sure which option was best.

3.2.2 Described strategies

Again, we found the two general types of strategies: elimination and additive. Of the 27 participants, almost all (26

of 27; 96%) eliminated alternatives during their decision making process, based on between one and nine ($M = 4.77$, $Mdn = 5$) attributes. Eight participants used JNDs. Adding up attribute values in a linear fashion was used by 18 of 27 (67%) participants. Of those 18 participants who used an additive strategy, 10 (63%) assigned weights to the attributes according to their subjective importance. Finally, 17 of the 27 participants (63%) combined elimination with an additive strategy.

3.2.3 Prediction accuracy

The degree to which the strategies described by the participants could predict their own choices (66%) was slightly, but not significantly, higher ($F(1, 25) = 3.96$, $p = .07$, $MS_e = 139.8$) in the ML condition (69%) than in the ET condition (63%). In the repeated trials, participants made the same decision in both instances in only 73% of the cases. Note that the prediction accuracy was considerably higher in the consistent trials (78%) than in the inconsistent trials (40%) ($F(1, 22) = 52.3$, $p < .001$, $MS_e = 296.3$). Moreover, the prediction accuracy was significantly higher in the trials where the participants selected an option (70%) than in the trials in which choice was deferred (53%, $F(1, 18) = 4.81$, $p = .042$, $MS_e = 812.7$).¹³

3.2.4 Information search

We did an in-depth analysis of the information search data to check for possible differences between ML and ET. As in the first experiment, we also verified whether the described strategies were reflected in the information search data. In particular, we focused on the following: (1) the time spent per trial, (2) the amount of information acquired, (3) the information considered by the participants compared to the information needed by the strategy they described, (4) the direction of the information search, and (5) the correlation between percentage of accesses and attribute rank. The scanpaths depicted in Figure 5 exemplify some of the results described in the following.

Due to some calibration problems that we detected only when analyzing the ET data, seven participants had to be excluded from all analyses involving information search data except for time. Half of the remaining 20 participants started with Mouselab. The ET data was analyzed using the software BeGaze (SMI). We analyzed fixation position, duration, and sequence (i.e., scanpath). Fixations of less than 100 ms were excluded from the analysis.¹⁴

¹³For all analyses concerning choices and deferrals, the one participant who never chose and the six participants who did not defer at least once in each condition were excluded.

¹⁴12BeGaze calculates fixations by subtracting saccades and eye

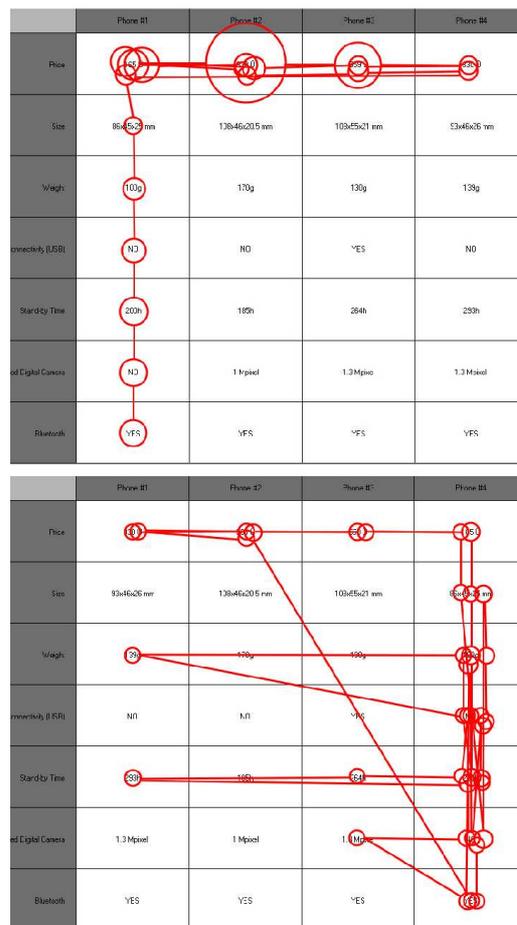


Figure 5: The scanpath of one participant in the ML condition (a) and in the ET condition (b) of Experiment 2. The size of the circles correspond to the time a box remained open in the ML condition and the fixation time in the ET condition. The trials were identical in both conditions with the exception that the positions of Phones 1 and 4 were swapped. The participant completed the trials in 44 sec (ML) and 17 sec (ET).

Time. In general, participants spent significantly more time per trial in the ML condition than in the ET condition (36.73 vs. 20.41 seconds, respectively, $F(1, 25) = 72.0$, $p < .001$, $MS_e = 52.36$). There was a significant interaction between condition and order ($F(1, 25) = 30.5$, $p < .001$, $MS_e = 52.36$), but in both orderings the effect of condition was significant and in the same direction.

The time in which a trial was completed did not de-

blinks from the original gaze stream. For a saccade to be detected, three conditions had to be satisfied: (a) peak values of velocities in the gaze stream were greater than 75° , (b) the single peak value of velocity lay in the middle 60% between the start and end of the event and (c) the duration of the event was more than 1 ms. An eye blink was detected when the conditions for saccades were satisfied and when the change in the pupil diameter exceeded an internally defined threshold.

pend on whether a phone was chosen or choice was deferred (29.47 vs. 29.49 seconds, respectively; $F(1, 22) < 1, p = .95$). However, participants needed significantly more time when they indicated that they were not sure which option was best as compared to those trials for which they indicated that none of the options was good enough (39.82 vs. 26.54 seconds, respectively, $F(1, 7) = 13.1, p = .009, MS_e = 51.39$).¹⁵

Amount of information. Our next analysis concerns the amount of information the participants accessed. First, we distinguished between the *total* number of accesses or fixations (i.e., including reacquisitions of the same information) and the number of *different* cells accessed. As expected, participants had significantly more total accesses in the ET condition than in the ML condition (41.83 vs. 22.35, respectively, $F(1, 18) = 44.5, p < .001, MS_e = 85.35$). The effect of condition interacted with the order ($F(1, 18) = 14.7, p < .001$); looking at the simple effects of condition for each order showed that this was the case for both orderings but the effect just failed to reach significance when ML was the first condition (ML first: $F(1, 18) = 4.00, p = .06$; ET first: $F(1, 18) = 55.2, p < .001$). However, the number of different cells accessed was very similar in the two conditions. On average, participants accessed 15.45 (59%) cells in the ML condition and 16.73 (63%) cells in the ET condition ($F(1, 18) = 3.01, p = .01, MS_e = 5.393$). Again, the effect of condition interacted with the order ($F(1, 18) = 15.0, p < .001$). When ET was the first condition, participants searched for significantly more information in the ET condition than in the ML condition ($F(1, 18) = 15.7, p < .001$). However, this was reversed for the opposite ordering, but here the difference between the ML and ET condition was not significant ($F(1, 18) = 2.28, p = .15$).

Second, we looked at the reacquisition rate, which is the percentage of accesses that were reaccesses of previously seen information (in the same trial). There was a significant difference between the two conditions, with a reacquisition rate of 27% in the ML condition and 57% in the ET condition ($F(1, 18) = 126, p < .001, MS_e = 72.34$). Again, the effect of condition interacted with the order ($F(1, 18) = 13.1, p = .002$), but in both orderings the effect was significant and in the same direction.

Information considered. Again, we compared the information accessed by the participants with the information that their described strategies needed for execution. Regardless of the condition, participants accessed about 50% more information than prescribed by their strategy. Interestingly, out of all the information that was needed by the strategy, only about 18% was *not* accessed.

Direction and variability of information search. As already mentioned in the first experiment, the PI has been

subject to some criticism (Böckenholt & Hyman, 1994; Stokmans, 1992). As a reaction, Böckenholt and Hyman (1994) developed a standardized version of the PI, the SM index.¹⁶ We calculated the SM index for each participant and each condition and found the following. In the ML condition, 17 of 20 (85%) participants had a SM score that indicated attribute-wise search whereas only one participant had a SM score that indicated alternative-wise search. Two participants had non-significant SM scores. In the ET condition, 14 participants searched attribute-wise and again only one participant (not the one of the ML condition) searched alternative-wise. In this condition, five participants had non-significant SM scores. There was no significant difference in the SM scores between the two conditions ($F(1, 18) < 1, p = .39$).

In only 14 (6%) of all trials of the ML condition (across all participants) and in none of the trials of the ET condition did the participants access an equal proportion of information about each alternative (cf. variability of search). This high degree of selectiveness was more pronounced in the ET condition than in the ML condition ($F(1, 18) = 34.8, p < .001, MS_e = 2.579$).

Frequency of access. As in Experiment 1, we compared the frequency of access with the attribute ranks assigned by the participants. There were no significant differences between the two conditions ($F(1, 19) < 1, p = .71$) and the correlation ($r = -0.91$) very closely resembled that found in Experiment 1 ($r = -0.83$). Interestingly, the time participants spent on a particular piece of information did not depend on the importance they assigned to the attribute containing this information ($F(3, 54) < 1, p = .90$ and $p = .88$ for the ML and ET condition, respectively).¹⁷

Summary. The analysis of the process data yielded the following results. First, participants needed significantly less time to complete a trial in the ET condition than in the ML condition. When participants deferred choice, they spent more time on a trial when they reported deferring because none of the phones was good enough than when they reported that they were unsure which phone was best. Second, participants had a significantly higher number of accesses (including reacquisitions) in the ET condition than in the ML condition.

¹⁶This index is a function of the differences between the observed alternative-wise and attribute-wise transitions. For any N , the mean is 0 and the variance is 1 when the search pattern is random. For a large number of transitions, the SM approximates a standard normal distribution, that is, unlike for the PI, extreme values have a lower probability of occurrence than intermediate values. The SM index is not without criticism either (e.g., Ball, 1997; Harte & Koele, 2001), but we felt that it is sufficiently informative for our purposes. Note that it is applicable only for matrices where the cells do not remain open once they have been clicked on and could thus not have been used for Experiment 1.

¹⁷This analysis was calculated based on only the first four ranks (thereby excluding one participant who selected less than four attributes).

¹⁵Only the 9 participants who indicated that they deferred for both reasons (on different trials) were used for this analysis.

However, there was no difference between the two conditions regarding the number of different cells accessed (i.e., depth of search). Consequentially, the reacquisition rate was far higher in the ET condition than in the ML condition. Third, when comparing information search and the described strategies we found that participants accessed significantly more information than their strategy required for execution, without any difference between the conditions. However, they obtained almost all the necessary information for their strategy to work. Fourth, the pattern of search also did not differ significantly between the two conditions. The search was generally more attribute-wise and selective (indicating noncompensatory processing), which was in line with the nature of the described strategies. However, the participants' search was significantly more selective in the ET condition than in the ML condition. Fifth and finally, participants' search for information reflected, by and large, their ranking of the attributes according to their importance.

3.3 Discussion of Experiment 2

In Experiment 2, we successfully replicated our finding that the strategies identified with IAPT have good predictive power. In 66% of the cases, the described strategies correctly predicted the participants' choices, which is very similar to the 73% we observed in Experiment 1. Moreover, it appears that many of the incorrect predictions can be attributed to inconsistent choices rather than to unreliable strategy descriptions: participants made consistent choices in only 73% of the trials and the prediction accuracy was considerably higher (i.e., 78%) when only the consistent trials were taken into account. Thus, it appears that some or even many of the incorrect predictions of the participants' strategies can be explained by inconsistent behavior during the choice phase.

Very similar to what we found in the first experiment, the described strategies were only partly reflected in the information search data. The analysis of the pattern, variability and depth of search measures did not lead to new insights, and, in an analysis slightly different to the one performed in the first experiment, we found that participants accessed a lot of information that was not needed by the described strategy. However, they rarely failed to obtain information that *was* required by their strategy, which demonstrates at least some convergence between the information search measures and the verbal protocol.

Regarding choice deferral, the participants' strategies were far less successful at predicting choice deferrals (i.e., 53%) than the choice of a concrete alternative (i.e., 70%). It seems that participants were better at giving reasons for their choices than for their deferrals.

The comparison between the two information search

techniques, Mouselab and eye tracking, yielded the following picture. Eye tracking was generally faster, that is, even though participants had a higher number of accesses, they needed less time to complete a trial. Furthermore, the information search was more selective (i.e., there was a higher variability of search) in the eye tracking condition. However, participants searched for virtually the same proportion of the total information in both conditions, and the difference in the number of accesses can almost completely be attributed to the fact that participants simply reaccessed some cells several times. Many of these reaccesses might have served the purpose of validating a tentative choice (which was often visible in the scanpath of the participants' eye movements), which corresponds to the validation stage reported by Russo and Leclerc (1994). Moreover, the pattern of search and the relation of attribute rank and frequency of access did not differ between the Mouselab and the eye tracking condition. Our results are quite similar to the findings of van Raaij (1977) and Lohse and Johnson (1996) except for the following: van Raaij's participants acquired more different items with eye tracking than with the information board, whereas our participants had a very similar depth of search in both conditions. Lohse and Johnson found a slight difference in search pattern (i.e., more alternative-wise search with eye tracking than with Mouselab) and their participants unexpectedly searched for less information with eye tracking. In contrast, we did not find any differences on these variables.

What can we now conclude about the use of eye tracking with IAPT? It appears that this methodology improves neither the exactness of the description of the cognitive processes nor the quality of the results concerning the information search. Although this method allows for a more natural way of searching for information, it does not provide more informative data than does Mouselab. With eye tracking, there is considerable noise in the information search data due to the fact that sometimes it is impossible to separate voluntary information acquisitions from random fixations that occurred while the participant was thinking. With Mouselab, the process of information acquisition seems to be more systematic, which could be a result of some reactivity of the method on the one hand (see Glöckner & Betsch, 2008), but which leads to data that is easier to interpret on the other hand. In sum, despite the technological innovations of the eye tracking technology, Mouselab is still much easier to set up and to use. Mouselab requires no calibration and works with virtually every participant, whereas eye tracking requires exclusion of some of the participants because no reliable calibration can be achieved. (In our experiment, this was the case for seven of the 27 participants; 26%.) In addition, with Mouselab, many participants can be run at the same time and even over the internet with a ready-to-use

program called MouselabWEB (Willemsen & Johnson, 2006). Given that the advantages of eye tracking were not very pronounced in our experiment, we conclude that Mouselab is the more convenient and efficient method for this kind of task.

4 General discussion and conclusions

In two experiments we have shown that our new interactive process tracing method is a valid technique for identifying human decision processes. We were able to replicate various findings in the related literature and achieved a detailed description of the strategies people used when making a purchase decision. Similar to Bettman (1970), Larcker and Lessig (1983), Einhorn et al. (1979), and Li et al. (2000), we showed that models constructed based on verbal reports describe the participants' behavior quite well.

A more critical finding that we observed in both experiments is that people's search for information often deviated from what would be expected given the described strategy.¹⁸ Moreover, it appears that the data obtained with Mouselab and eye tracking are on a rather general level and, consequently, are not specific enough to allow for discrimination among candidate decision strategies. This casts some doubt on the general usefulness of information search techniques, at least in this context. It may even be that the link between information search and cognitive processes is less pronounced than commonly assumed.¹⁹ Thus, we believe that it is sensible to use ver-

¹⁸It should be noted that this is certainly not the first study that revealed a mismatch between the process that is expected from the identified strategy and the process that is actually observed. Rieskamp and Hoffrage (2008), for instance, found that participants who were classified as selecting a weighted additive strategy did not search for information alternative-wise as one would expect from the description of their strategies but instead searched for information attribute-wise. Following Tversky (1969), these authors speculated that participants, when applying a WADD strategy, did not compute a score for each alternative sequentially but instead computed several scores in parallel, one for each alternative, by looking up information attribute-wise and by using the information of each additional attribute to update the scores. This procedure appears cognitively more demanding, because all scores have to be maintained in memory. However, it has the advantage that at any point during the evaluation, all alternatives are comparable on a subset of attributes, so that when making inferences under time pressure, a decision can be made on the basis of the preliminary scores. Likewise, Rieskamp and Otto (2006) also found that participants searched attribute-wise, even though their inferences could best be predicted by WADD. Finally, Johnson, Schulte-Mecklenbeck, and Willemsen (2008) found a mismatch between the search order prescribed by the priority heuristic (Brandstätter, Gigerenzer, & Hertwig, 2006) and their participants' information search as observed with Mouselab.

¹⁹A possible reason for this is that Mouselab alters the way information is searched and processed. For instance, Glöckner and Betsch (2008) found that under time pressure, participants switched from compensatory to non-compensatory processing only when Mouselab was

used. In contrast, when an "open" matrix was used (i.e., no covered information), participants used an (automatic) WADD strategy and they did this extremely fast (i.e., 1.5 s on average). The authors suspect that the well documented switch from compensatory to non-compensatory processing when under time pressure might be partially induced by the method rather than being something typical for human decision making. However, the fact that we did not observe a difference in the search patterns between Mouselab and eye tracking makes this explanation less plausible, at least for our experiments. Moreover, our choice problems were much more complex than those used by Glöckner and Betsch (2008) (i.e., more attributes and alternatives, many continuous attributes instead of dichotomous attributes) and we did not impose time pressure.

bal protocols in addition to the search measures to obtain data from two different sources that, it seems, highlight two qualitatively different aspects of the decision making process. For IAPT this means that in particular Phases 1 and 3 are crucial for the detection of cognitive decision processes. However, we nevertheless think that the use of information search techniques is still worthwhile when integrated into a multimethod approach such as IAPT, where the data of one method can be validated with the data of the other.

In conclusion, our findings demonstrate that IAPT is a useful tool for the description of decision processes. In the future, this method could be used in other domains and with different participant populations to learn more about domain specificity and inter-individual differences in this context. Moreover, in addition to the purely descriptive use of IAPT, we can also imagine it being used for applied purposes. For instance, the IAPT technique could prove beneficial for the creation of purchase environments, especially regarding the presentation of product information (e.g., selection and positioning of attributes presented to consumers). Another possibility would be to use the obtained findings for the development of decision support systems, such as interactive choice aids that can be implemented in consumer websites (e.g., Edwards & Fasolo, 2001; Häubl & Trifts, 2000). These choice aids facilitate the process of choosing by directly assisting the consumers in the execution of typical decision strategies (e.g., by providing tools for quickly eliminating alternatives or calculating overall values). Thus, IAPT does not only provide valid descriptions of decision strategies, it also has rich potential for applications.

References

- Abelson, R., & Levi, A. (1985). Decision making and decision theory. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology. 1. Theory and method*. New York: Random House.
- Arch, D. C., Bettman, J. R., & Pakkar, P. (1978). Subjects' information processing in information display board studies. *Advances in Consumer Research*, 5, 555–560.

used. In contrast, when an "open" matrix was used (i.e., no covered information), participants used an (automatic) WADD strategy and they did this extremely fast (i.e., 1.5 s on average). The authors suspect that the well documented switch from compensatory to non-compensatory processing when under time pressure might be partially induced by the method rather than being something typical for human decision making. However, the fact that we did not observe a difference in the search patterns between Mouselab and eye tracking makes this explanation less plausible, at least for our experiments. Moreover, our choice problems were much more complex than those used by Glöckner and Betsch (2008) (i.e., more attributes and alternatives, many continuous attributes instead of dichotomous attributes) and we did not impose time pressure.

- Ball, C. (1997). A comparison of single-step and multiple-step transition analyses of multiattribute decision strategies. *Organizational Behavior and Human Decision Processes*, *69*, 195–204.
- Bettman, J. R. (1970). Information processing models of consumer behavior. *Journal of Marketing Research*, *7*, 370–376.
- Billings, R. S., & Marcus, S. A. (1983). Measures of compensatory and noncompensatory models of decision behavior: Process tracing versus policy capturing. *Organizational Behavior and Human Performance*, *31*, 331–352.
- Böckenholt, U., & Hynan, L. S. (1994). Caveats on a process-tracing measure and a remedy. *Journal of Behavioral Decision Making*, *7*, 103–117.
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*, 409–432.
- Bröder, A. (2000). A methodological comment on behavioral decision research. *Psychologische Beiträge*, *42*, 645–662.
- Bröder, A., & Newell, B. (2008). Challenging some common beliefs: Empirical work within the adaptive toolbox metaphor. *Judgment and Decision Making*, *3*, 205–214.
- Brucks, M. (1988). Search Monitor: An approach for computer-controlled experiments involving consumer information search. *Journal of Consumer Research*, *15*, 117–121.
- Dhar, R. (1997). Consumer preference for a no-choice option. *Journal of Consumer Research*, *24*, 215–231.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582.
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668–1674.
- Duchowski, A. T. (2002). A breadth-first survey of eye tracking applications. *Behavior Research Methods, Instruments, and Computers*, *1*, 1–16.
- Edwards, W., & Fasolo, B. (2001). Decision technology. *Annual Review of Psychology*, *52*, 581–606.
- Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, *86*, 465–485.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Ford, J. K., Schmitt, N., Schlechtman, S. L., Hulst, B. M., & Doherty, M. L. (1989). Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes*, *43*, 75–117.
- Garcia-Retamero, R., Hoffrage, U., & Dieckmann, A. (2007). When on ecue is not enough: Combining fast and frugal heuristics with compound cue processing. *Quarterly Journal of Experimental Psychology*, *60*(9), 1197–1215.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.
- Glöckner, A., & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition: Learning, memory, and cognition*, *34*, 1055–1075.
- Goodman, L. A. and Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*, 732–764.
- Harte, J. M., & Koele, P. (1995). A comparison of different methods for the elicitation of attribute weights: structural modeling, process tracing, and self-reports. *Organizational Behavior and Human Decision Processes*, *64*, 49–64.
- Harte, J. M., & Koele, P. (2001). Modelling and describing human judgement processes: The multiattribute evaluation case. *Thinking and Reasoning*, *7*, 29–49.
- Häubl, G., & Trifts, V. (2000). Consumer decision making in online shopping environments: The effects of interactive decision aids. *Marketing Science*, *19*, 4–21.
- Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, *24*, 383–403.
- Huber, O., Wider, R., & Huber, O. W. (1997). Active information search and complete information presentation in naturalistic risky decision tasks. *Acta Psychologica*, *95*, 15–29.
- Johnson, E. J., Schulte-Mecklenbeck, M., & Willemsen, M. C. (2008). Process models deserve process data: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, *115*, 263–272.
- Larcker, D. F., & Lessig, V. P. (1983). An examination of the linear and retrospective process tracing approaches to judgment and modeling. *Accounting Review*, *85*, 58–77.
- Li, S., Shue, L., & Shiue, W. (2000). The development of a decision model for liquidity analysis. *Expert Systems with Applications*, *19*, 271–278.
- Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organizational Behavior and Human Decision Processes*, *68*, 28–43.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than

- you can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: an information search and protocol analysis. *Organizational Behavior and Human Performance*, 16, 366–387.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.
- Payne, J. W., Braunstein, M. L., & Carroll, J. S. (1978). Exploring predecisional behavior: An alternative approach to decision research. *Organizational Behavior and Human Performance*, 22, 17–44.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Riedl, R., Brandstätter, E., & Roithmayr, F. (2008). Identifying decision strategies: A process- and outcome-based method. *Behavior Research Methods*, 40, 795–807.
- Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell. In G. Gigerenzer, P. M. Todd, and the ABC Research Group, *Simple heuristics that make us smart* (pp. 141–167). New York: Oxford University Press.
- Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, 127, 258–276.
- Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, 135, 207–236.
- Russo, J. E. (1978). Eye fixations can save the world: A critical evaluation and a comparison between eye fixations and other information processing methodologies. In H. K. Hunt (Ed.), *Advances in consumer research* (pp. 561–570). Ann Arbor, MI: Association for Consumer Research.
- Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory and Cognition*, 17, 759–769.
- Russo, J. E., & Leclerc, F. (1994). An eye-fixation analysis of choice processes for consumer nondurables. *Journal of Consumer Research*, 21, 274–290.
- Russo, J. E., & Rosen, L. D. (1975). An eye fixation analysis of multialternative choice. *Memory and Cognition*, 3, 267–276.
- Sismeiro, C., & Bucklin, R. E. (2004). Modeling purchase behavior at an e-commerce web site: A task-completion approach. *Journal of Marketing Research*, 41, 306–323.
- Stokmans, M. (1992). Analyzing information search patterns to test the use of a two-phased decision strategy. *Acta Psychologica*, 80, 213–227.
- Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance*, 23, 86–112.
- Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76, 31–48.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281–299.
- van Gog, T., Paas, F., Merriënboer, J. J. G. van, & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, 11, 237–244.
- van Raaij, F. W. (1977). Consumer information processing for different information structures and formats. *Advances in Consumer Research*, 4, 176–184.
- Wedel, M., & Pieters, R. (2007). A review of eye-tracking research in marketing. In N. Malhotra (Ed.), *Review of marketing research, Volume 4* (pp. 123–146). New York: M. E. Sharpe Inc.
- White, C. M., & Hoffrage, U. (in press). Testing the tyranny of too much choice against the allure of more choice. *Psychology and Marketing*.
- Willemsen, M. C., & Johnson, E. J. (2006). *Mouselabweb: Monitoring information acquisition processes on the web*. Retrieved October 14, 2008, from <http://www.mouselabweg.org>.