Université de Lausanne

Faculté des sciences sociales et politiques

On Consumer Decision Strategies:

New Approaches for Studying and Aiding Preferential Choices

Thèse présentée à la Faculté des sciences sociales et politiques

de l'Université de Lausanne pour obtenir le grade de docteur en psychologie

par

Nils Reisen

Lausanne

2009

UNIL | Université de Lausanne
Faculté des sciences
sociales et politiques

IMPRIMATUR

Le Conseil de la Faculté des sciences sociales et politiques de l'Université de Lausanne, sur proposition d'un jury formé des professeurs

- Fabrizio Butera, professeur à l'Université de Lausanne, co-directeur de thèse
- Fred Mast, professeur à l'Université de Berne, co-directeur de thèse
- Ulrich Hoffrage, professeur à l'Université de Lausanne, co-directeur de thèse
- Jérôme Rossier, professeur à l'Université de Lausanne
- Michael Herzog, professeur à l'Ecole Polytechnique Fédérale de Lausanne

autorise, sans se prononcer sur les opinions du candidat, l'impression de la thèse de Monsieur Nils Reisen, intitulée :

« On Consumer Decision Strategies : New Approaches for Studying and Aiding Preferential Choices» .

Lausanne, le 5 mai 2009

Le Doyen de la Faculté

Professeur
Alain Clémence

In psychology, invoking "strategies" to explain funny data is the last refuge of the clueless.

Steven Pinker

Acknowledgements

First of all, I want to thank my two main supervisors Ulrich Hoffrage and Fred Mast for their support and guidance. Special thanks go to Ulrich Hoffrage for his constant and always immediate assistance. His door is always open, both literally and figuratively. I extend my appreciation to my three colleagues Chris White, Jan K. Woike, and Sebastian Hafenbrädl. They were always open for questions and we had many enriching discussions. I also want to thank all the other colleagues who provided help with the planning, execution, and analysis of my experiments.

Moreover, I would like to thank the following people for their help with the experiments. For Experiment 1, my thanks go to Giovanni Rivera Diaz, Ada Lezama Lugo, Huseyin Cumhur Tekin, and Eren Vardarli for their help with preparing the material, writing the program, and collecting the data, and Chris M. White for help with the simulations. For Experiment 2, I thank Lucas Sinclair for programming MouseWorkshop, Dario Bombari for his help with the eye tracking equipment, Gregory Affolter, Richard Ciapala, Julien Finci, Gabriella Sinicco, and Vasko Vitanov for their help with the data collection and Felix Reisen for his help with the programming of the data analyses. For Experiment 3, I thank Joanes Grandjean for his good work as an experimenter and his helpful comments, and, last but not least, all the participants of my experiments for their cooperation.

Furthermore, I am indebted to Chris M. White, Jan K. Woike, John Antonakis, Christian Zehnder, Marika Angerfelt, Jon Baron and four anonymous reviewers for helpful comments on several earlier versions of this dissertation as well as on a manuscript, which is the basis of Part I. Likewise, my thanks go to the members of the dissertation committee for their helpful input.

I also want to thank my partner, Mónica E. Castañeda Tinoco, my family and all my friends for their constant support during these almost four years.

Finally, I am grateful for financial support provided by the Schweizerischer Nationalfonds (Grant numbers 105511–111621/1 and 100011–116111/1).

Abstract

This dissertation focuses on the strategies consumers use when making purchase decisions. It is organized in two main parts, one centering on descriptive and the other on applied decision making research.

In the first part, a new process tracing tool called *InterActive Process Tracing* (IAPT) is presented, which I developed to investigate the nature of consumers' decision strategies. This tool is a combination of several process tracing techniques, namely Active Information Search, Mouselab, and retrospective verbal protocol. To validate IAPT, two experiments on mobile phone purchase decisions were conducted where participants first repeatedly chose a mobile phone and then were asked to formalize their decision 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, Mouselab and eye tracking were directly compared with respect to their impact on information search and strategy description. Only minor differences were found between these two methods. I 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.

In the second part, a prototype of a decision aid is introduced that was developed building in particular on the knowledge about consumers' decision strategies gained in Part I. This decision aid, which is called the *InterActive Choice Aid* (IACA), systematically assists consumers in their purchase decisions. To evaluate the prototype regarding its perceived utility, an experiment was conducted where IACA was compared to two other prototypes that were based on real-world consumer decision aids. All three prototypes differed in the number and type of tools they provided to facilitate the process of choosing, ranging from low (Amazon) to medium (Sunrise/dpreview) to high functionality (IACA). Overall, participants slightly preferred the prototype of medium functionality and this prototype was also rated best on the dimensions of understandability and ease of use. IACA was rated best regarding the two dimensions of ease of elimination and ease of comparison of alternatives. Moreover, participants choices were more in line with the normatively oriented weighted additive strategy when they used IACA than when they used the medium functionality prototype. The low functionality prototype was the least preferred overall. It is concluded that consumers can and will benefit from highly functional decision aids like IACA, but only when these systems are easy to understand and to use.

Keywords: Decision strategies, process tracing, Mouselab, eye tracking, preferential choice, consumer decision making, decision aids, online marketing

Contents

List of Tables

List of Figures

Chapter 1

General introduction

Decisions are made throughout our entire lives and in virtually every moment. The most frequent among them are partly or entirely automated and do not involve much cognitive effort, for example when changing gears while driving a car or whether or not to carry an umbrella when leaving the house. More complex are decisions where the consequences affect the future of a single person or a group of persons in a more serious way (e.g., career decisions, managerial decisions). These decisions often involve some time and effort to be solved and can result in great personal or monetary costs if they do not lead to the desired outcome. In the present dissertation, I study consumer purchase decisions, which fall somewhere between these two extremes. In most cases, they are the result of conscious deliberation, and bad decisions can entail negative—albeit not dramatic—consequences. In this first chapter, I give a brief introduction into the field of decision making to provide the theoretical basis for the research presented here, followed by an overview of my work and the contributions made to the literature on decision making.

A decision is a means of realizing a goal, which can often be achieved in several ways. For instance, a person with the goal of communicating on the road may face the decision of buying either a basic mobile phone or a smart phone that also allows for sending e-mails. Independent of the goal, the decision maker has to choose between at least two *alternatives*. In the simplest form, the decision is between doing something or doing nothing. The non-action is usually referred to as remaining at the *status quo* or *deferring choice* (e.g., Iyengar & Lepper, 2000; White & Hoffrage, 2009; White, Reisen, & Hoffrage, 2009). Alternatives must be mutually exclusive and can be *objects* like different restaurant for dinner, *actions* like reading or going to the cinema, or *long-term strategies* like dieting or saving money.

The different characteristics that describe an alternative are called *attributes*. For instance, if one makes a choice between two different mobile phones, a basic phone and a smart phone, the attributes are the features of each of the phones. Whereas the basic phone is cheap, small and well suited for basic communication (i.e., calling and text messaging), the smart phone offers many additional functions such as mobile internet and e-mail. Which of the attributes is of greater importance to the decision maker depends on his or her subjective preferences. Consequently, different people may make different decisions when faced with the same choice.

Decisions can be made by an *individual* or by a *group of persons* and are either *single* decisions or *scenarios*. In a single decision, the decision maker has to decide only once. This can be a student's decision whether or not to enroll in a certain class. A scenario, in contrast, is a series of decisions in which the outcome of a first decision affects further decisions. These can be decisions of an emergency physician who has to decide whether to operate immediately or to make some laboratory tests beforehand. If she decides to make the tests, the obtained results will act as a new basis for the second decision and so on. A similar distinction exists between whether a decision is made only once or repeatedly. The decision of whether to become a physician is usually taken only once in a life-time. In contrast, the choice of a mobile phone is made several times during one's life, at least for many people.

When the person who makes the decision knows all the possible alternatives right from the beginning, we speak of a *given set of alternatives* or *options*. In the case of an *open set of options*, on the other hand, the decision maker first has to find out what the alternatives are. For instance, if you do not want to spend your evening at home, you will have to make up your mind what your options for that night are. Experiments in decision making have been concerned primarily with single decisions with a given set of options (e.g., gambles).

Moreover, decisions are characterized by the number of attributes the alternatives have. *Onedimensional* decisions are decisions in situations where each alternative has only one attribute. *Multidimensional* or *multiattribute* decisions, in contrast, involve alternatives that are described on two or more attributes. Typical for the latter is a conflict that arises when one alternative is better on attribute A and the other is better on attribute B. This requires making tradeoffs between the attributes, which is cognitively more demanding and often experienced as painful (e.g., Weber, Baron, & Loomes, 2001).

As mentioned above, decisions can differ considerably with regard to the *cognitive costs*. *Routine* decisions have the lowest cognitive costs, which are decisions where the decision maker simply matches the current decision onto decisions made in the past and, if a similar decision is found, chooses as before. *Reflected* decisions, in contrast, involve high cognitive costs. Here, the decision maker first has to identify possible alternatives, search for information on these alternatives and then integrate this information to finally make a decision. It is this latter type of decision that is of special interest for the field of decision making in general and for the present work in particular.

A final but important aspect is the distinction between *judgement* and *choice*. A choice is a selection of an alternative out of a set of two or more alternatives, whereas a judgement demands an allocation of values to each alternative, for example on a point scale or by assigning a price one would be willing to pay (Jungermann, Pfister, & Fischer, 1998; Huber, 1982). In this dissertation, the focus is on decisions as choices. I therefore now describe some theories of choice.

1.1 Theories of choice

The study of people's decision behavior can be divided into three broad lines of theories, *normative*, *descriptive* and *prescriptive*, which are usually summed up under the term *Behavioral Decision Theory* (BDT). The normative theory of decision making is "the theory of how we should choose among possible actions under ideal conditions" (Baron, 2000, p. 223). It is based on the assumption that the decision maker is rational and follows the principle of maximization (i.e., achieving the best possible outcome). However, people often do not act in the way prescribed by normative theory. In fact, in many cases only partly rational reasoning is observed. This is due to the fact that the cognitive capacity of the decision maker is not sufficient for the complex calculations required by normative theory, or—in some situations—it is simply not used (Jungermann et al., 1998). The first author who took this fact into consideration was Herbert Simon (1955; 1956) with his concept of *bounded rationality*. He argues that models of human judgment and decision making should be built on what we actually know about the mind's capacities rather than on fictitious competencies (Simon, 1987). Specifically, "[b]ecause of limits on their computing speeds and power, intelligent systems [human beings, computers] must use approximate methods to handle most tasks. Their rationality is bounded." (Simon, 1990, p. 6). Moreover, the optimal strategy is often unknown or unknowable and therefore the normative model fails in such real-world situations. He introduced the concept of *bounded rationality* and proposed to replace *maximizing* (i.e., choosing the option with the highest utility) with *satisficing* (i.e., choosing the first option that is good enough). Originating from this major objection, modern decision theory turned away from the normative model toward a more realistic description of human decision behavior. This line of research, descriptive decision theory, is the study of how people who are not trained in normative theory make decisions in real-life situations, thereby taking into account the natural limitations of the human mind. In particular, the work of the two psychologists Daniel Kahneman and Amos Tversky documented people's violation of the rules of normative theory. They showed that people often rely on simple *heuristics* rather than on complex calculations when making decisions (Kahneman, Slovic, & Tversky, 1982; Tversky & Kahneman, 1981). Many authors contributed to this line of research, which conceptualizes the human being as an information processing system that has to cope with many different environments (e.g., Gigerenzer, Todd, & the ABC Research Group, 1999; Hogarth & Karelaia, 2007; Payne, Bettman, & Johnson, 1993). Both the normative and the descriptive theory are of importance for each other. On the one hand, normative research provides theories and models of optimal decision behavior that can be used for the development of decision aids. Observing people's actual decision behavior, on the other hand, gives insights about people's deviations from normative theory and can ultimately help to fit the decision aids to the needs and habits of the decision makers.

The third class of decision theories, prescriptive decision theory or *decision analysis*, originated from normative and descriptive theory and provides guidelines for how a real person should act (Kühberger, 1994). Decision analysis provides techniques for structuring the decision problem to achieve the best possible outcome and is meant to help people who are faced with difficult decisions like managers and physicians (for a detailed overview see Eisenführ & Weber, 1994). To illustrate, have a look at a famous example: Charles Darwin facing the question of whether or not to marry (Darwin, 1958). To structure the problem, he identified the pros and cons for each of the two alternatives and wrote them on a piece of paper with the heading "This is the Question:"

MARRY

Children – (if it please God) – constant companion, (friend in old age) who will feel interested in one, object to be beloved and played with – better than a dog anyhow – Home, and someone to take care of house – Charms of music and female chitchat. These things good for one's health. Forced to visit and receive relations *but terrible loss of time.* My God, it is intolerable to think of spending one's whole life, like a neuter bee, working, working and nothing after all. – No, no won't do. – Imagine living all one's day solitarily in smoky dirty London House. – Only picture to yourself a nice soft wife on a sofa with good fire, and books and music perhaps – compare this vision with the dingy reality of Grt Marlboro' St.

Not MARRY

No children, (no second life) no one to care for one in old age — What is the use of working without sympathy from near and dear friends — who are near and dear friends to the old except relatives. Freedom to go where one liked – Choice of Society *and little of it.* Conversation of clever men at clubs. – Not forced to visit relatives, and to bend in every trifle – to have the expense and anxiety of children – perhaps quarrelling. *Loss of time* – cannot read in the evenings – fatness and idleness – anxiety and responsibility – less money for books etc. – if many children forced to gain one's bread. – (But then it is very bad for one's health to work too much)
Perhaps my wife won't like London; then the sentence is banishment and degradation with indolent idle fool –

From this, Darwin concluded that he should marry. Many years later, this method of writing down the pros and cons of the alternatives in separate columns was adopted by Janis and Mann (1977) who developed a decision aid called *balance sheet*.

1.1.1 The normative theory of choice

The main assumption of normative models is the *homo economicus.* He is characterized by three issues: (i) he is disposed of the complete information about all alternatives, (ii) he is unboundedly sensible regarding the differences between the alternatives, and (iii) he is rational, that is, he is able to order his preferences and actions and his choice follows the principle of maximization. Given the above premises he maximizes in any form his personal utility (Kühberger, 1994). The most famous example of such a theory is *Subjectively Expected Utility theory* (SEU) by John von Neumann and Oskar Morgenstern (1947), which is described in the following.

Subjectively expected utility theory

Most of the research in BDT is done using gambles. In such a decision the determining variables are (i) the utilities of the consequences and (ii) the probabilities of obtaining these consequences (Huber & Huber, 2003). Depending on whether probabilities and consequences are objective or subjective, four types of theories can be distinguished, which are shown in Table 1.1.

Table 1.1: Utility theories

		Values of the Consequences	
		objective	subjective
Probabilities	objective	expected value	expected utility
	subjective	subj. expected value	subj. expected utility

(adapted from Jungermann et al., 1998, p. 199)

SEU theory is fundamental for BDT.[1] Here, the decision maker tries to maximize the value for him- or herself, hence his or her subjective utility. The (subjectively) expected value can be computed mathematically by multiplying the probability of winning by the monetary value of payoff. Imagine an unbiased coin tossed only once. When it lands on tails, the player is paid $2, otherwise the player gets nothing. The probability of winning is exactly .5, so the expected value of this gamble is .5*$2 or $1. If the game were played several times, the player would, on average, win $1 per play. The formula for calculating the expected outcome is as

[1]Originally from normative theory, SEU theory was also regarded as *the* descriptive theory of human decision behavior. Because of its foundation in normative theory, SEU theory is described in this chapter. Further descriptive theories are presented in Section 1.1.2 (p. 6).

follows:

$$(S)EV = \sum_i p_i * v_i \tag{1.1}$$

where EV is the expected value, i are all the different outcomes, p_i is the probability of the ith outcome and v_i is the value of the ith outcome. Hence, $p_i * v_i$ is the expected outcome of one trial and $EV = \sum_i p_i * v_i$ is the expected outcome over all trials (Baron, 2000).

Analogously, the (subjectively) expected utility is obtained by replacing values (v_i) with utilities (u_i):

$$(S)EU = \sum_i p_i * u_i \tag{1.2}$$

To be able to compare the mathematical predictions of expected utility theory with the behavior of real decision makers, von Neumann and Morgenstern (1947) specified a set of axioms that underlie rational decision making. The following four axioms are central:

Comparability/Weak ordering. A decision maker can compare the alternatives and prefers either option X to option Y, or option Y to option X, or is indifferent ($X \sim Y$).

Transitivity. A rational decision maker preferring option X to option Y and option Y to option Z should also prefer option X to option Z.

Cancellation. A choice between two alternatives should be based only on those outcomes that differ, not on outcomes that are the same for both alternatives.

Continuity. When a decision maker prefers option X to Y and Y to Z, then there is always a probability p such that $Y \sim [X, p; Z, 1-p]$.

When at least one of these principles is violated, expected utility is not necessarily maximized. Famous violations are the *Allais Paradox* (Allais, 1953), *Ellsberg's Paradox* (Ellsberg, 1961), and reversals of preference due to changes in the method of preference elicitation (see Seidl, 2002, for an overview). These violations show that in many situations expected utility theory does not adequately describe how people make decisions. Thus, it seems that both paradigm (i.e., the gamble) and subject (i.e., homo economicus) are too narrow and that the EU model can describe human behavior only in exceptional cases (Kühberger, 1994).

1.1.2 Descriptive theories of choice

SEU theory is not only a normative theory, it was also the first descriptive theory of decision making. However, with a steadily growing number of findings that clearly demonstrated people's deviations from SEU principles the research focus shifted toward new theories that could explain and predict these contradictory findings. Here, two general types of models are used. *Isomorphic* models aim to describe process, input, and outcome of cognitive information processing, whereas *paramorphic* models try to relate input (i.e., information about the decision alternatives) to output (i.e., the final decision), without claiming to provide a valid representation of the underlying process (Ramaprasad, 1987). In this dissertation, the focus is on isomorphic models, that is, the goal is to obtain valid and accurate descriptions of human decision strategies. Several techniques exist that allow for the construction of such process models. These techniques, called process tracing techniques, are designed to uncover the cognitive processes that are involved in decision making, such as the acquisition and the integration of information about the choice alternatives. In this context, cognitive processes are understood as the decision maker's processing of the information upon which the decision is based. In other words:

> Cognitive process research focuses on specific influences on a person's cognitive information processing. Its focus is micro, on the elements of a person's cognitive information processing. These are, for example: perceiving and recognizing stimuli, remembering and searching for information, inducing rules, recognizing patterns, formulating concepts, and applying all of these in sensing, formulating, and solving problems (Ramaprasad, 1987, p. 140).

Paramorphic models, in contrast, can be obtained with structural modeling techniques. They are not of direct importance for the present dissertation and are therefore only briefly mentioned in Chapter 2 (p 15). I now describe some descriptive approaches.

Prospect theory

The most important revision of SEU is *prospect theory* or its newer form *cumulative prospect theory* (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). It accounts for almost all of the available data concerning decisions under risk and has inspired many other similar attempts. Prospect theory is descriptive in that it explains how and why our choices deviate from the normative model of SEU theory. It has two main parts, one concerning probability and the other utility. In the first phase called *editing*, the decision maker encodes and transforms the given problem and creates a mental representation according to certain rules. Then, in the *evaluation* phase, a subjective value is assigned to each of the edited options and finally one of the options is selected. The basic assumption remains that choices are made by multiplying a subjective probability by a utility. What is new here is that humans distort probabilities and think about utilities as changes from a reference point, which is usually the individual's *status quo*. This reference point can easily be affected by other factors, such as how a problem is presented. A famous effect that depends on the presentation of the problem and that can be explained by PT is the so called *framing effect*, which refers to the phenomenon that already slight differences in the way the problem is described can lead to contrary choice (e.g., Asian disease, Tversky & Kahneman, 1981).

Prospect theory is the most widely accepted alternative to SEU theory. It provides a more accurate description of how people actually make decisions and can also be applied rather easily to a range of common situations. Furthermore, it is capable of explicitly predicting many violations of SEU theory (cf. Plous, 1993). Compared to SEU, prospect theory is better in lotteries with extreme probabilities, that is, probabilities that are close to 0 or 1, and is suitable especially when one is interested in a good trade-off between frugality and explanative power. However, being an "as-if" model, prospect theory remains silent about the processes underlying decision making and still falls short in explaining some empirical findings (e.g., intransitivities, cf. Jungermann et al., 1998). A more detailed description of PT would be beyond the scope of this dissertation. For an overview, see Baron (2000) or Jungermann et al. (1998).

Heuristics and other decision strategies

SEU and prospect theory represent the human mind as a machine computing probabilities and utilities. The problem with this assumption is that the more complex a decision gets, the less likely it is that a human decision maker is able or willing to calculate the expected values for all possible alternatives. A new line of descriptive theory that takes the limitations of the human mind explicitly into account is the approach of *decision heuristics*.

Heuristics are decision strategies that are very simple to execute and that often use only a fraction of the available information, but which—in spite of their simplicity—often yield very close approximations to the "optimal" answers suggested by normative theories (Gigerenzer et al., 1999; Payne et al., 1993). However, these "rules of thumb" or "shortcuts" can also result in systematic biases and inconsistencies (i.e., deviations from normative theory; Kahneman et al., 1982; Tversky & Kahneman, 1981). Whereas research in the tradition of Kahneman and Tversky focuses more on the biases that result from the use of heuristics, the research by Gerd Gigerenzer and the ABC Research Group demonstrates that heuristics can lead to very good decisions, even when only a fraction of the available information is used (Gigerenzer et al., 1999; Gigerenzer, 2007, Gigerenzer, Hoffrage, & Goldstein, 2008).

A famous example for the latter approach is the *take-the-best heuristic* proposed by Gigerenzer and Goldstein (1999). This heuristic first looks up the values on the most valid cue, where cue validity is defined as the relative proportion of correct inferences among all inferences in which this cue discriminated between the alternatives. If the two alternatives differ on this cue the information search is stopped and the alternative with the higher cue value is selected. If they do not differ, take-the-best proceeds with the second most valid cue, and so on. Take-the-best is a *lexicographic* heuristic: it orders the cues or attributes according to their validity or importance (similar to the ordering of words in a dictionary). Originally formulated for inferential choices (e.g., which of two cities has more inhabitants), take-the-best has also been generalized to preferential tasks (e.g., which company to buy stocks from; see Section 3.1.1 on page 34 for details). A heuristic for preferential choices that follows a similar logic is the *eliminations-by-aspects heuristic* (Tversky, 1972). It first determines the most important attribute and then compares the values of all alternatives on this attribute to a cutoff value (e.g., maximum acceptable price). All alternatives that fail to reach the cutoff on this attribute are eliminated. This is continued with the second most important attribute and so on until only one alternative remains.

The *weighted additive strategy*, in contrast, is a decision strategy that demands substantial computational processing of the information. In contrast to a heuristic, it considers the values of each alternative on all the relevant attributes and all the relative importances or weights of the attributes to the decision maker. The alternative with the highest overall value is chosen. This decision strategy is also known as the *Multi-Attribute-Utility Theory* (MAUT; Keeney & Raiffa, 1976), which is one of the best-known decision models in the literature (see Section 5 on page 67 for more details).

Many more heuristics and decision strategies are described in the literature, which vary in the extent to which they make tradeoffs among attributes. A strategy is *compensatory* when good values on one attribute can offset bad values on another. If no such compensation is possible then the strategy is *noncompensatory*. The weighted additive strategy is compensatory and take-the-best and eliminations-by-aspects are noncompensatory. Furthermore, take-the-best and elimination-by-aspects are examples for one-reason decision making. That is, in some cases the decision is based solely on one reason or cue. The meta-decision of which decision strategy to use depends on the problem characteristics such as the importance of the decision and the number of alternatives and attributes. The more important the problem, the more likely it is that the decision maker uses a compensatory rule. Noncompensatory strategies, in contrast, are especially used in decisions with many alternatives and attributes, in situations with time pressure, or when the decision maker simply does not want to devote a lot of work into the decision (Jungermann et al., 1998). Research suggests that, especially in more complex decision situations, decision makers use not only one but several different heuristics or strategies in the course of the choice process. For instance, a person facing a decision with a hundred possible alternatives will first want to reduce the set of options to a manageable amount. This can be done easily with a noncompensatory heuristic, such as elimination-by-aspects. When only three or four alternatives remain, she may want to proceed in a more exhaustive manner and chooses in a second step the compensatory and more effortful weighted additive strategy. This particular combination of decision strategies has been observed frequently (Ball, 1997; Bettman, 1979; Billings & Marcus, 1983; Ford, Schmitt, Schlechtman, Hults, & Doherty, 1989; Gensch, 1987; Olshavsky, 1979; Payne, 1976;

Westenberg & Koele, 1994; Wright & Barbour, 1977; see, however, Glöckner & Betsch, 2008). Further combinations of different rules have been proposed by H. Montgomery and Svenson (1976), Svenson (1979) and Payne et al. (1993).

1.2 Overview and contributions of the dissertation

Decision strategies, or, in other words, the cognitive processes that are at work when humans make decisions are the main topic of the present dissertation. Specifically, I focus on two major questions. The first is of a purely descriptive nature: How can the decision strategies people use when making preferential choices be investigated and described? The second, in turn, has an applied character: How can our knowledge about these decision strategies be used to support human decision making? The research described here is thus organized into two main parts, one devoted to each of these two questions.

In the first part, I introduce a new process tracing tool called *InterActive Process Tracing* (IAPT), which I developed together with my co-workers at the University of Lausanne. This tool is the result of a combination of several process tracing techniques that have been widely used for the descriptive study of decision making, namely, Active Information Search, Mouselab, and retrospective verbal protocol. In Chapter 2, I give a general introduction into process tracing techniques followed by the description of the new method. To evaluate IAPT, two experiments were conducted, which are described in Chapters 3 and 4, respectively. Because IAPT was complemented with a further process tracing technique (namely, eye tracking) in Experiment 2, a brief introduction into the topic of the recording of eye movements is given at the beginning of Chapter 4. Note that a reduced version of Part I is published in the Journal of Judgment and Decision Making (Reisen, Hoffrage, & Mast, 2008).

Part II is concerned with the question of how the insights from past research in general and my own in particular can be used to develop a decision aid, which systematically assists consumers in their purchase decisions. In Chapter 5, I give an overview of current (online) decision aids, followed by the description of the prototype of a decision aid called *InterActive Choice Aid* (IACA). The intention behind developing the IACA was to create a decision aid that assists the natural process of human decision making by providing tools that facilitate the execution of resource intensive tasks. To evaluate this prototype regarding its perceived utility and general acceptance, an experiment was conducted where IACA was compared to two other prototypes of real-world consumer decision aids. This third experiment is described in Chapter 6. In the final chapter, conclusions are drawn about the general insights of the presented research as well as its contributions to the literature.

1.3 Contributions

This dissertation makes the following contributions to the literature:

- A new methodology for the descriptive research of cognitive decision strategies is developed, extended, and evaluated (i.e., InterActive Process Tracing).

- Retrospective verbal protocols are validated as a measure of cognitive processes: decisions predicted on the bases of these verbal protocols are compared to observed choices.

- The validity of information search measures for the detection of cognitive processes is investigated by contrasting information search data and verbal protocol data.

- Two widely used process tracing techniques (i.e., Mouselab and eye tracking) are directly compared in terms of their convergent validity.

- A new tool for aiding consumer decisions is developed and evaluated, which is directly based on insights gained from the descriptive study of decision making (i.e., InterActive Choice Aid).

Part I

Identifying Decision Strategies

Chapter 2

Introduction to Part I

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; Dawes, 1979; Einhorn, Kleinmuntz, & Kleinmuntz, 1979; Ford et al., 1989; Harte & Koele, 1995; Payne, 1976; Riedl, Brandstätter, & Roithmayr, 2008; Svenson, 1979; Westenberg & Koele, 1994). Structural modeling aims to uncover psychological processes by relating the provided information to the decisions or judgments, typically via multiple linear regression analysis. Inferences are drawn from the resulting model about the participants' decision strategy. Despite its popularity, this approach has been criticized for ignoring the predecisional phase, that is, the processes that take place between stimulus presentation and final decision. For example, Svenson (1979) came to the conclusion 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 predecisional 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 et al., 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; Williamson, Ranyard, & Cuthbert, 2000).[1]

[1] For the sake of completeness, further process tracing techniques are the Information Search Laboratory (ISLab) by Cook and Swain (1993) and Search Monitor by Brucks (1988).

In the following, I briefly describe these process tracing methodologies and discuss their strengths and weaknesses. I then present a new tool called *InterActive Process Tracing* (IAPT), which was developed to identify the decision processes underlying preferential choice. IAPT uses various elements of the process tracing measures cited above to combine their strengths and simultaneously overcome some of their weaknesses. I subsequently describe two experiments in which IAPT was successfully applied to identify participants' decision strategies. Finally, Part I concludes with a discussion of the findings and an outline of avenues for future research.

2.1 Process tracing techniques

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

A range of techniques have been developed within the process tracing paradigm, each of them having both strengths and weaknesses (see Table 2.1). 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 alternatives 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).

In the real world, however, decision problems very rarely come as pre-structured as in the typical information-board experiments. Outside the laboratory, people not only have to construct their choice strategy on-line, but they also have to structure the problem on-line (Lopes, 1990). Furthermore, a pre-structured arrangement may bias participants' information processing because they are told indirectly what information is considered important for the task (Brucks, 1988; Huber et al., 1997). To overcome this disadvantage, AIS is a promising next step in the development of process tracing techniques. Similar to Mouselab and eye tracking, this method 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., 1997). Specifically, the participants receive a minimal description of the decision situation and have to query the experimenter for any further information.

Table 2.1: Strengths and weaknesses of four process tracing techniques

Strenghts	**Weaknesses**
Mouselab	
+ Convenient to use. + A large amount of data: which and how much information is retrieved and the sequence of the information acquisition.	– 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.
Eye Tracking	
+ 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.	– 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.
Active Information Search (AIS)	
+ Enhanced realism: participants are less affected by the experimental setup.	– Less exact monitoring of the information acquisition process than with Mouselab. – Only data concerning the search for information, but no data concerning information integration.
Retrospective Verbal Protocol	
+ Rich and detailed information: information search *and* integration. + No interference with on-line decision making.	– Doubts that people can introspectively access their cognitive processes. – Reactivity: forgetting and fabrication. – Very time-consuming analysis.

The most robust finding of the studies within the information search paradigm is that there is a general tendency to use noncompensatory strategies in judgment and choice. This is the case when the task is getting more complex (i.e., an increase in the amount of information to process) or when time pressure is induced. In particular, the depth of search is reduced and the search becomes more variable when the information load increases. Moreover, participants tend to use strategies that consist of two steps. First, they quickly eliminate unattractive alternatives with a noncompensatory algorithm, where unfavorable values on some attributes cannot be compensated for by favorable values on other attributes. In a second step, the remaining options are compared in detail with the help of a compensatory algorithm (Ball, 1997; Bettman, 1979; Billings & Marcus, 1983; Ford et al., 1989; Gensch, 1987; Olshavsky, 1979; Payne, 1976); Westenberg & Koele, 1994; Wright & Barbour, 1977; see, however, Glöckner & Betsch, 2008).

A major weakness of the information search techniques is, however, 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).

2.1.2 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. Given that the particularities of the design of IAPT excluded the use of concurrent protocols, in the following I focus mainly on retrospective protocols. In a classic paper, Nisbett and Wilson (1977) questioned the assumption that people have introspective access to their cognitive processes. They concluded that people's ability to observe and report upon higher order mental operations is often small or even not existent:

> "People often cannot report accurately on the effects of particular stimuli on higher order, inference-based responses. Indeed, sometimes they cannot report on the existence of critical stimuli, sometimes cannot report on the existence of their responses, and sometimes cannot even report that an inferential process of any kind has occurred (p. 233)."

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 (e.g., due to an increased load on working memory), 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 my 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. This procedure has been shown to increase the completeness of the verbal protocol (see van Gog, Paas, van Merriënboer, & Witte, 2005, for an overview).[2] 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.

2.2 The method of InterActive Process Tracing (IAPT)

Given that each of the four process tracing techniques described above has weaknesses and limitations, my co-workers and I 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). In particular the fact that it is difficult to detect different decision strategies within one choice process with process or outcome measures alone calls for the use of a complimentary method, such as verbal protocols (Riedl et al., 2008). For instance, Payne (1976) used a combination of an information board and a concurrent verbal protocol in an apartment choice task. He found that the more complex the task became, the more partici-

[2]Interestingly, Russo et al. (1989) were one of the first to use this method but without the positive effects found in other studies.

pants shifted from compensatory to noncompensatory processing and eliminated alternatives early in the process to winnow down the number of options. Biggs, Bedard, Gaber, and Linsmeier (1985) used a similar design for studying decision behavior of bank loan officers (i.e., experts). Here, too, increased task size led to more noncompensatory strategies, whereas increased similarity of alternatives led to more compensatory strategies. Participants in both studies demonstrated adaptive behavior that is contingent on the characteristics of the task (cf. Payne et al., 1993).

A major feature of the new method described here is that an attempt is made to detect the cognitive processes interactively with the participant, which is why it is called *InterActive* Process Tracing. In the experiments, participants first selected the attributes they considered important (AIS) and then they made a series of choices (Mouselab in Experiments 1 and 2, eye tracking in Experiment 2), which was followed by a retrospective protocol. Here, the participants were asked to give a detailed description of how they made the decision. Note, however, that the last phase of IAPT 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.[3] I am aware that these changes in the procedure might reduce the validity of the verbal protocols. However, the described strategies can be used to retrospectively predict[4] 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.

2.2.1 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

[3]Consequently, no tape-recording during the collection of the verbal protocol and no subsequent coding was necessary because participants provided a "ready to use" process model that could be used directly to retrospectively predict their choices.

[4]Note that because the strategies were calculated only after the participants' choices, the correct term in this context would be postdiction rather than prediction. However, given that prediction is the more standard terminology it will be used throughout this dissertation.

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 experiments described in the following two chapters, IAPT was used to address the question of whether people are indeed able to gain introspective access to their cognitive processes, and ultimately, to what extent those verbal protocol data are instrumental in constructing process models that can accurately predict the observed choices. Moreover, in Experiment 2, IAPT was developed further by integrating yet another process tracing measure, namely eye tracking, and some attention was devoted to the phenomenon of choice deferral.

Chapter 3

Experiment 1: Validation of the method of InterActive Process Tracing

In Experiment 1, IAPT was put to a first test. The central question was whether people are able to accurately describe their cognitive processes, and whether IAPT is a good method for identifying these processes. The hypotheses are as follows.

3.1 Hypotheses

3.1.1 Prediction accuracy

Building on the findings by Bettman (1970), Einhorn et al. (1979), Larcker and Lessig (1983), and Li et al. (2000), it is reasonable to assume that people have a fairly good idea of (and are able to verbalize) the decision strategy they use to select a product. If asked to provide an executable decision strategy, they should thus be able to come up with an algorithm that is able to retrospectively predict their own choices in a large number of the cases.

Hypothesis 1: *The strategies described by the participants will predict their own choices better than a random strategy (i.e., chance).*

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 one phone dominated the others on a given trial. Thus, a high number of correct predictions does not necessarily imply that participants were able to accurately describe their strategies. Therefore, another benchmark against which the number of correct predictions could be compared is the percentage of correct predictions that results from using a certain participant's strategy to predict the choices of all other participants. Assuming that each participant's strategy is unique, it is expected that a particular participants' strategy does not predict other participants' choices well.

HYPOTHESIS 2: *A particular participants' strategy will predict his or her own choices much better than the choices made by other participants.*

As a third benchmark, besides chance and cross-participants comparisons, is the fit obtained when modeling the observed choices with several variants of two established strategies from the literature. These strategies have been shown to lead to good fits (Dawes, 1979; Gigerenzer & Goldstein, 1999). However, given the uniqueness of the strategies described by the participants, it is expected that these still have a higher predictive accuracy than the different versions of WADD and TTB.

HYPOTHESIS 3: *The described strategies will predict the observed choices better than the different variants of the Weighted ADDitive strategy and the take-the-best heuristic.*

3.1.2 Information search

Process tracing studies are based on the assumption that the information search data reflects how the information about the decision alternatives is processed and used to form an evaluation. From these data, insight into various particularities of the evaluation process can be gained, because different decision strategies often require different types of information and acquire them in a different order. Of special interest here is whether the strategy is compensatory or noncompensatory and which information is used or not used (Harte & Koele, 2001). However, a concern with this type of data is that, strictly speaking, it only tells us something about the information search, and hence the conclusions about the underlying cognitive processes are more or less speculative. Given that two different procedures are used (i.e., Mouselab and retrospective verbal protocol) to identify these processes, it is possible to verify whether the way in which participants search for information is in agreement with the strategies they describe. Here, I focus on three main questions. First, are the described strategies reflected in the direction of the participants' search for information? Second, do participants 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 participants access information on the selected attributes reflect the attributes' ranking that they established in the first phase of the experiment?

To examine the direction of the participants' information search, three measures are commonly used: the pattern of information acquisition expressed by the *Payne Index* (PI, Payne, 1976), the variability, and the depth of the information search (see, e.g., Ford et al., 1989; Harte & Koele, 2001; Payne, 1976). The PI 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 an attribute-wise search is indicative of noncompensatory strategies. Formally, it compares the total number of within-alternative transitions to the number of within-attribute transitions (neglecting diagonal transitions, i.e., transitions from one alternative to another one and, *simultaneously*, from one attribute to another):

$$PI = \frac{N_{alternative} - N_{attribute}}{N_{alternative} + N_{attribute}} \tag{3.1}$$

where $N_{alternative}$ is the number of within-alternative transition and $N_{attribute}$ is the number of within-attribute transitions. A score of 1.0 represents a fully alternative-based search whereas a score of –1.0 represents a fully attribute-based search.

The variability of search indicates the degree of selectiveness of the participant's search by focusing on the amount of information examined per alternative. It is measured by the standard deviation of the proportion of cells accessed. A variability of search of zero means that the participant searched for the same amount of information on all alternatives, which is typical for compensatory decision making. The more the standard deviation departs from zero, the more selective is the search. This, in turn, points to a more noncompensatory strategy. In addition, the depth of search is the percentage of the total information actually inspected by the participant. Compensatory strategies such as the weighted additive strategy need all the available information for execution whereas noncompensatory strategies in many cases make a decision based on only a fraction of the information.

Therefore, it is expected that the strategies used by the participants are reflected in the direction of their information search as measured by the Payne Index, the depth of search and the variability of search. Specifically:

HYPOTHESIS 4: *Elimination strategies will lead to an attribute-wise search and additive strategies will lead to an alternative-wise search (as indicated by the Payne Index).*

HYPOTHESIS 5: *Elimination strategies will lead to a higher variability of search than additive strategies.*

HYPOTHESIS 6: *Elimination strategies will lead to a lower depth of search than additive strategies.*

Moreover, it is possible to test whether the elimination of alternatives as described by the participants' strategies is reflected in their information search. Specifically, it is assumed that as soon as an alternative is eliminated because its value on one of the attributes fails to reach the threshold, the participant should not acquire any more information about that alternative.

HYPOTHESIS 7: *Participants will stop searching for information on an alternative as soon as their strategy prescribes its elimination.*

Finally, it can be tested whether there is a relation between an attribute's rank and the frequency with which information about this attribute is accessed. This can be done by assessing whether information about attributes that are reported to be more important is acquired more frequently than information about less important attributes.

HYPOTHESIS 8: *Participants will access more information on higher-ranked attributes than on lower-ranked attributes.*

3.2 Method

3.2.1 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).

3.2.2 Task and stimuli

In each of 30 choice trials, participants had to select 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. In this way, it was ensured that participants based their decisions on the informations they sought rather than on inferences they made about the products based on previous experiences with a certain brand or model. The attributes that were used in this and also the following experiment were technical characteristics of mobile phones that could typically be found on shopping websites (see Tables 3.1 and 4.1).

3.2.3 Design

Participants were each 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. The rationale for this was to leave it entirely up to the participants which information they wanted to consider. 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.

3.2.4 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.

Phase 1: Selection of attributes

Attribute selection and ordering (a feature from AIS) as well as the 30 choice trials of Phase 2 were performed using a computer-based process tracing measure that combines central elements of AIS and Mouselab. Phase 1 was the first interactive part of IAPT.

Participants were asked to select 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.[1]

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.

[1]This did not happen very often. Participants were informed that only technical (i.e., objective) attributes were available and some pre-testing was done to ensure the completeness of the list.

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 3.1), 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.[2] There were no constraints regarding the amount of or the order in which the information was considered. A choice could be made at any time during a given trial and participants could proceed to the next trial only after having selected one of the options. It was not possible to go back to earlier trials.

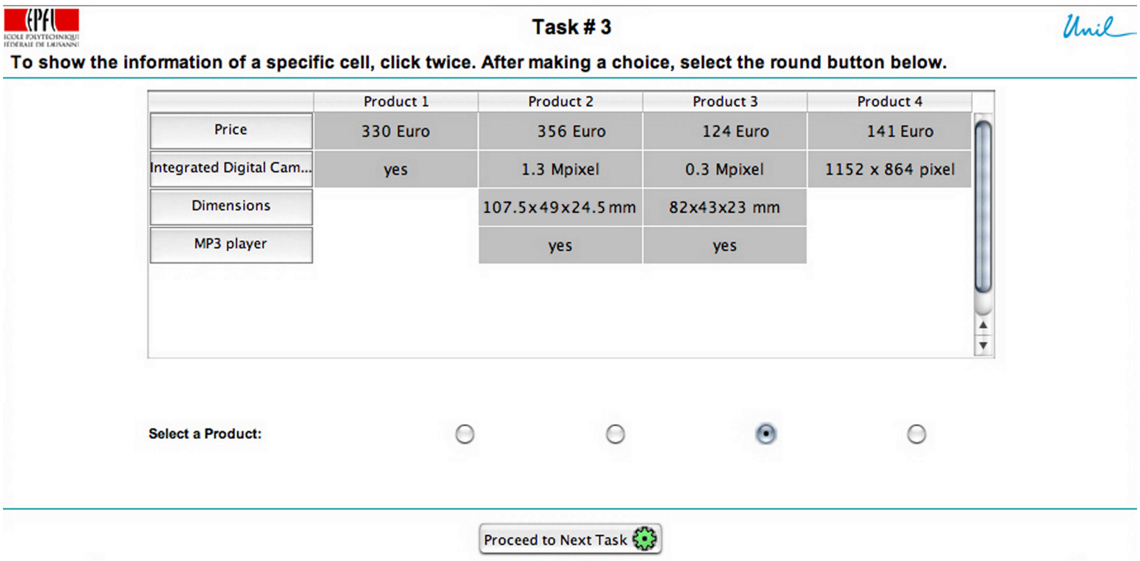

Figure 3.1: Screen-shot of the computer-based process-tracing measure used in Experiment 1

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. That is, participants described their decision strategies in terms of a step-by-step list, similar to a computer program. When an element of their strategy was too imprecise, that is, when a step could not be executed by a computer, they were asked to be more specific. For instance, when

[2] Note that this is different from the standard form of Mouselab, where the cells close as soon as the mouse is moved away. I think that this form is easier to use for participants and, for the current purpose, I found no reason to stick to the standard procedure.

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, screen-shots of the information board of five of the trials were presented. These screen-shots were taken when the participants had made a choice (a procedure known as cued retrospective reporting; van Gog et al., 2005). These cuing trials that were different for each participant were selected 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 their 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*.

3.2.5 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.

3.3 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. Conditions did not differ significantly with respect to the number of attributes participants decided to include in the information board (5.13 and 6.33 of 33, respectively;

$t(19) = 1.49, p = .15$).[3] Overall, they 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%). Twenty-four attributes were requested by between 29% and 3% of the participants and four attributes were never used (i.e., voice command, Office, WAP, and external display). The frequency with which each attribute was chosen is listed in Table 3.1.

[3]Because Levene's test for the equality of variances proved significant, the degrees of freedom were adjusted accordingly.

Table 3.1: Number of participants who chose each attribute

	Choice frequency			
Attribute	Without list	With list	Total	Total %
Price	9	12	21	68
Integrated digital camera	11	6	17	55
Size	9	7	16	52
MP3 Player	9	3	12	39
Bluetooth	3	6	9	29
Stand by	4	5	9	29
Internal display	5	4	9	29
Image	6	3	9	29
Talk time	2	5	7	23
Weight	4	3	7	23
VibraCall alert	0	6	6	19
SMS	1	5	6	19
GPRS	3	3	6	19
Stereo FM	0	5	5	16
MMS	2	2	4	13
Band support	1	3	4	13
Internal memory	2	2	4	13
Java games	2	1	3	10
Windows Mobile	2	1	3	10
External flash memory	2	1	3	10
Color	2	1	3	10
Integrated speakerphone	1	1	2	6
Email support	0	2	2	6
Infrared	0	2	2	6
UMTS (3G) compatibility	1	1	2	6
Connectivity (USB)	1	1	2	6
Video streaming	0	2	2	6
Video clip playback with sound	0	1	1	3
Speech recording	0	1	1	3
Voice command	0	0	0	0
Office	0	0	0	0
WAP	0	0	0	0
External display	0	0	0	0

3.3.1 Described strategies

The strategies were classified according to several dimensions. Specifically, for each strategy it was determined whether a certain element was present or not. The elements were: elimination of alternatives, adding up attribute values, assigning weights to attributes, and *Just-Noticeable-Differences* (see below). In general, two types of strategies could be identified: elimination strategies and additive strategies. Strategies of the first type eliminate alternatives from the consideration set based on attribute values, that is, when a particular attribute value does not reach the acceptance threshold specified by the participant (for an example, see Table 3.2, participant 37). These strategies follow a similar logic as do lexicographic strategies like the take-the-best heuristic (TTB, Gigerenzer & Goldstein, 1996; Gigerenzer, Hoffrage, & Kleinbölting, 1991) or the Elimination-By-Aspects strategy (EBA, Tversky, 1972). The number of attributes that were 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 3.2, participant 5).

Table 3.2: Three examples of participants' strategies. Participant 5 used a purely additive strategy, the strategy of participant 37 was exclusively based on elimination, and participant 32 combined the two features.

Participant 5:	1)	Look at the following attributes: Video clip playback with sound, Stereo FM, 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.
Participant 32:	1)	Eliminate all phones that do not have SMS and whose standby time is less than 300 hours. If one phone remains then choose it.
	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.
Participant 37:	1)	Eliminate all phones that do not have SMS and VibraCall. If one phone remains then choose it.
	2)	If more than one phone remains, select the cheapest phone.
	3)	If two or more products are equal in price, choose the smallest phone.

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

3.3.2 Prediction accuracy

The degree to which the strategies described by the participants could predict their own choices was calculated. The averaged percentage of correct predictions across all 30 trials was 73% (Figure 3.2, 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 and Hypothesis 1 is thus supported. The percentage of correct predictions that resulted from using a certain participant's strategy to predict the choices of all other participants was 34% across all participants (Figure 3.2, 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 is in line with Hypothesis 2 and gives further evidence for the uniqueness of the participants' strategies and indicates that they cannot be replaced easily by each other.

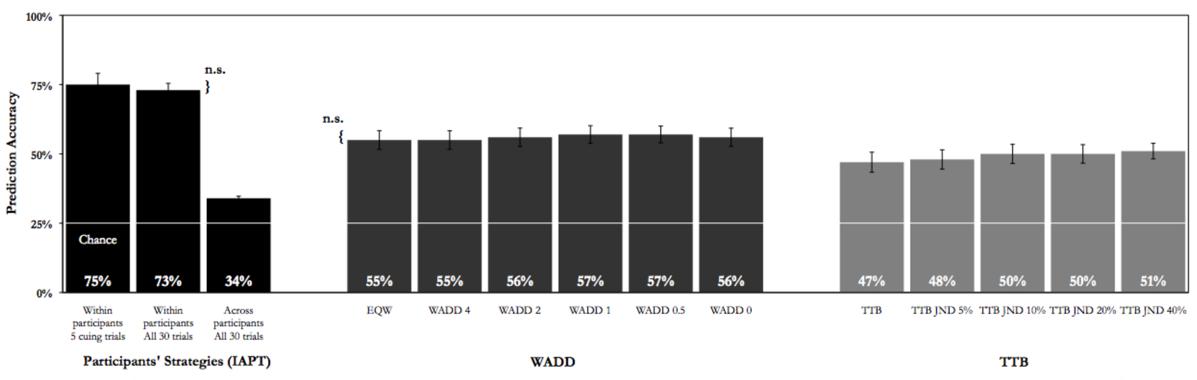

Figure 3.2: Percentage of choices correctly predicted by various decision strategies

Note—The vertical bars denote the standard errors; EQW = EQual Weighting, WADD = Weighted ADDitive, TTB = Take-The-Best, JND = Just Noticeable Difference.

Given that the sample consisted of more male (29) than female (8) participants, it would be interesting to test whether there are any gender differences with respect to the prediction accuracy of the described strategy. Unfortunately, after the exclusion of the six participants only four (13%) of the remaining 31 participants were female, a number that is too low for obtaining meaningful results. When calculating the prediction accuracy separately for each gender, the prediction accuracy was found to be slightly higher for male participants than for female participants (74% vs. 69%, respectively). The main result, however, remains the same.

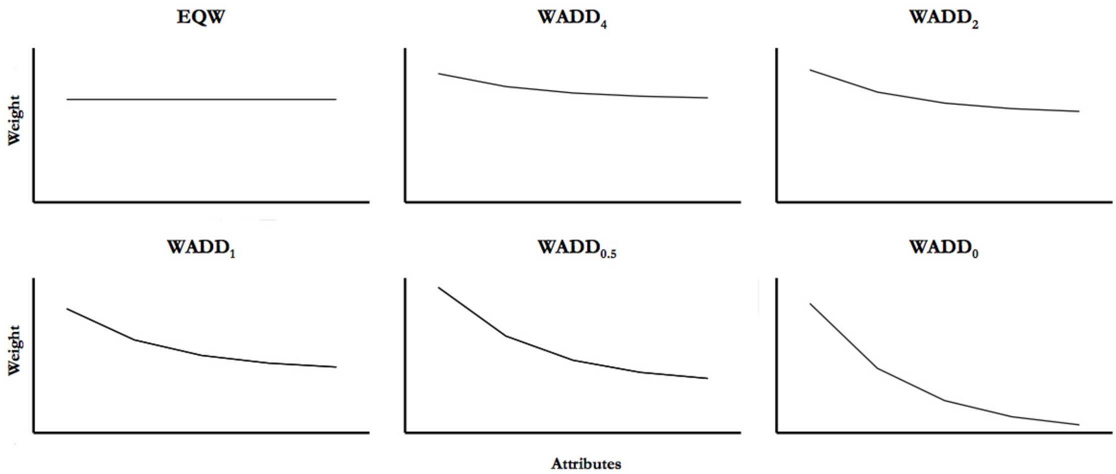

Figure 3.3: The six variants of WADD

In this experiment, six different variants of the Weighted ADDitive (WADD) strategy were used. This strategy is computationally demanding and is considered to be "the traditional gold standard for rational preferences" (Gigerenzer & Goldstein, 1999, p. 26; see also Keeney & Raiffa, 1976; Shah & Oppenheimer, 2008). Moreover, five variants of the take-the-best (TTB) heuristic (Gigerenzer & Goldstein, 1996) were used (with different Just-Noticeable-Differences), a lexicographic strategy that applies one-reason decision making and thus its execution requires much less processing (although they each require a similar level of pre-processing to determine the weights or the cue order). 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.[4] The variants differed with respect to the skewness of these weights. At one extreme, EQual Weights was used. At the other extreme, a set of noncompensatory weights was used, 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 the attribute that was ranked second highest was bigger than the sum of all following weights, and so on ($WADD_0$). Specifically, 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

[4]To be able to sum up the values across the different attributes, they were first normalized by performing z-transformations. In addition, the attributes *weight*, *dimensions*, and *price* were multiplied by -1, because lower values on these attributes are generally perceived as being better.

variants ($WADD_4$, $WADD_2$, $WADD_1$, 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[5] (see Figure 3.3).

Take-the-best was originally formulated for tasks in which two objects had to be compared to each other on a given criterion, as for example "Which nation obtained more gold medals in the last Summer Olympics, Germany or France?" This heuristic first looks up the values on the most valid cue, where cue validity is defined as the relative proportion of correct inferences among all inferences in which this cue discriminated between the alternatives (e.g., the population of each country). If the two alternatives differ on this cue the information search is stopped and the alternative with the higher cue value is selected. If they do not differ, TTB proceeds with the second most valid cue, and so on. Rieskamp and Hoffrage (1999; 2008) generalized TTB from inferential tasks to preferential tasks, and from two-alternative to multi-alternative choice tasks. If two or more alternatives are tied, that is, if they have the same value on the attribute of the highest importance, and no other alternative has an even higher one, 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). Because of its highly selective and attribute-wise information search, TTB can be considered very simple to use once the cue-ordering has been determined.

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. Consider, for instance, someone who cares most about price. After finding out that Phones A, B, C, and D cost 158, 159, 299, and 339 SFR, it is not plausible to assume that he or she will choose Phone A simply because a price of 158 SFR is more attractive than a price of 159 SFR. To capture an insight from early research on psychophysics, versions of TTB were created 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. Five levels of JNDs were used that were applied to all selected attributes, namely 0%, 5%, 10%, 20%, and 40%. The five corresponding strategies are referred to as TTB_0, TTB_5, TTB_{10}, TTB_{20} and TTB_{40}, respectively. For calculating these JNDs, the standard of reference was the alternative with the most attractive attribute value (in the respective trial). For some attributes, this was the alternative with the highest value (e.g., battery life), for others it was the one with the lowest value (e.g., price). For instance, if the most important attribute of a particular participant was price, and the cheapest phone in a given trial cost 100 SFR, TTB_{20} would have eliminated all phones that were more expensive

[5] 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 minimize 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.

than 120 SFR.

The 30 choices of each participant were predicted separately using each of the 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 1. As can be seen in Figure 3.2, 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 ($F(1, 44) = .367, p = .87$).[6] 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 $t(30)$'s > 5.7, all ps $< .001$), which is in favor of Hypothesis 3. Even when the best fitting linear model and the best fitting lexicographic model for each participant were selected, the fit of the reported strategy was better (tied) for 20 (6) of the 31 participants when compared to their best fitting linear strategy, and it was better (tied) for 26 (1) when compared to their best fitting lexicographic strategy. When selecting the particular strategy that predicted each participant's choices best, be it the described strategy, a variant of WADD or a variant of TTB, and calculating the overall fit, the fit was 74% and hence only slightly but significantly higher than the fit achieved with the described strategies alone ($t(30) = -2.193, p = .04$). As a last step, exactly the same analysis was performed again, but this time excluding the described strategies. Now the fit was 66% and hence higher than the overall fit of any of the variants of WADD and TTB, but at the same time lower than the fit achieved when applying IAPT. This difference was significant ($t(30) = 2.850, p = .007$).

3.3.3 Information search

Direction of information search

A problem with the PI is that 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 0. Therefore, instead of taking zero as the reference point to distinguish alternative-wise from attribute-wise search, the expected value of a random search in a particular matrix was used. To obtain these chance PIs, 10,000 random sequences of information search for each participant and each trial were simulated, with the number of boxes opened by the simulation being equal to the number of boxes opened by the participant in the respective trial. For each sequence, the PI was calculated and the mean of these PIs then served as the values for the chance PIs. It turned out that participants' chance PIs ranged between -0.03

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

and 0.62. Twenty-two (71%) participants had an observed PI that differed significantly from their chance PI, which indicated an attribute-wise search, and 5 (16%) of the participants had an observed PI that indicated an alternative-wise search (the remaining 4 [13%] 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).

Across all trials, the variability of search ranged between 0.02 and 3.49 with an overall mean of 1.11. This indicates that participants tended to inspect an uneven amount of information on each alternative. Across all participants, in 35% of the trials the proportion of cells accessed was equal on each alternative (i.e., a standard deviation of zero).[7]

On average, each participant accessed 76% of the information (i.e., depth of search), ranging between 47% and 100% (SD = 17%). Similarly to the variability of search, this points to the use of more noncompensatory decision strategies (Ford et al., 1989).

As a next step, the pattern (i.e., the PI), the variability and the depth of the participants' search for information were compared to that predicted by the described strategies. Regarding the PI, the participants were expected to favor attribute-wise search when their strategy was based on elimination whereas they should search more in an alternative-wise fashion when using an additive strategy. And, indeed, for the eight participants who used strategies based on elimination alone, seven had a significant PI score that indicated attribute-wise search. However, the one and only participant who used a purely additive strategy also had a score that indicated attribute-wise search. For the remaining 18 participants with significant PI scores no claims can be made because these participants used mixed strategies (i.e., elimination *and* additive). When looking at the variability of search, very similar scores were found for all participants regardless of their described strategy. Even worse, the only participant who described an exclusively additive strategy had only two trials with a complete search and hence far less than many other participants who described elimination or mixed strategies. Finally, looking at the depth of search, no significant differences could be found between the participants using elimination-based strategies and all other participants ($t(18) = 1.04, n.s.$), and, curiously, the one participant using an exclusively additive strategy had a depth of search that was lower than most of those found for the participants with elimination strategies. Therefore, these data provide some evidence in favor of Hypothesis 4 but fail to support Hypotheses 5 and 6.

[7] A general problem with this measure is, however, that the standard deviations are greater when one or more alternatives are eliminated early in the process and the subsequent information search concentrates only on the remaining alternatives (i.e., when the strategy has two phases, of which one is more noncompensatory and one is more compensatory). On one hand, this is a characteristic of noncompensatory decision making and hence this fact correctly reflects the processes of the first phase of the decision strategy. On the other hand, however, this measure fails to reflect a possible compensatoriness of the strategy in the second phase. Thus, it should be interpreted carefully because it fails to distinguish between different phases of the decision making process.

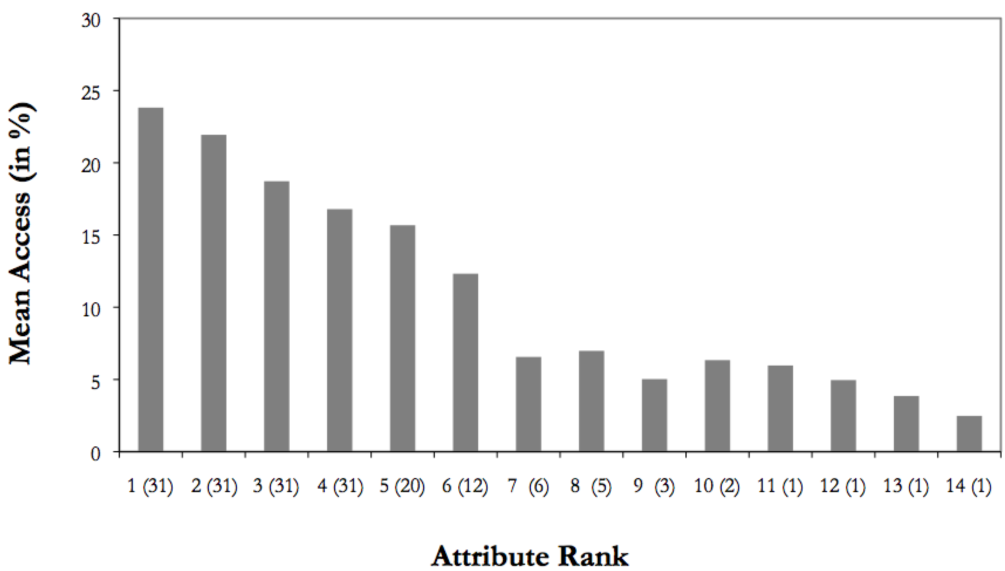

Figure 3.4: Mean proportion of accesses per attribute rank

Note—The numbers in brackets below an attribute rank indicate how many participants used the corresponding number of attributes or more.

Eliminations and information search

Participants stopped search on a particular alternative after its elimination in one third (33%) of the 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. Hence, Hypothesis 7 is not fully supported.

Frequency of access

Generally, the more important an attribute was rated on average, the more often it was accessed by the participants (see Figure 3.4). To determine whether the number of accesses per rank was significantly different from each other, a within-participant one-way ANOVA was conducted with attribute rank as independent variable. Only the first four ranks were used for the analysis because this was the minimum number of attributes selected by individual participants. The linear trend was highly significant ($F(1,30) = 18.9, p < .001$). Then, the correlation between attribute rank and frequency of access was calculated separately for each participant, which was subsequently standardized by means of a Fisher transformation and, finally, the mean over all participants was calculated. This standardized mean was re-transformed and resulted in a correlation of -0.83.[8] This result thus supports Hypothesis 8.

[8]Moreover, the correlation between (1) each participant's correlation between attribute rank and frequency of access and (2) the number of attributes this participant accessed was calculated. This correlation was .025,

It should be noted, however, that attribute importance was confounded with the vertical position on the screen, which may have artificially enhanced this effect.

As a final step, it was verified for how many participants the frequency of accesses decreased monotonously from the highest ranked attribute to the lowest. For 8 (26%) participants the trend was strictly monotonous (i.e., each higher ranked attribute was accessed more frequently than any of the lower ranked ones), for 12 (39%) it was monotonous (i.e., each higher ranked attribute was accessed more frequently than or as frequently as any of the lower ranked ones, and for the remaining 11 (35%) it was not monotonous, though, as Figure 3.4 indicates, most of the violations of the monotonous decrease occurred relatively low in the attribute hierarchy.

3.4 Discussion

The main finding is that people facing a consumer choice situation are indeed able to verbally formalize the strategy they used to make their decisions. The strategies identified with IAPT correctly predicted the observed choices in 73% of the cases, which is far greater than chance. Male participants appeared to be slightly better at describing their strategies than were female participants (i.e., 74% vs. 69%, respectively), but given the very small number of female participants these numbers are not very meaningful. Moreover, the identified strategies were able to predict the actual choices much better than several variants of linear and lexicographic strategies. Thus, these 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 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 on the part of the participants. 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 it was

indicating that participants who used only a small number of attributes did not spuriously inflate the first correlation.

found that linear models outperformed the people on whom these models were based (for a review, see Dawes, Faust, & Meehl, 1989). Moreover, in the second experiment described in Chapter 4, 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.

Still another reason for the mismatch between described strategy and observed process could be that the participant's description of his or her strategy—including the criteria for elimination—resulted from an inductive inference, that is, from an attempt to characterize the conditions under which a specific alternative is eliminated. This description should not be confused with the strategy the participant used when performing the choices—maybe such a strategy did not even exist in the first place and the description has just been constructed post-hoc, after the experimenter requested to do so. In fact, things may even be more complicated in that participants' memories of thoughts during information acquisition will most likely have mixed with their post-hoc attempt to identify patterns and statistical regularities in the set of choices they made. Likewise, it cannot be excluded that participants used configural strategies (Garcia-Retamero, Hoffrage, & Dieckmann, 2007) in Phase 2 but did not bother to report this in Phase 3 as such strategies are complex and thus hard to describe. In other words, the idea that IAPT merely leads to a post-hoc description of the participants' behavior in the preceding choice task instead of truly revealing the underlying strategies cannot be excluded. A possible remedy for this problem is to let participants rank the alternatives in terms of their desirability as opposed to asking them to simply choose one (cf. Riedl et al., 2008). This improvement should be considered in further studies.[9]

3.4.1 Information search vs. described strategies

Many of the described strategies are in line with previous research stating that people often start with a noncompensatory 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 (e.g., Bettman, 1979; Billings & Marcus, 1983; Ford et al., 1989; Gensch, 1987; Olshavsky, 1979; Payne, 1976; Wright & Barbour, 1977, see, however, Glöckner & Betsch, 2008). Such a two-step strategy poses a challenge for any attempt to contrast the strategies and the choices they predict with the information acquisition data. And in fact, the correspondence between described strategies and measures reflecting the information search revealed mixed findings. First, the three measures examining the direction of the information search all indicate that

[9]Unfortunately, this point was raised only after both experiments had been conducted and therefore it is not considered in the following experiment.

participants engaged in noncompensatory processing (at least to a certain extent), which is consistent with previous research (Ford et al., 1989). In particular, for participants who used purely elimination-based strategies, the Payne Index points in the predicted direction of attribute-wise search, but this was not the case for the one participant who reported to have used purely additive strategies: his search was also more attribute-wise than alternative-wise (but note that only very few participants used pure elimination or additive strategies and that this analysis is thus based on a very small sample). The two measures variability and depth of search did not provide any meaningful data when compared to the described strategies. Second, in two-thirds of the trials, participants looked up information for alternatives that they should have already eliminated according to the strategy they described. Third and finally, participants' search for information reflects, by and large, their ranking of the attributes according to their importance.

On the one hand, it seems that the participants have a clear idea of the importance of the attributes they selected and their search processes reflect the ranking they made at the beginning of the experiment. Moreover, the information search measures generally showed more noncompensatory search, which is in line with the participants' strategies because most of these contained elements that suggest noncompensatory processing. In particular, almost all of the participants who used elimination-based strategies searched in an attribute-wise fashion, a search pattern that is to be expected with this type of decision strategy. On the other hand, beyond the general tendency for noncompensatory search, no correspondence could be found between the described strategies and the two measures of variability and depth of search. Also, the information search frequently deviated from the expected pattern given the strategies described by the participants. For instance, participants did not consistently stop information search on an alternative after their strategy prescribed its elimination (only in 33% of the trials). 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), it is reasonable to assume that the decision maker first acquires a certain amount of information and then decides on the strategy (or just certain parameters of it such as thresholds) he or she is going to use.

Overall, the first test of IAPT yielded satisfactory results. In Experiment 2, I sought to further develop and eventually improve this new process tracing method by integrating eye tracking technology. This experiment is presented next, after a brief introduction into the recording of eye movements.

Chapter 4

Experiment 2: InterActive Process Tracing and eye tracking

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 and accurate and need almost no conscious control. Hence, the possibility to track a person's eye movements is expected to yield very reliable and complete information about his or her information search. When an object is examined, the eye fixates an area for about 100–300 ms and then moves on to the next area. The very rapid movements between the fixations are called *saccades* and last for about 30–120 ms (Lohse & Johnson, 1996; Rayner, 1998). The sequence of fixations and saccades across a stimulus is often referred to as the *scanpath*.

Most eye tracking systems work in the following way. A camera points to one or both eyes and constantly sends an infrared light that is reflected on the pupil. The eye tracking software detects the center of the pupil (on the basis of the contrast between pupil and iris) as well as the reflection of the infrared light. The orientation of the eye is calculated based on the relative position of these two points, a procedure which is often referred to as the pupil and corneal reflection method. For the system to work, a calibration is necessary for each participant, which usually takes a few minutes. For the analysis, fixations (i.e., position and duration, among others), saccades, and blinks are detected and can be displayed as raw data as well as visual outputs (e.g., a film of the scanpath).

Whereas with older systems it was often necessary to use bite bars to ensure that the participants do not move their heads during the experiment, modern eye trackers can do without this rather unpleasant measure to make accurate recordings of eye movements. Some high speed eye trackers (operating at a sampling rate of 1250 Hz or even 2000 Hz with a gaze position accuracy of 0.25°–0.5°) use a chin rest for stabilizing, but there are also (slower) systems on the market where the participants can sit freely in front of the experimental screen. These systems can compensate for minor head movements and still have an acceptable speed (typically between 50 and 120 Hz) and accuracy ($< 0.5°$). They are completely unobtrusive because the eye tracking device is often integrated in the screen used for stimulus presentation and hence almost invisible. Therefore, the participants can work on the task as if they were working on a normal computer. Still another type of eye trackers are head-mounted devices, which work at various speeds (50 Hz–500 Hz) and with a gaze position accuracy of around 0.5°. Here, the camera is attached to the head of the participant and the recording is done with a tablet PC, which can be carried in a backpack. These systems are useful for studies in the field (e.g., a marketing study in a supermarket).

Researchers' optimism regarding the recording of eye movements is manifested in a large body of studies which have been conducted in a great variety of disciplines. Among others, eye tracking is used in neuroscience (e.g., attention), psychology (e.g., reading, scene and face perception, visual search, and problem solving), industrial engineering and human factors (e.g., aviation, driving, visual inspection), marketing/advertising (e.g., copy testing, print advertising), and computer science (e.g., gaze as a pointing modality) (for excellent reviews, see Duchowski, 2002, Rayner, 1998, and Wedel & Pieters, 2007). The first eye tracking study to be published was conducted as early as in 1924 by Nixon, who hid in a box behind a curtain to observe eye movements of consumers studying print ads. For the research on cognitive decision strategies, in particular Edward Russo and co-workers made use of this technology. In one paper, Russo and Rosen (1975) used eye tracking to examine cognitive strategies in multialternative choice. Their participants had to choose from sets of 6 used cars. The fixation sequences suggested that participants engaged in many pairwise-comparisons, which was confirmed by verbal protocols. Similarly, Russo and Dosher (1983) presented their participants with simple binary choice problems (e.g., granting a scholarship to one of two applicants) and recorded their eye movements. Again, the interpretations of the eye fixation data were validated with cued retrospective verbal protocols. The authors found that participants processed the available information predominantly in an attribute-wise manner. Even for simple gambles, where computations within alternatives were required, participants searched at least as often attribute-wise as alternative-wise. In addition, participants used procedures to simplify the calculations and to reduce cognitive effort, which, however, led to a (relatively small) increase in errors. In a study on consumer decision making, Russo and Leclerc (1994) recorded the eye fixations of typical shoppers in a simulated supermarket environment. The participants were videotaped through a one-way mirror and were hence not aware that their eye fixations were being recorded. The authors found evidence for three stages in which the process of decision making could be divided. The stages were interpreted as (1) orientation, (2) evaluation, and (3) verification. Finally, in another study where eye movements were observed through a one-way mirror, van Raaij (1977) found that consumers engage in many comparisons between pairs of alternatives.

Based on the assumption that simply looking at product information presented in a matrix (i.e., eye tracking) is more realistic and less under conscious control than using a computer mouse for uncovering cells (i.e., Mouselab), it is reasonable to assume that there will be less reactivity to the experimental method with eye tracking than with Mouselab. However, and despite the popularity of eye tracking technology, very few studies directly compared different information search techniques, with two important exceptions.

In van Raaij's (1977) study, 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 one 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. 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. Generally, the differences between the two methods increased with task size and the participants were more accurate in the gambles task when eye tracking was used. The authors came to the conclusion that the recording of eye movements has several advantages: it elicits faster reaction times, hence it appears that it is less demanding for the participants; it leads to more accurate task performance in choices between gambles, and this was particularly true when processing demands increased; and 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 offers advantages not offered by other methods. Moreover, he argued for a simultaneous use of eye tracking and verbal protocols.

Having the results of van Raaij (1977) and of Lohse and Johnson (1996) in mind, a second experiment was conducted where eye tracking was integrated into IAPT to test for possible influences of the research method on the participants' cognitive processes and behavior, and, ultimately, whether this new method could still be improved by the use of eye tracking in addition to or instead of Mouselab. 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, or, in other words, 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 the first experiment), I explicitly wanted to give the participants the possibility to defer choice in any given set. I think that this is essential for the type of choice situation examined in the present experiments because in real life people frequently (e.g., more than 95% of the time, Sismeiro & Bucklin, 2004) people decide not to buy any of the options available in a certain (online) store. As a theoretical framework the Two-Stage, Two-Threshold (2S2T) framework of choice deferral by White and Hoffrage (2009) was used, which can account for all documented choice deferral phenomena, including the counter-intuitive finding that people defer choice more often when faced with a larger choice set (Iyengar & Lepper, 2000). It assumes two possible reasons or stages for choice deferral. Either none of the presented options is judged as being sufficiently attractive (Stage 1; absolute threshold), or the decision maker is not certain which of the options is best (Stage 2; relative threshold). Note that in this experiment the possibility to defer choice was introduced primarily to increase the reality of the purchase situation and not to study the phenomenon of choice deferral in depth. Therefore, the investigation of choice and deferral behavior is only of explorative character and no hypotheses are formulated regarding this topic. For a specification of the framework and related experiments, see White and Hoffrage (2009) and White et al. (2009).

4.1 Hypotheses

4.1.1 Prediction accuracy

Identical to Experiment 1, it is expected that participants can provide an accurate description of their choice strategies. Based on the assumption that eye tracking allows for a less obtrusive and hence more natural information search than does Mouselab, it is reasonable to hypothesize that there will be less reactivity to the experimental method when this technology is used. Albeit this is a rather speculative assumption, this decrease in reactivity could be manifested in a higher prediction accuracy of the described strategies when eye tracking is used rather than when Mouselab is used.

Hypothesis 9: *The strategies described by the participants will predict their own choices better than a random strategy.*

Hypothesis 10: *The prediction accuracy will be higher in trials where eye tracking is used than in trials where Mouselab is used.*

Given that in the first experiment the described strategies made incorrect predictions in about a quarter of the cases, the question arose of how these prediction errors came about (see the Discussion of Experiment 1 on p. 38). One possible explanation was that people are not always consistent when faced with the same choice twice. Thus, rather than a lack of validity of the achieved descriptions, inconsistencies in the participants' strategy use could be the reason for the observed incorrect predictions. In other words, participants made errors during the execution of their strategies. To test whether participants are consistent, in Experiment 2 some of the trials are presented twice and whether participants make identical choices in both instances is observed. If it is assumed that the execution errors observed in Experiment 1 were due to inconsistencies on the side of the participants rather than due to invalid strategy descriptions, then the prediction accuracy should be higher in the consistent trials than in the totality of the trials.

HYPOTHESIS 11: *The prediction accuracy will be higher in the consistent trials than in the totality of the trials.*

4.1.2 Information search

The acquisition of information needs far less effort with eye tracking than with Mouselab, which leads to two expectations. First, participants are expected to need less time to complete a trial when using eye tracking than when using Mouslab, and, second, they should access more information in the former condition than in the latter.

HYPOTHESIS 12: *The time to complete a trial will be shorter in the eye tracking condition than in the Mouselab condition.*

HYPOTHESIS 13: *The number of accesses will be higher with eye tracking than with Mouselab.*

Similarly, expectations concerning not only the total amount but also the number of reacquisitions of previously accessed information can be formulated. Due to the more effortful information acquisition with Mouselab, participants are likely to memorize the acquired information better with this method than with eye tracking, which in turn reduces the need for reacquisitions (cf. Lohse & Johnson, 1996; Russo, 1978; van Raaij, 1977). Hence, the reacquisition rate is expected to be higher in the eye tracking condition than in the Mouselab condition.

HYPOTHESIS 14: *The reacquisition rate will be higher with eye tracking than with Mouselab.*

In the first experiment, participants stopped search on alternatives that they should have eliminated according to the strategy they described in only a minority of the cases. In other words, very often they acquired more information than their strategy needed for execution. This is in line with Karelaia (2006) who found that people have the tendency to look for confirming information before choosing. Following from this, it is expected that participants will search for more information than their strategy needs for execution.

HYPOTHESIS 15: *Participants will access more information than the strategy they describe needs for execution.*

Finally, the hypotheses regarding the direction of the information search and the relation between attribute rank and frequency of access are identical to the ones formulated in Experiment 1.

HYPOTHESIS 16: *Elimination strategies will lead to an attribute-wise search and additive strategies will lead to an alternative-wise search (as indicated by the Payne Index).*

HYPOTHESIS 17: *Elimination strategies will lead to a higher variability of search than additive strategies.*

HYPOTHESIS 18: *Elimination strategies will lead to a lower depth of search than additive strategies.*

HYPOTHESIS 19: *Participants will access more information on higher ranked attributes than on lower ranked attributes.*

4.2 Method

4.2.1 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 ($SD = 3$).

4.2.2 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 randomly[1] from the pool of phones used in the first experiment (except for one which disappeared from the market in the meantime).

[1] This 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, two orders were excluded: (1) the order that was the exact reverse of the original order and (2) all the orders where two phones were next two each other in the same order as in the first condition. In addition, the first trial of the first condition was never repeated at all and the last trial of the first condition was never repeated as the first trial in the second condition.

4.2.3 Apparatus

For Phases 1 and 2 of IAPT a computer-based process tracing measure was used, which was 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 (see Figure 4.1).

Figure 4.1: Screen-shot of the computer-based process-tracing measure used in Experiment 2

The iView X^{TM}Hi-Speed eye tracker was used, 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. For stimulus presentation, a 17-inch screen was used 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.

4.2.4 Design and procedure

Each participant experienced both of the two conditions, Mouselab (ML) and eye tracking (ET) (in Phase 2 of IAPT), with the order counterbalanced (13 participants began with ML and 14 with ET). Each condition consisted of 12 trials. Thirteen of 27 participants started with ML and the remaining 14 with ET. 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 1 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. The total duration of the experiment was thus about an hour.

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 now the image of the phone was replaced 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, the size of the cells was increased 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 the aim was to keep the situation as natural and realistic as possible, the participants were informed about this limit only when the number of attributes they selected exceeded this number. Apart from the fact that the cells were initially masked 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 it was 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.

4.2.5 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 ($SD = 6$).

4.3 Results

On average, participants selected 7.15 (22%) of the 33 available attributes. The attributes that were selected most often were price (96%), size (85%), stand-by time (59%), and digital camera (56%). Twenty-three attributes were requested by between 52% and 4% of the participants and six attributes were never used (external display, Java games, Office, video playback, video streaming, and voice command). Table 4.1 lists the attributes and the frequency with which they were chosen.

All analyses regarding differences between the conditions were done using a mixed design ANOVA including the within-participants variable of condition and the between-participants variable of order.

4.3.1 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 in the respective set. This is in line with most of the literature on choice deferral (e.g., Dhar, 1997; White & Hoffrage, 2009). The difference between the conditions is not significant ($F(1, 25) = .201, p = .66, MS_e = 99.01$), but the order had a significant effect on the frequency with which participants deferred choice ($F(1, 25) = 6.49, p = .02, MS_e = 880.6$). When ET was used first participants deferred significantly more often (41%) than when Mouselab was used first (20%). The deferral option was used by all but two participants (93%). In most of the deferrals (86%, across conditions) participants indicated that none of the available options was good enough, which corresponds to a Stage 1 deferral in the 2S2T framework.

4.3.2 Described strategies

Again, two general types were found: elimination strategies and additive strategies. 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 (30%) introduced 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

Table 4.1: Number of participants who chose each attribute

Attribute	Chosen by % of Participants	Mean Rank
Price	96	2.19
Size	85	3.13
Stand-by Time	59	3.75
Integrated Digital Camera	56	5.53
Internal Display	52	5.36
Weight	52	4.14
Color	48	6
Bluetooth	41	5.18
MP3 player	41	5.27
Connectivity (USB)	33	3.22
Phone Name	33	5.56
Internal Memory	26	5.57
Stereo FM	11	4
Talk Time	11	6.67
UMTS(3G) Compatibility	7	2.5
Band Support	7	4.5
Email Support (POP3+IMAP4)	7	8
Infrared	7	3.5
SMS	7	5
VibraCall Alert	7	4
External Memory	4	10
GPRS	4	6
Integrated Speakerphone	4	3
MMS (Photos+Text+Sound)	4	7
Speech Recording	4	5
WAP	4	9
Windows© Mobile	4	10
External LCD Display	0	–
Java Games	0	–
Office©	0	–
Video Playback with Sound	0	–
Video Streaming	0	–
Voice Command	0	–

the 27 participants (63%) combined elimination with an additive strategy. Table 4.2 displays three exemplary strategies.

4.3.3 Prediction accuracy

The degree to which the strategies described by the participants could predict their own choices (66%) was slightly, but not significantly, higher ($F(1,25) = 3.96, p = .07, MS_e =$

Table 4.2: Three examples of participants' strategies

Participant 11:	1)	Eliminate all phones that have a stand-by time of less than 250 hours and that are more expensive than 200 euros. If no phone remains then defer choice (because none of the phones is good enough). If one phone remains then choose it.
	2)	If more than one phone remains, calculate a WADD model with all attributes for the remaining phones (without the attribute size). Use the following weights: price and stand-by: 4, mp3: 2, camera and USB: 1. Choose the phone with the highest overall value.
	3)	If there is a tie, choose the cheapest phone if the difference in stand-by is less than 20 hours. If this difference is greater than 20 hours then choose the phone with the higher stand-by time.
Participant 12:	1)	Eliminate all phones that are more expensive than 200 euros, that are not at least tri-band phones, that do not have an integrated speakerphone and that are heavier than 120 grams. If no phone remains then defer choice (because none of the phones is good enough). If one phone remains then choose it.
	2)	If more than one phone remains then take the lightest phone.
	3)	If there is a tie then choose the phone with bluetooth.
	4)	If there is still a tie then choose the cheaper phone.
Participant 26:	1)	Calculate the ratio price/talk time for all phones. Take the phone with the best ratio if it has also the most or equal connectivity features (i.e., UMTS, e-mail, USB, GPRS, Bluetooth).
	2)	If no phone can be chosen this way, take the phone with the most connectivity features, if the ratio price/talk time is not the worst of the four.
	3)	If still no phone can be chosen, take the phone with the second most connectivity features.
	4)	If there are more than one second best phones regarding connectivity then take the phone with the best ratio price/talk time.

139.8) in the ML condition (69%) than in the ET condition (63%, average = 66%). In the repeated trials, participants made the same decision in both instance in only 73% of the cases. This supports Hypothesis 9 but not Hypothesis 10. The prediction accuracy was considerably higher in the consistent trials (78%) than in the inconsistent trials (40%) ($F(1, 22) = 52.3, p < .001, MS_e = 296.3$), which is in line with Hypothesis 11.

Similar to what was found in the first experiment, the prediction accuracy was higher for male than for female participants (68% vs. 59%, respectively). However, the number of female participants was again very low (five out of 27), and therefore no meaningful conclusions can be drawn from this data.

It was also verified whether the participants' strategies were better at predicting choices,

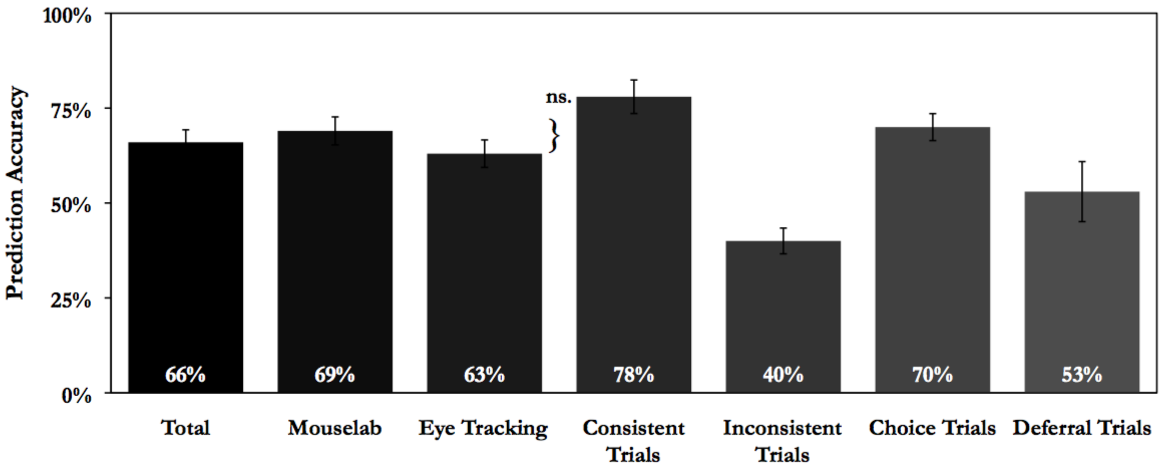

Figure 4.2: Percentage of choices correctly predicted by the participants' decision strategies

Note—The vertical bars denote the standard errors.

choice deferrals, or whether there was no difference between these two decisions. The prediction accuracy was significantly higher in the trials where the participants selected an option (70%) then in the trials in which choice was deferred (53%, $F(1, 18) = 4.81, p = .04, MS_e = 812.7$)[2] (see Figure 4.2).

4.3.4 Information search

An in-depth analysis of the information search data was done to check for possible differences between Mouselab and ET. As in the first experiment, it was also verified whether the described strategies were reflected in the information search data. In particular, I 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 4.4 exemplify some of the results described in the following.

Due to some calibration problems that were 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 and the other 10 with ET. The ET data was analyzed using the software BeGaze (SMI). Fixation position, duration, and sequence (i.e., scanpath) were analyzed. Fixations of less than 100 ms were

[2]For all analyses concerning choices and deferrals, the one participant who never chose and the six participants who never deferred were excluded.

excluded from the analysis.[3]

Time

In general and in line with Hypothesis 12, participants spent significantly more time per trial in the Mouselab 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 depend on whether a phone was chosen or choice was deferred (29.47 vs, 29.49 seconds, respectively; $F(1,22) = .005, p = .95$). However, participants needed significantly more time for Stage 2 deferrals than for Stage 1 deferrals (39.82 vs. 26.54 seconds, respectively, $F(1,7) = 13.1, p = .009, MS_e = 51.39$).[4] This is in line with the assumptions of the 2S2T framework, which predicts that Stage 2 processing only occurs after Stage 1 processing and so Stage 2 deferrals should take longer than Stage 1 deferrals.

Amount of information

The next analysis concerns the amount of information the participants accessed. First, it was 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 predicted by Hypothesis 13, 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 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 ($F(1,18) = 3.01, p = .10, MS_e = 5.393$). This corresponds to 59% (ML) and 63% (ET) of the total information that was available (i.e., depth of search). 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

[3]BeGaze calculates fixations by subtracting saccades and eye 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 $> 75°/s$, (b) the single peak value of velocity lay in the middle 60% between start and end of the event and (c) the duration of the event was > 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.

[4]Only the 9 participants who made Stage 1 *and* Stage 2 deferrals were used for this analysis.

was not significant ($F(1, 18) = 2.28, p = .15$).

Second, the reacquisition rate was assessed, which is the percentage of accesses that were reaccesses of previously seen information (in the same trial). In line with Hypothesis 14, there was a significant difference between the two conditions, with a reacquisition rate of 27% in the Mouselab 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

The information accessed by the participants (IA) was compared with the information that their described strategies needed for execution (SP). In particular, the number of times a participant accessed a cell contrary to the prescription of his or her strategy (CA), and the number of times a participant did not access a cell although his or her strategy prescribed it (CNA). Generally, participants in both conditions accessed more information than their strategy needed, which is evidence in favor of Hypothesis 15. In the Mouselab condition, participants accessed an average of 15.45 cells although their strategies required only 10.63. This difference is significant ($F(1, 18) = 45.2, p < .001, MS_e = 5.156$). Moreover, there was a significant interaction between condition and order ($F(1, 18) = 4.93, p = .04$), but the effect was significant and in the same direction in both conditions. The average number of cells accessed contrary to the requirement of the strategy (CA) was 5.80 as opposed to only 0.98 cells that were not looked up although the described strategy required it (CNA) ($F(1, 18) = 35.5, p < .001, MS_e = 6.570$). The same holds true for the ET condition, where participants accessed an average of 16.73 cells although their strategy required only 10.65 cells ($F(1, 18) = 21.6, p < .001, MS_e = 17.06$). Analogous to the Mouselab condition, the mean CA was 7.48 as opposed to a mean CNA of 1.40 ($F(1, 18) = 30.6, p < .001, MS_e = 12.06$). Apart from CA ($F(1, 18) = 7.24, p = .02, MS_e = 3.858$)[5], these measures did not differ significantly between the two conditions (IA: $F(1, 18) = 1.85, p = .19, MS_e = 8.724$; $SP : F(1, 18) = .03, p = .86$; CNA: $F(1, 18) = .971, p = .34$).

Direction and variability of information search

As already briefly mentioned in the first experiment, the PI has been subject to some criticism (Böckenholt & Hynan, 1994; Stokmans, 1992). For the cases where the number of attributes is not equal to the number of alternatives the PI for a completely random search is unequal to zero. In addition, the PI varies as a function of the number of transitions in a particular trial. Therefore, the PI can lead to inaccurate conclusions and the values of the index observed under

[5] A significant interaction between order and condition was found ($F(1, 18) = 14.8, p = .001$). Only when the participants started with ET the difference in CA was significantly higher in the ET condition than in the Mouselab condition.

different combinations of attributes and alternatives or even different numbers of transitions are not directly comparable (Böckenholt & Hynan, 1994). Still another problem with the PI is that extreme values have a higher probability of occurrence than intermediate values. As a reaction, Böckenholt and Hynan (1994) developed a standardized version of the PI, the SM index. This 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. With this index, extreme values have a lower probability of occurrence than intermediate values.[6] Formally:

$$SM = \frac{\sqrt{N}((AD/N)(r_a - r_d) - (D - A)}{\sqrt{a^2(D-1) + (D^2(A-1)}} \tag{4.1}$$

where N is the total number of transitions, A is the number of alternatives, D is the number of attributes, r_a is the number of alternative-wise transitions and r_d is the number of attribute-wise transitions. The SM index was calculated for each participant and each condition, resulting in the following. In the Mouselab 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 ($F(1, 18) = .777, p = .39, MS_e = 31.20$).

The variability of search was calculated separately for each condition and the number of trials in which the proportion of cells accessed was equal on each alternative (i.e., a standard deviation of zero) was counted. This was found in only 14 (6%) of all trials of the ML condition (across all participants) and in none of the trials of the ET condition. The mean standard deviation across all trials and participants was 3.17 in the ML condition and 6.17 in the ET condition. This difference is significant ($F(1, 18) = 34.8, p < .001, MS_e = 2.579$), showing that the search in the ET condition was much more selective than in the ML condition. The variability of search was also generally higher here than in the first experiment.

The significant SM scores were also compared to the described strategies to check for the degree of correspondence between the described strategies and information search. In the ML condition, five of the six participants who used purely elimination-based strategies had a significant SM score that indicated attribute-wise search. However, the only participant who used a purely additive strategy also had a score that indicated attribute-wise search. The

[6] This index is not without criticism either (e.g., Ball, 1997; Harte & Koele, 2001), but I felt that it is sufficiently informative for the present purposes. Note that it could not be used in Experiment 1 because it is not applicable when the cells of the information board remain open once they have been clicked on.

remaining 13 participants with significant SM scores used mixed strategies (i.e., elimination *and* additive), and therefore their data cannot be interpreted in this context. In the ET condition, virtually the same pattern emerged. Four of the six participants who used purely elimination-based strategies had a significant SM score indicating attribute-wise search. But again, attribute-wise search was also found for the participant with the purely additive strategy. The remaining 11 participants with significant SM scores used mixed strategies. Thus, although the search pattern and the strategy descriptions are in line for those participants who used purely elimination-based strategies, this does not hold true for the participant who used an additive strategy. However, it would be audacious to base a conclusion on only one participant.

Moreover, the participants in the ML condition who used exclusively elimination-based strategies tended to have a lower variability of search than the participants with the mixed strategies. This is contrary to what would be expected. In the ET condition, the mean variability of search was almost identical. Curiously, the participant using the additive strategy had a relatively low variability of search in the ML condition and a relatively high variability of search in the ET condition. This might indicate a limited validity of this measure. In contrast, and more promising, the depth of search was lower for participants with elimination-based strategies (in both conditions) than for participants with mixed strategies. The participant with the additive strategy again showed a strange behavior having one of the highest depths of search overall in the ML condition and a lower depth of search than the participants with mixed strategies in the ET condition. Altogether, Hypotheses 16 and 18 are somewhat supported by the data but there is no evidence in favor of Hypothesis 17.

Further assumptions about the underlying decision strategies used by the participants can be derived by looking at the two measures search pattern and variability of search in combination. In the present case, every participant with significant SM score indicating attribute-wise search had a variability of search greater than zero. This suggests that a strategy similar to the elimination-by-aspects strategy (Tversky, 1972) was used. The one participant who had a more alternative-wise search pattern also had a variability of search larger than zero which points to the use of the conjunctive strategy (cf. Harte & Koele, 2001).

Frequency of access

In a last analysis, I wanted to know whether the frequency with which participants accessed information on the selected attributes reflected the attributes' rank ordering that they had established in the first phase of the experiment. First, the mean frequency of accesses per rank was calculated separately for both conditions. However, the correlations between access rate and attribute rank (see below) were not significantly different in both conditions ($F(1, 19) = .139, p = .71$). Hence, only the results averaged over both conditions are reported here. Overall and in accordance with Hypothesis 19, it was found that the less important an

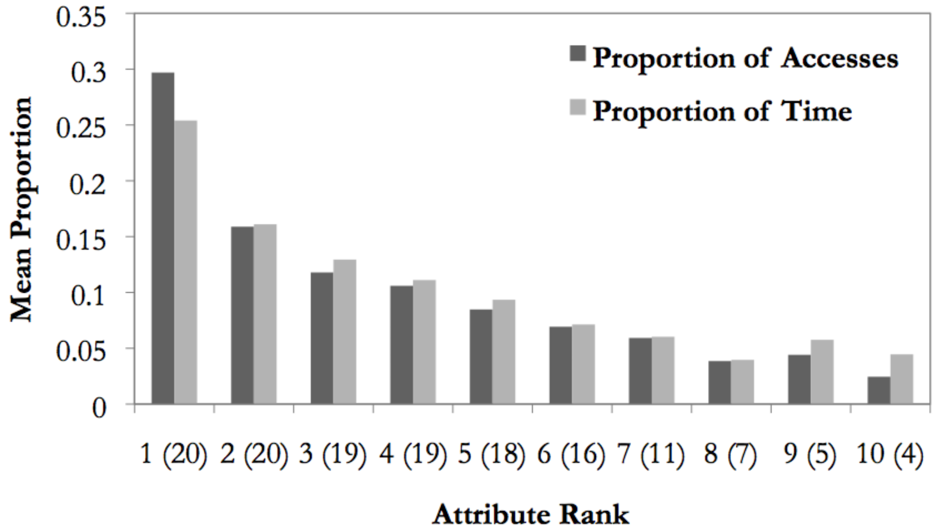

Figure 4.3: Mean proportion of accesses and box opening/fixation times per attribute rank

Note—(1) *Proportion of accesses* refers to the frequency with which a particular attribute rank was accessed. (2) *Proportion of time* refers to the total amount of time spent on a particular attribute rank. (3) The numbers in brackets following an attribute rank indicate how many participants used the corresponding number of attributes or more.

attribute was rated, the less frequently it was accessed by the participant (Figure 4.3). The linear trend was highly significant ($F(1, 30) = 45.4, p < .001$). The standardized mean correlation was -0.91 and hence very closely resembled that found in Experiment 1 (r=-0.83).[7,8]

As a final step, it was verified for how many participants the frequency of accesses decreased monotonously from the highest ranked attribute to the lowest. Looking at each participant separately, only 3 (10%) participants had a strictly monotonous trend (i.e., each higher ranked attribute was accessed more frequently than any of the lower ranked ones), and one single participant (3%) had a monotonous trend (i.e., each higher ranked attribute was accessed more frequently than or as frequently as any of the lower ranked ones). For the majority of participants (16 of 20, 52%) the trend was not monotonous, although the violations were small in many cases and the general trend of decreasing frequency of accesses with decreasing attribute importance could still be observed. Averaged across all participants the trend was not monotonous either, though, as Figure 4.3 indicates, the only violation of the monotonous

[7]Moreover, the correlation between (1) each participant's correlation between attribute rank and frequency of access and (2) the number of attributes this participant accessed was calculated. This correlation was -.12, indicating that participants who used only a small number of attributes did not spuriously inflate the first correlation.

[8]But note that, as in Experiment 1, attribute importance was confounded with the vertical position on the screen.

decrease occurred relatively low in the attribute hierarchy. In this context, I also looked at the latency of search (Ford et al., 1989; Harte & Koele, 2001), that is, the total time spent on each attribute rank (averaged across all participants). Not surprisingly, exactly the same pattern was found as with the access rates (see Figure 4.3). However, the time a box was opened in the Mouselab condition and the fixation time in the ET condition did not differ significantly between the attribute ranks ($F(3, 54) < 1, p > .88$ for both conditions).[9] In other words, the time participants spent on a particular piece of information did not depend on the importance they assigned to the attribute containing this information.

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 they were unsure which phone was best than when they reported that none of the phones was good enough. Second, participants had a significantly higher number of accesses (including reacquisitions) in the ET condition than in the ML condition. 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 it was 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 (i.e., indicating noncompensatory processing), which was in line with the nature of the described strategies. However, the variability of search was significantly higher 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.

A look at the scanpath of Participant 7 in Figure 4.4 is a good example for some of the results described above. What can be seen here is that this participant examined almost exactly the same information (with a few exceptions) in both conditions. However, he had many more fixations in the ET condition than in the Mouselab condition, of which most were reacquisitions of already accessed cells. The times the boxes were open in the Mouselab condition were clearly longer than the fixation times in the ET condition. The direction of his information search was very similar in both conditions.

[9] This analysis was calculated based on only the first four ranks (thereby excluding one participant who had less than four ranks).

(a) (b)

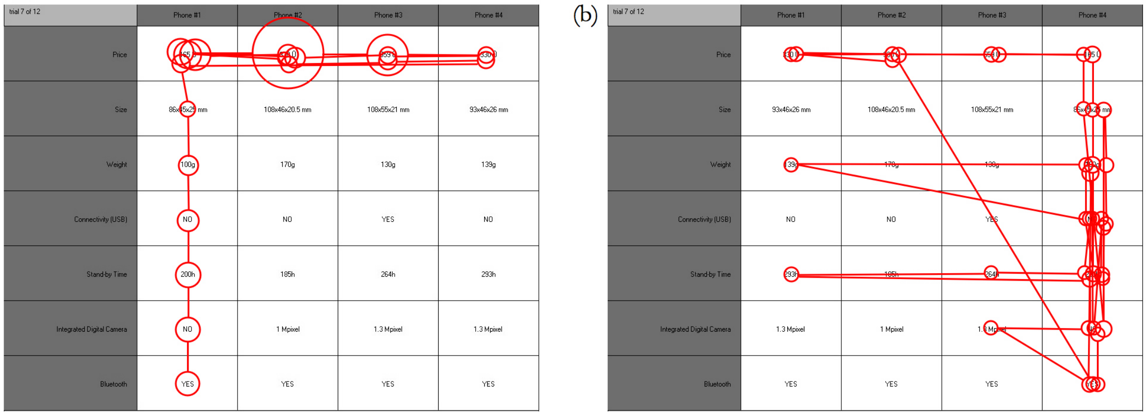

Figure 4.4: Scanpaths of one participant in the ML condition (a) and in the ET condition (b)

Note—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).

4.4 Discussion

In Experiment 2, the finding that the strategies identified with IAPT have good predictive power was successfully replicated. In 66% of the cases the described strategies correctly predicted the participants' choices, which is very similar to the 73% that were observed in Experiment 1. Again male participants were better at describing their strategies than were female participants (68% vs. 59%, respectively) but, as in Experiment 1, the number of female participants was too low to obtain meaningful results. However, the question of whether gender differences regarding the description of choice strategies exist is worthwhile and should be addressed in further studies using more balanced samples.

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 was 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, it was 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 there was some evidence for the 2S2T framework (White et al., 2009), for instance, the fact that Stage 2 deferrals took significantly longer than Stage 1 deferrals. Interestingly, the participants' strategies were far less successful at predicting a choice deferral (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, the difference in the number of accesses can almost completely be attributed to the fact that participants simply reaccessed some cells several times, because participants searched for virtually the same proportion of the total information in both conditions. 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 described by Russo and Leclerc (1994). Moreover, no differences were found between the two information search measures regarding the pattern of search and the relation of determined rank order and frequency of access. These 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 my participants had a very similar depth of search in both conditions. Lohse and Johnson found a slight difference in search pattern (i.e., with more alternative-wise with eye tracking) and their participants unexpectedly searched for less information with eye tracking. In contrast, I did not find any differences on these variables.

What can now be concluded about the use of eye tracking with IAPT? Given that eye movements are less under conscious control than the hand movements required by Mouselab, it is reasonable to assume that eye tracking is less prone to reactivity than Mouselab. However, it appears that the former methodology improves neither the exactness of the description of the cognitive processes nor the quality of the results concerning the information search. In other words, eye tracking did not provide more informative data than did 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. In contrast, the process of information acquisition with Mouselab 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. Moreover, given that the results obtained with both methods were very similar is thus evidence in favor of the latter method. In other words, Mouselab was validated with eye tracking. In sum, despite the technological innovations of the eye tracking technology, Mouselab is still much easier to set up and to use. When using Mouselab, no calibration is necessary and it works with virtually every participant whereas when using eye tracking some of the participants have to be excluded because no reliable calibration can be achieved (in my 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 my experiment, I conclude that Mouselab is the more convenient and efficient method for this kind of task. It should be noted, however, that there is another, very recent process tracing technique called *Flashlight* (Schulte-Mecklenbeck, Murphy, & Hutzler, 2008), which has the potential to mimic laboratory based eye tracking research online. Here, the information on the screen is blurred so that it cannot be read by the participant. However, a small area (i.e., a circle) around the mouse cursor is sharp, similar to a flashlight in the dark. To search for information, participants just have to move the mouse over the areas of interest. Flashlight seems to be a promising alternative here because it is as easy to set up and use as is Mouselab but, similar to eye tracking, it comes closer to a natural information search by providing less structure. Moreover, given that the information search is done with the mouse and not with the eyes, Flashlight may even provide less noisy data than eye tracking because arbitrary fixations that occur during "thinking breaks," for example, are less likely to occur.

In sum, Experiments 1 and 2 demonstrate that IAPT is a valid technique for the identification of consumers' decision strategies. Moreover, the new process tracing technique is not only useful for descriptive research in decision making, but as is shown in the following, the obtained findings also prove beneficial for applied purposes such as the development of decision aids. These systems systematically assist consumers in the execution of typical choice strategies like the ones described above. I have developed such a decision aid, which is described in the subsequent second part of the present dissertation. After an introduction into the topic of decision support systems in general and web-based decision aids in particular, the new choice aid is presented in detail and an experiment that was conducted for its evaluation in terms of usability and general acceptance is described.

Part II

Aiding Preferential Choices

Chapter 5

Decision support systems and web-based decision aids

Broadly speaking, a *decision support system* (DSS) is a "computer-based system that aids the process of decision making" (Finlay, 1994, p. 29). Emerging from both information technology and decision analysis, DSS aim to improve managerial decision making in terms of efficiency and effectiveness. They are used by one or several decision makers and come into operation in many different fields, such as management, medicine, government, or engineering.[1]

A typical application of DSSs is to assist *multi-attribute decision making* (MADM),[2] that is, decisions under certainty where the decision maker chooses among several alternatives, which are described on a varying number of attributes. How many alternatives and attributes are actually considered depends on the decision problem at hand, and, to some extent also on the decision maker. For instance, for the choice of a new apartment, only a few alternatives (< 10) may be considered, whereas a human resources specialist often has to choose among hundreds of job applicants. To represent the (subjective) importance of an attribute relative to the others, each of them receives a weight. These weights are either assigned by the decision maker or elicited by special methods (e.g., ranking, ratio weighting, etc.). MADM problems are usually represented in a matrix format, with the attributes in the rows and the alternatives in the columns (or vice versa). This form of representation has the advantage that it facilitates the detection of the conflict typically inherent in this type of decision problems. Given that often one alternative is better than the other(s) on some attributes but worse on others, the decision maker is forced to make trade-offs to resolve this conflict (Yoon & Hwang, 1995).

In contrast to *descriptive* models of decision making that aim at merely describing human behavior without normative implications (see Part I), MADM models are *prescriptive* models of decision making. The purpose of these models is to provide instruments for making rational

[1]The field of DSS is very broad and diverse, but a more detailed description would be beyond the scope of this work. For good overviews see Finlay (1994) and Holsapple (2008).

[2]Variants of this term are: multiple attribute, multiple objective, multiobjective, multiple criteria or multicriteria (D. L. Olson, 2008).

decisions.

Decision aids are tools within a DSS where multi-attribute analysis is used to support selection decisions (D. L. Olson, 2008). They are intended to help the decision makers to make better decisions from a normative point of view. Typical multi attribute approaches that have been integrated into decision support systems are multiattribute utility theory (MAUT; Keeney & Raiffa, 1976), TOPSIS (Hwang & Yoon, 1981), or the analytic hierarchy process (AHP; Saaty, 1980).

One of the best known and most widely used methods of MADM is the *Weighted ADDitive* method (WADD), also known as the *Simple Additive Weighting* method (SAW). This method, which is based on MAUT, multiplies each attribute value with the corresponding weight and then adds up these weighted values to obtain an overall score for each alternative (cf. Part I, Chapter 3.3.2 on p. 32). The alternative with the highest overall score is chosen. To be able to add up items with different measurement units (e.g., price vs. size), the attribute values have to be normalized. This can be achieved by various methods (see, for example, Yoon & Hwang, 1995). Formally:

$$V_i = \sum_{j=1}^{n} w_j * r_{ij}, \qquad i = 1, \ldots, m \tag{5.1}$$

where w_j are the attribute weights, r_{ij} the normalized attribute values, i the alternatives and j the attributes. An example for a system that is based on MAUT is the software *Logical Decisions*[3] (for more examples, see D. L. Olson, 2008).

The *Technique for Order Preference by Similarity to Ideal Solution* (TOPSIS, Hwang & Yoon, 1981) follows a similar logic. This method first calculates normalized attribute values and weights and then determines the euclidean distance of each alternative to the positive- and negative-ideal solutions (i.e., a hypothetical alternative with the best/worst attribute ratings attainable for the specific set of alternatives under consideration). TOPSIS then chooses the alternative with the maximum similarity to the positive-ideal solution.

A different approach is the analytic hierarchy process by Thomas L. Saaty, which has received much attention since its development in 1980. A complex problem is structured into a hierarchy of attributes, with the overall goal at the top (e.g., buy a mobile phone that will satisfy me), the attributes describing the alternatives in the middle (e.g., price, display size, etc.), and the choice alternatives at the bottom. To determine the relative importance of the attributes, the decision maker has to make pairwise comparisons of all attributes and all alternatives on a nine-point scale. These ratings are then used to calculate each alternative's contribution to the overall goal. An additional feature of AHP is that after the pairwise comparisons the decision maker can perform a sensitivity analysis (i.e., a "what if" analysis) by changing attribute weights and observing how different weights affect the ranking of the

[3]www.logicaldecisions.com

alternatives. Today, the AHP is used as a principal component in many (commercial) decision support systems (e.g., *Expert Choice*,[4] *Decision Lens*,[5] and *Decision Simplifier*[6]).

5.1 Decision aiding on the internet

The internet is a very well suited platform for decision aids due to three reasons. First, it is easily accessible and being accessed by a tremendous amount of people. As of today, the internet has almost 1.5 billion users worldwide[7] and the number of purchases that are made on the internet is constantly growing. According to Forrester research, European e-commerce will reach €263 billion in 2011, which translates to an average spending increase per customer from €1,000 in 2006 to €1,500 in 2011 (Favier & Bouquet, 2006). Second, it stores enormous amounts of information, the basic ingredient for making decisions. Third, the possibilities of today's technology allow for technically advanced implementations of decision aids, which are comparable to its offline counterparts in speed and functionality. In fact, many of the commercial decision support systems now have a web-based version.

These positive features are some of the reasons for the success of online shopping. As opposed to traditional brick-and-mortar stores, online retailers do not face physically imposed limits regarding the number of alternatives per product category. This is certainly an advantage for both the retailer (i.e., competitive edge) and the consumer (i.e., a large choice set), but it entails a problem that is very well known to everyone who has done information search on the world wide web: information overload. For example, at the time of writing, Amazon.com featured more than 2'300 products only in the category "point & shoot digital cameras." On the one hand, having a large choice set is good, because the probability of finding a product that corresponds to one's needs is higher. White and Hoffrage (2009) refer to this positive effect as the "allure of more choice." However, having many alternatives to choose from can also be problematic. In a large choice set, there are often many very similar alternatives, which makes it more difficult to decide which of the products is best. There is hence a "tyranny of too much choice" (White & Hoffrage, 2009). Moreover, as apposed to a traditional shop, an online retailer has no vendor who can assist the consumers to find their way when confronted with a large number of products and product features.

The objective of online decision aids is to overcome the tyranny of too much choice and even, at least to a certain extent, to replace the vendor. They present and structure the available information according to the input provided by the users and can substantially increase the fit between a particular person's information needs and the available information (Ariely, 2000). Given that preferences are often constructed on the spot and that the selection

[4]www.expertchoice.com
[5]www.decisionlens.com
[6]www.decisionduck.com
[7]www.internetworldstats.com/stats.htm, accessed in 2009

of decision strategies is contingent on the task at hand, these information needs tend to vary not only between but also within consumers (over time).

Nowadays, several tools facilitating consumer decision making exist on the internet. Häubl and Trifts (2000) call these tools *interactive decision aids*, where the term (machine) *interactivity* refers to systems that have the following characteristics: "reciprocity in the exchange of information, availability of information on demand, response contingency, customization of content, and real-time feedback" (p. 5; see also Alba et al., 1997; Ariely, 2000; Zack, 1993). The underlying idea of these tools is that "resource-intensive but standardizable, information processing tasks are performed by a computer-based system, thus freeing up some of the human decision maker's processing capacity" (Häubl & Trifts, 2000, p. 6). These kinds of decision aids are also referred to as recommendation agents, which can be defined as systems that "elicit the interests of preferences of individual users for products, either explicitly or implicitly, and make recommendations accordingly" (Xiao & Benbasat, 2007, pp. 138–139). In their overview of interactive decision aids, Häubl and Trifts (2000) used a slightly different categorization and identified two general types: *recommendation agents* and *comparison matrices*. As can be seen in the description of the choice aid presented below, this distinction is very well suitable for the present purpose and I will therefore use it in the remainder of this dissertation.

Recommendation agents (also called *recommender systems* [Adomavicius & Tuzhilin, 2005] or *electronic sales assistants* [Miles, Howes, & Davies, 2000]) help consumers to view products they are very likely to be interested in. These can either be recommendations based on past search behavior or recommendations based on preferences specified by the consumer in terms of a pre-selection of products. For instance, the recommendation agent used in a study by Häubl and Trifts (2000) generates a personalized list of products based on the consumers' attribute importance weights. Using these weights, the agent applies the WADD rule to the alternatives in the choice set and ranks them by overall value. In addition, the consumers can specify acceptance thresholds for the attributes and also determine the desired maximum size of the choice set.

In contrast, comparison matrices display the product information in an attributes-by-alternatives matrix, which allows for a good side-by-side comparison of the products. Examples for comparison matrices can be found on the websites of shopping.com, dpreview, or O_2 Germany. Although it is a promising development, simply displaying product information in a matrix is neither directly aiding the consumers nor does it guarantee that the decision process used is in line with rational standards and procedures. For instance, Fasolo and McClelland (1999) found that although people often tend to look at much or even all of the available information, they do not aggregate the information in a way that would be consistent with a (weighted) additive rule (especially when the attributes were positively correlated). This calls for a decision aid that directly supports the compensatory in-depth comparison of alternatives. At the time being, however, there is no such system available for private consumers, at least to my knowledge.

Decision aids belonging to the two types presented above can now be found on many if not most consumer websites. Edwards and Fasolo (2001) discuss some further tools, which incorporate various ideas and procedures of decision analysis. However, by the time the present work was written, almost every site they present has disappeared. Moreover, some of the (promising) features they mention are not contained in today's web decision aids anymore (e.g., assigning weights to attributes or specifying the preferred range of variation for each attribute). I now first give an overview of decision aids implemented in today's consumer websites and then discuss some related literature.

5.1.1 Web-based decision aids: An overview

The most prominent decision aids that can be found currently are recommendation agents, which can be divided into two types: *collaborative filtering* and *content filtering* agents. Recommendation agents of the first type are based on the users' past browsing and buying behavior and are useful especially for products that are rather difficult to describe on attributes other than price, such as books and music. For instance, one of the major online retailers, Amazon, features a recommendation page where consumers are provided with a list of articles that might be of interest to them. This list is compiled with the help of recommendation algorithms, which elicit the users preference implicitly (e.g.,"people who also bought") or explicitly (via techniques such as question answering, rating options, critiquing, or conversational interaction; Peintner, Viappiani, & Yorke-Smith, 2008) and make suggestions accordingly. The resulting list of recommendations can often be sorted according to certain criteria such as price or customer rating. There is a vast literature on this type of recommendation agents, which use complex algorithms (e.g., collaborative filtering; Linden, Smith, & York, 2003) to find the products that correspond best to the preferences of the users (for a recent review, see Xiao & Benbasat, 2007). However, these recommendation agents are only of minor interest for the present purpose and a discussion of this field would be beyond the scope of this dissertation.

In contrast, content-filtering recommendation agents allow the consumers to view a list with alternatives that correspond to certain criteria (via a pre-selection feature). These recommendation agents are typically used for technical products such as mobile phones or digital cameras. For instance, Amazon provides the possibility to specify the set of displayed alternatives by indicating desired attribute values (e.g., digital cameras that have a resolution of 8 megapixels). This is not surprising, given that the number of products within one product category can sometimes be quite large. Many of today's online retailers (e.g., shopping.com, bestbuy.com) have similar elimination or "winnowing-down" features. However, the attributes that can be used for elimination vary widely from one site to the next and it is quite possible that a decision maker might get very different choice sets from one site to another (Martin & Norton, 2008). In addition to this, many sites feature comparison matrices. Here, the users can sometimes sort the alternatives by attribute values (e.g., price) and reduce the size of the matrix by hiding attributes and/or alternatives. Apart from that, however, no further manipulations such as changing the vertical position of the attributes are possible.

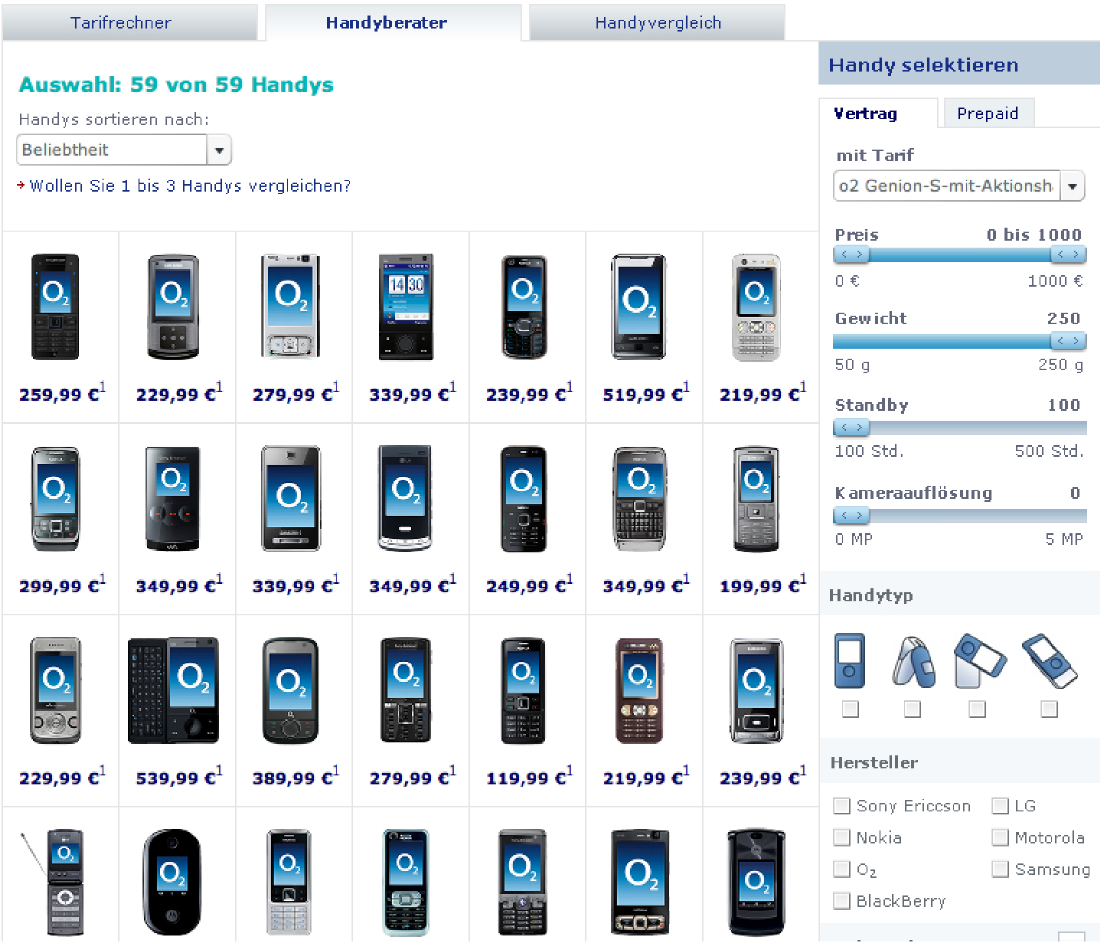

Figure 5.1: The main screen of the O_2 *Handyberater*

A choice aid with a relatively advanced pre-selection tool is the so called "Handyberater" (mobile phone advisor), which can be found on the websites of several german mobile phone carriers (i.e., T-Mobile, E-Plus, and O_2). In this tool (a java applet) the users define cut-offs for continuous attributes such as price or weight by setting sliders to the maximum or minimum acceptable attribute level. For discrete attributes (usually "yes" or "no"), the desired presence of a feature is indicated by checking a box. After each user input, the eliminated alternatives disappear from the screen (see Figure 5.1). This is an advantage as compared to other sites, where the page has to be re-loaded each time a threshold is set or changed. Along with this tool a comparison matrix is provided where a maximum of three alternatives can be compared.

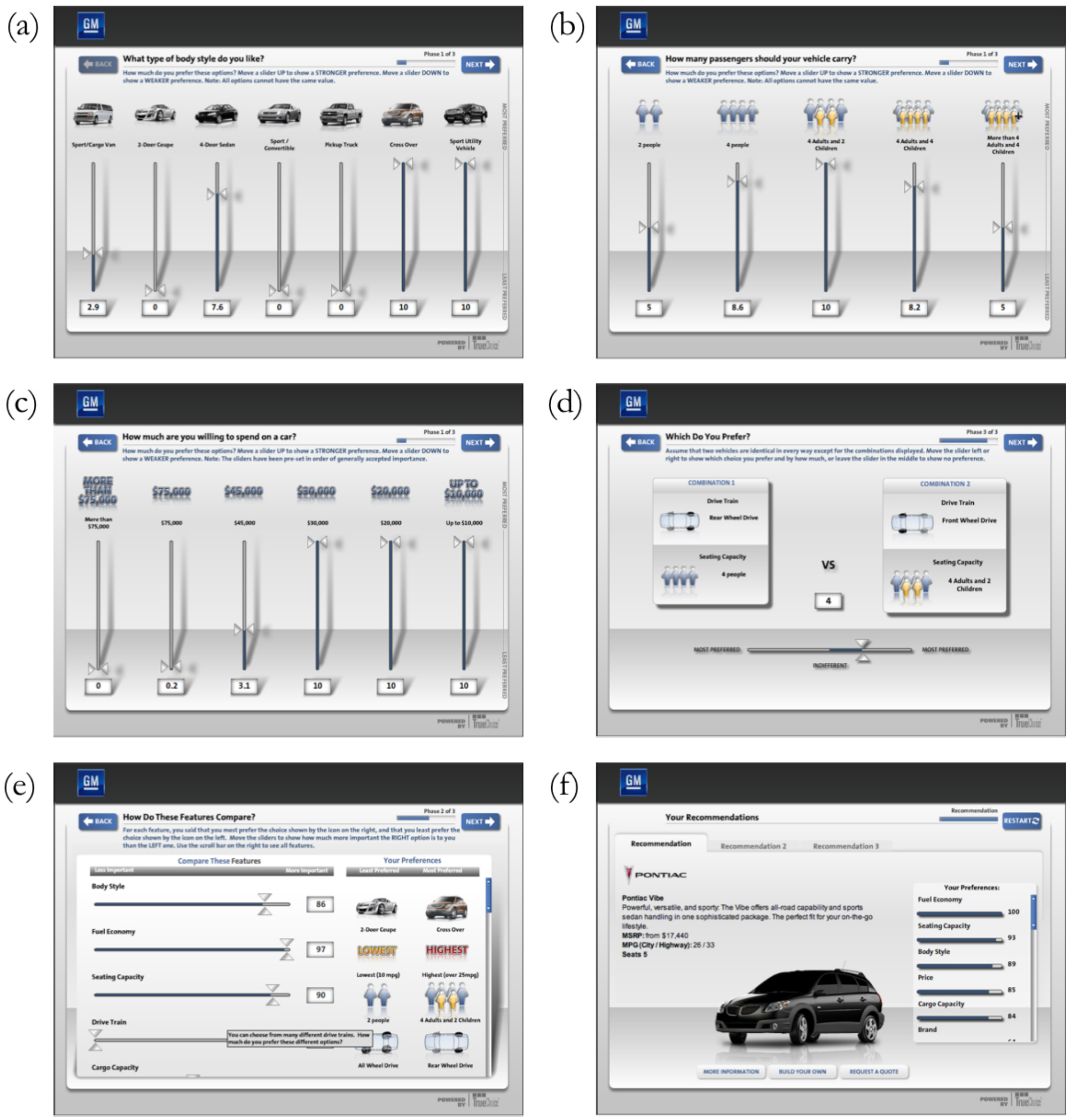

Figure 5.2: Several screens of GM's *Shopping Advisor*

Even more advanced and fundamentally different in its use is General Motor's *Shopping Advisor* for choosing a car, where the users are walked through a step-by-step process with three phases. In the first phase (Figure 5.2 a, b, c), the potential car buyers are asked to indicate their preferences regarding several attributes (e.g., body style, passenger and cargo capacity, and price) by assigning numbers between zero and ten to several discrete values of the attribute (e.g., capacity for 2, 4, 6, 8, or > 8 passengers).[8] Then, in the second phase,

[8]In most cases, the values have a preset of five and to be able to proceed the value of at least one alternative

the users may specify importance weights for each attribute (between zero and hundred, with the default being 50) (Figure 5.2 d). Finally, in the third phase, the users are asked to make pairwise comparisons of sets of two attribute values (assuming everything else to be constant). Specifically, they have to make active trade-offs between attributes that they have rated as important (see Figure 5.2 e). Following that, the system displays the users' attributes weights and recommends three cars (Figure 5.2 f). Unfortunately, however, the users are not told to what degree the three recommended alternatives match their preferences and there is no possibility to make a sensitivity analysis in terms of changing the attribute weights calculated by the system. From what one can tell by using the tool, the underlying mechanism of this choice aid is probably an adaptive conjoint analysis (Johnson, 1985; 1991).

5.1.2 Related literature

Electronic product recommendations, even more than human recommendations, can have a significant influence on the likelihood of purchasing a product. For instance, Senecal and Nantel (2004) found that consumers using a recommendation agent bought recommended products twice as often as consumers who did not receive any recommendations. But not only whether or not but also which product is bought can be influenced by the use of a decision aid. Häubl and Trifts (2000) conducted an experiment to investigate the impact of a recommendation agent and a comparison matrix on consumer decision making. Generally, both tools led to smaller but higher quality consideration sets. Moreover, 93% of the participants selected a nondominated alternative when using a recommendation agent as opposed to only 65% without assistance. Aided participants also switched far less often (21%) to another alternative after their purchase than unaided participants (60%). The authors concluded that these two decision aids simultaneously increase decision quality and decrease effort.

In a study by Jedetski, Adelman, and Yeo (2002) participants were asked to choose several products on two consumer shopping sites (varied between-participants). On one website (Jango) the alternatives were presented in a simple list, whereas the other (CompareNet)[9] provided tools to sort alternatives by specific criteria and to compare the products side by side. Jedetski et al. (2002) found that participants used more compensatory decision strategies and were more satisfied[10] when using the CompareNet than when using Jango (as indicated by a self-report). The number of alternatives also had an effect on the type of decision strategy used: when the number of alternatives increased more participants used noncompensatory strategies. However, this study has some flaws. The participants were presented with a list containing six very different decision strategies (three of them compensatory and three noncompensatory) *before* they made their choices and, after each choice, were asked to select

has to be changed.

[9]Both websites discussed in this paper do not exist anymore in the form described by the authors.

[10]This satisfaction was not affected by the decision strategy used.

the strategy that best described their decision process. Hence, when making their choices, participants could have simply followed one of the presented decision strategies motivated by several reasons: to appear consistent, to please the experimenter, etc. In addition, one probably important strategy for this kind of task was omitted, the elimination-by-aspects strategy (Tversky, 1972).

Ariely (2000) conducted a series of experiments on the costs and benefits of control over the information flow. He found that a high level of control over the displayed information was beneficial in terms of a better matching of preferences, better memory and knowledge about the examined domain, and a higher confidence in the judgment. People also liked the interface more than a very simple one where no control over the information flow was given. However, these benefits were not always present. Given that the use of a system with high information control itself entails a higher demand on processing capacities due to its increased complexity, some learning about how to use the system is necessary to be able to exploit its advantages. Without this learning, the use of such a system can even lead to a decrease in the ability to utilize the presented information. Another finding by Ariely (2000) was that systems with high information control led to a better memory for the information and the organization of the information in memory. Moreover, with these systems participants also had a higher knowledge of the structure of the environment, that is, the relationship between the values of the different attributes.

Finally, Işıklar and Büyüközkan (2007) propose a formal MADM approach for the selection of mobile phones. First, in a questionnaire, the users rate a set of 16 attributes in terms of their importance on a five-point scale. Then, AHP is used to determine the relative weights, and, finally, a rank order of the alternatives is established with TOPSIS. However, although it is an interesting approach, so far it has not been evaluated in terms of usability or user acceptance.

5.1.3 Caveats of web decision aids: The trust issue

Decision aids like the ones described above can greatly facilitate a consumer's way through the data laden hyperspace. However, these increasingly complex and sophisticated systems become less and less transparent. This reduced transparency may raise consumers' suspicions about the degree of altruism of these systems. As pointed out by Häubl and Murray (2006), such choice aids may be designed not only to support consumers but also to serve the sellers interests such as increasing sales, selling particular items and so on (see also Senecal & Nantel, 2004). Thus, consumers might transfer the image they have of real sales people to their electronic counterparts. This is perhaps one of the reasons why recommendation agents still are not as prevalent as they could, or perhaps should, be (according to A. L. Montgomery, Hosanagar, Krishnan, & Clay, 2004, they are used by only 10% of online shoppers). Xiao and Benbasat (2007) propose that the credibility of the provider or the decision aid (e.g., Amazon) might influence the trust users have in the system. Consequently, the same decision aid could be met with different degrees of trust, depending on the hosting website. However, they were not able to find experimental data that supports this hypothesis.

5.2 The InterActive Choice Aid (IACA): A web-based decision aid for online shopping

The rationale for the creation of a decision aid is to improve decision making, or, in other words, to help the decision maker to achieve a good decision outcome, while keeping the effort of the decision process low. That is, a decision aid tries to reduce the trade-off between effort and accuracy implied by the notion of adaptive decision making (Payne et al., 1993; see above on p. 19). However, there is evidence that people are primarily concerned with reducing the effort while maintaining accuracy at an acceptable level. Therefore, to induce the use of normatively oriented strategies such as WADD, the decision aid has to sufficiently reduce the effort needed to execute such a strategy (Todd & Benbasat, 1991; 1992; 1994a; 1994b; 2000).[11] When designing a decision aid, emphasis should thus be placed on the reduction of effort needed to execute a normatively desirable strategy. In addition, the system should help the users to cope with decisional conflict, that is, difficult trade-offs between attributes (Kottemann & Davis, 1991).

As mentioned in the previous section of this chapter, Häubl and Trifts (2000) characterize consumer decision making as a two-step process in which, first, the choice set is screened and promising alternatives are retained, and, second, the retained alternatives are compared in depth (for similar characterizations see Edwards & Fasolo, 2001; O'Keefe & McEachern, 1998). This two-step process is virtually identical to what was found in the experiments described in Part I of the present dissertation. Based on this knowledge about the way consumers make decisions, I wanted to create a choice aid, which systematically helps consumers to make purchase decisions. Rather than imposing a possibly objectively-ideal but unnatural decision procedure on the users (e.g., AHP-based decision aids), the intent was to assist the *natural* process of human decision making by providing explicit support for the execution of the users' decision strategies (i.e., elimination strategies and additive strategies). The rationale for this was to ensure that users accept and understand the system and, ultimately, can benefit from it. A very recent study by Al-Natour, Benbasat, and Cenfetelli (2008) provides positive evidence for this assumption. These authors found that decision aids were rated better in terms of usefulness and trustworthiness when their process was perceived to be similar to the one of the

[11] However, Chu and Spires (2000) have investigated this claim and found that rather than focusing exclusively on effort reduction, a decision maker trades off effort and quality to select the decision strategy. In that sense, decision aids can also cause users to choose a higher effort strategy to increase the accuracy of the decision.

users (see, however, Aksoy & Bloom, 2001). By reducing the effort of the decision process, the choice aid aims to promote more compensatory and hence more normative decision making.

The result of this endeavor, the *InterActive Choice Aid* (IACA), therefore has two phases: a *pre-selection phase*, where undesirable alternatives can be quickly eliminated by setting limits based on attribute values, and a *comparison phase*, where the users can perform side-by-side comparisons of the alternatives. The system has several features to support decision making in both stages and it is interactive in that its output depends on the user's interaction with the system. In contrast to many DSSs, IACA does not focus exclusively on the normative side of decision making (i.e., assisting the users in making decisions that are in line with rational standards and procedures), but it also explicitly takes into account our knowledge about the decision process. In the following, I describe an existing prototype of IACA, which was developed for the evaluation of the system. The screenshots presented below are all taken from this prototype. Some additional specifications of the prototype that are of importance for the evaluation experiment in particular but not for the choice aid in general are listed in Section 6.2.1 (p. 87). The description of IACA is followed by a discussion of its strengths and weaknesses. After that, an experiment is described that was conducted for its evaluation.

5.2.1 Phase 1: Pre-selection of alternatives

In Phase 1, the users can reduce the size of the choice set quickly and conveniently by indicating desirable ranges of values on some or all attributes. For example, many consumers have a good idea of how much they are willing to pay for a product. The possibility to exclude too expensive products from the list of shown alternatives would hence facilitate their information search considerably. In IACA, there are two ways to set such exclusion criteria, depending on the type of the attribute (i.e., continuous vs. discrete).

For continuous attributes such as price, size or stand-by time, IACA features sliders that can be moved to the desired cut-off-level. For instance, if a customer's maximum acceptable price was 100 SFR, he or she would set the slider to this value. The endpoints of the sliders are always the lowest and the highest attribute value, respectively, in the entire choice set. The value corresponding to the actual position of the slider is displayed to its right. For discrete attributes such as whether the phone features a music player or Bluetooth, the users define the acceptance threshold by checking a box for the attributes of interest (i.e., present/absent). As can be seen in Figure 5.3, the sliders and check boxes are on the left side of the screen with the field containing the alternatives is to its right. Note, that the pre-selection phase of IACA is very similar to the *Handyberater* described in Section 5.1.1 (p. 71).

At the outset, all sliders and check boxes are set so that all alternatives are included. For each alternative, the name and a link to another website with further details is provided. When the users move the mouse over the name of a particular product, a window containing

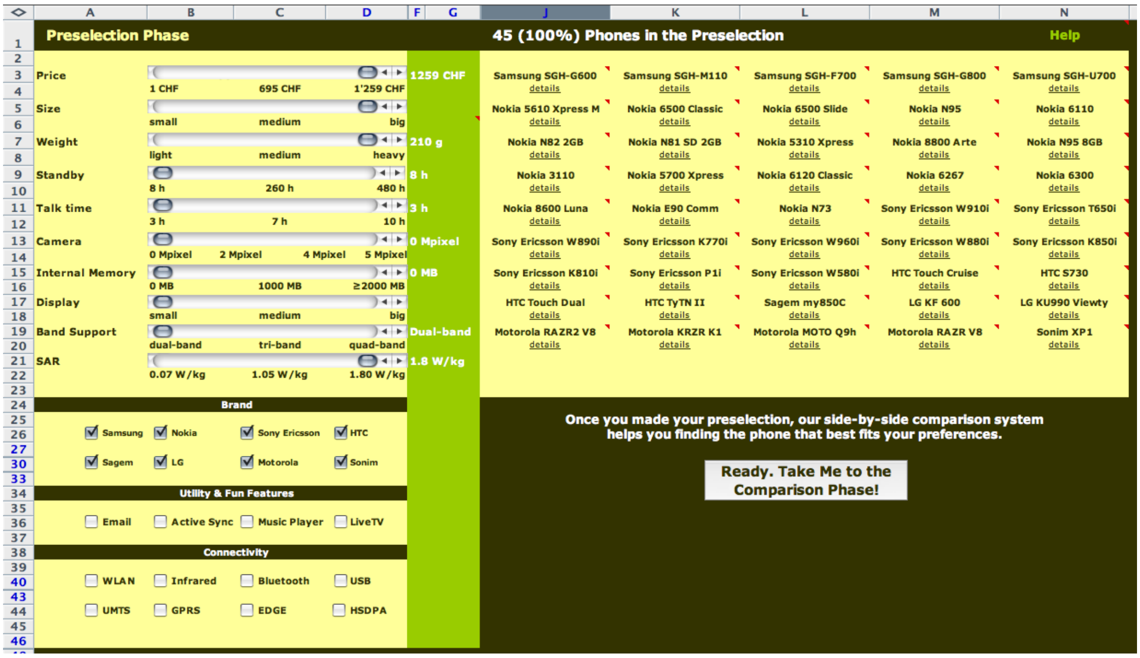

Figure 5.3: The pre-selection phase of IACA at the outset

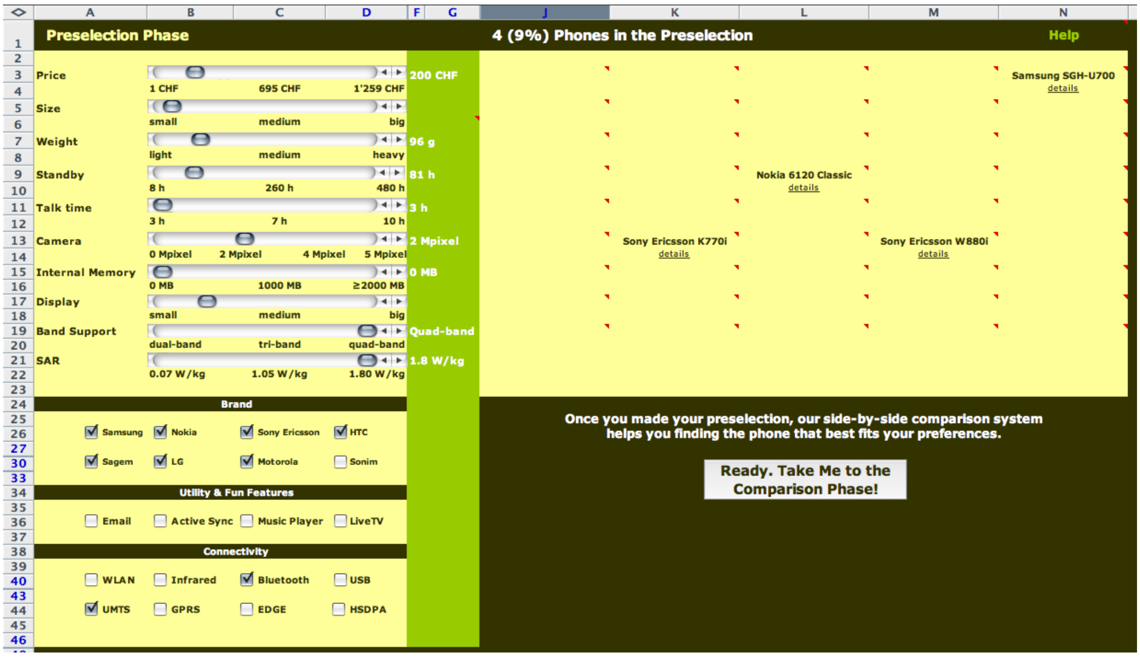

Figure 5.4: The pre-selection phase of IACA after some eliminations

the product picture pops up.[12] When a slider is moved or a box is checked, all alternatives that fail to meet the acceptance threshold on the respective attribute are eliminated. Eliminated

[12]I would have liked to display pictures of the products in addition to the product name and without any action required from the users, but this was technically not feasible for the prototype used (see Chapter 6 on p. 87).

alternatives disappear from the screen (see Figure 5.4).[13] In addition, on the top left of the field containing the alternatives the number of phones that are currently in the choice set is indicated. When the users feel that the choice set has a manageable size for making more detailed comparisons, they can proceed to Phase 2 (i.e., comparison of alternatives) by clicking a button labeled "Ready. Take Me to the Comparison Phase!"

5.2.2 Phase 2: Comparison of alternatives

The second phase of IACA, comparison of alternatives, can be considered a very simple spreadsheet-based decision support system (Seref & Ahuja, 2008). As is shown in Figures 5.5 to 5.7, the screen is dominated by an attributes-by-alternatives matrix with the product features in the rows and the products in the columns. Because the users can adapt this matrix interactively to their personal needs and preferences, it is called the *Interactive Comparison Matrix* (ICM). A list displayed to the left of the ICM contains all available attributes for the particular product category currently in use.

At the top of each column (except for the very first column), product name and picture are displayed along with a link to another webpage containing more details (the same as in Phase 1). Below that, an overall value for each product is shown (see below for details regarding the calculation of that value). The rank of a particular product, which depends on this overall value is displayed above the picture of the phone. Moreover, the three phones with the highest overall value are marked "Best Phone," "Second Best Phone," and "Third Best Phone," respectively. A phone can be selected for purchase by clicking on "Buy Phone" on the bottom of the matrix.

When the users first get to this site, the system displays only the products that were not eliminated in the pre-selection phase. However, at this moment the matrix does not contain any attribute values. The rationale for this was to try to prevent the system from influencing the users too much regarding which information should be considered important (cf. the attribute selection phase of the experiments in Part I). To see information on one of the attributes of the list, the users write the name of the attribute in the leftmost column of the ICM.[14] Upon that, this attribute's values are shown for each product. Note that the vertical position of the attribute plays an important role here. As indicated by a red bar labeled "importance factor" to the left of the first column, more important attributes should be placed higher up (i.e., "extremely important") and less important attributes lower down (i.e., "not so important") the column. Moreover, this importance rating is not simply a rank

[13] Again, this is in part due to the limitations of the prototype. In the final application and for the cases where the initial choice set is not very large (i.e., all alternatives fit conveniently on one page), the picture of the eliminated alternatives would not disappear but just be reduced in contrast. However, when the choice sets are so large that the alternatives are spread out over many pages, this is no longer feasible because one goal of elimination, that is, reducing the information search, would not be achieved given that the amount of pages to visit would remain the same as before the elimination.

[14] In the final application of IACA this will be a drag and drop function and no writing will be required.

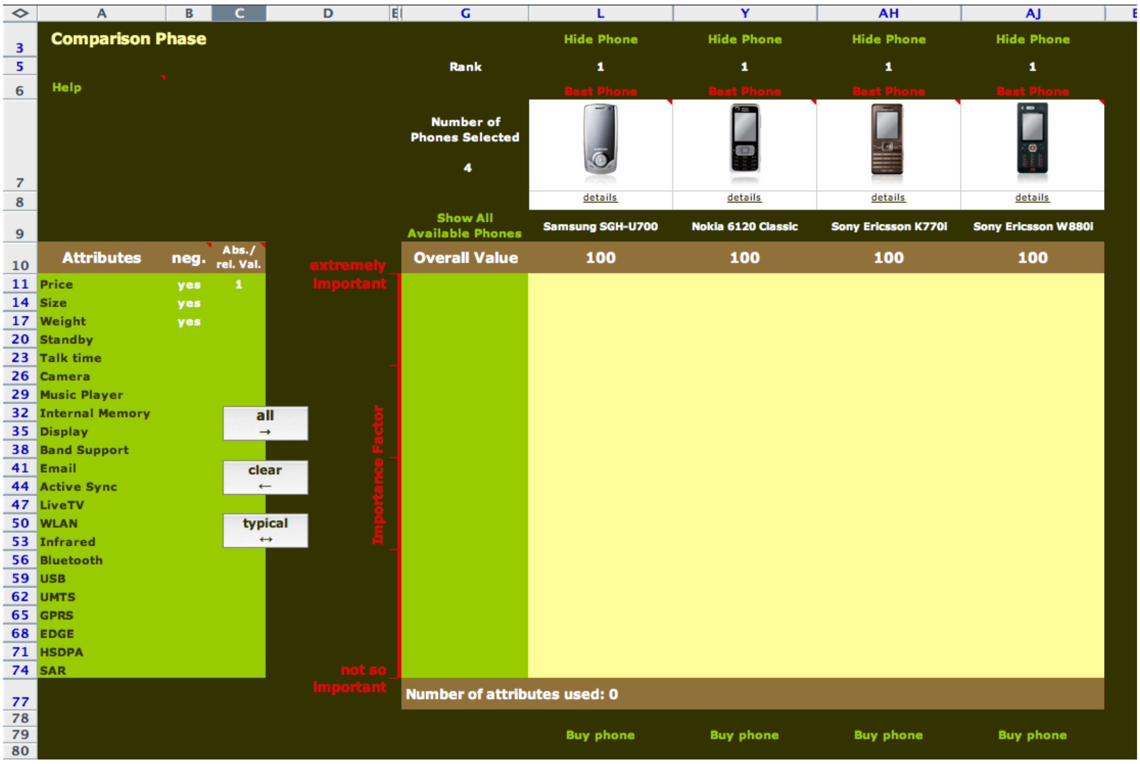

Figure 5.5: The comparison phase of IACA at the outset

ordering. Hence, bigger distances between two attributes mean bigger differences in subjective importance. By making the decision of where to place the attributes, the users define their subjective attribute weights. For each alternative, the weights of each attribute are multiplied with the corresponding attribute values. The sum of these values is the overall value shown for each product. The system hence executes a WADD strategy based on the user input. To be able to calculate this overall value across different attributes with different measurement units, the attribute values have to be normalized first. This normalization allows inter- and intra-attribute comparisons. As in the experiments reported in Part I, the attribute values were normalized by performing z-standardizations on them.[15] The position of the attributes can be changed throughout the entire decision process to see how different weights affect the overall values of the alternatives.

The users could also specify for each attribute whether a high value is desirable (by writing "yes" in the space to the right of the respective attribute) or undesirable (by leaving the space blank). The default is chosen so that most of the consumers would agree (e.g., a lower weight is more desirable than a higher weight but a higher stand-by time is more desirable than a lower stand-by time). This option has the advantage that each user can adapt the system to his or her particular needs (e.g., somebody who for a particular reason wants a very heavy

[15] See below for more ways of standardizing attribute values.

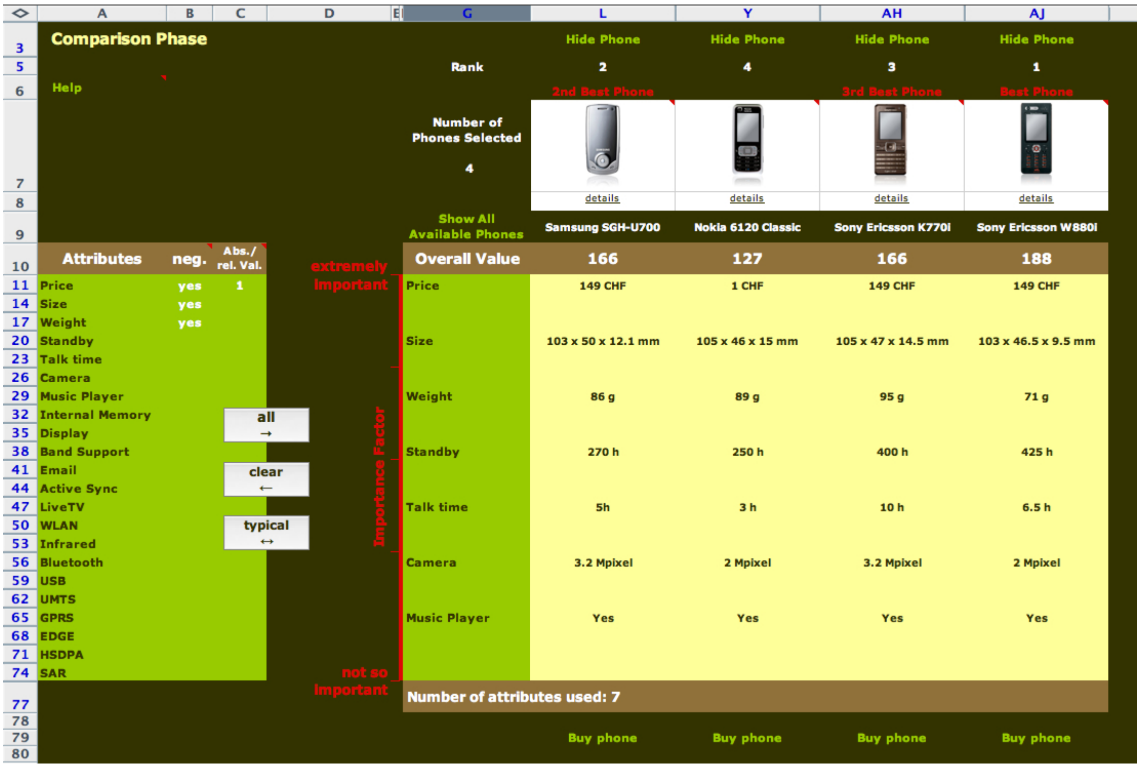

Figure 5.6: The comparison phase of IACA with some attributes added

phone would change from the default). Another possibility the users have here is to change between two different ways the information about the attributes is presented. In the standard setting, absolute attribute values are displayed (e.g., 149 CHF for price, 95g for weight etc.; see Figure 5.6). In the other setting, relative attribute values are displayed. These values are percentages relative to all other products in the entire choice set. For instance, in the choice set used in the experiment described below, a stand-by time of 250 hours corresponds to 52% of the maximum available stand-by time (see Figure 5.7). The rationale for introducing this representation was to facilitate alternative comparison in particular for users who are not experts in the respective product domain. Particularly for attributes such as size or weight, the users can quickly see whether the product is relatively big or small, light or heavy when compared to the other phones they can choose from.

To speed up the attribute selection process, three buttons were introduced to the right of the attribute list. The first, labeled “all,” allows the users to move all available attributes at once into the matrix whereas the second, “clear,” removes all attributes from the matrix. The button labeled “typical” moves a selection of attributes in the matrix, which is based on the behavior of past consumers. This last button was meant to assist consumers who have limited knowledge about the product category and who are therefore not sure which attributes to consider. Finally, there are two more options. First, the users can eliminate a

Figure 5.7: The comparison phase of IACA with relative values

particular product from the matrix by clicking on "Hide Phone" (at the top of each column). Second, the phones that have been eliminated in the pre-selection phase can be included into the matrix by clicking on "Show All Available Phones" (in the first column of the ICM).

With IACA, I have tried to create an environment that facilitates the search for and the integration of product information, the two cardinal components of consumer choice. To what degree this goal has been achieved and what the shortcomings are becomes apparent in the following discussion of the strengths and weaknesses of the new system.

5.2.3 Strengths and weaknesses of the pre-selection phase

To avoid information overload, which is often inevitable given the very large choice sets that are common today, IACA enables the users to quickly eliminate undesirable alternatives (e.g., too expensive) by defining what is an acceptable alternative (i.e., via the sliders and check boxes). In contrast to other systems that have an elimination feature, in IACA the users can base their exclusions on all available attributes (simultaneously) as opposed to only on a few, which are determined by the system designers. This guarantees a high system adaptivity across users. The reason why sliders and check boxes were used, which have a directly visible effect on the size and the composition of the choice set, was to provide accurate and timely feedback about the effects of the threshold setting. For instance, if most of the phones are below 200 CHF, not many eliminations would take place before the users reach this threshold (when starting from the most expensive phone). Similarly, when almost no phone has an infrared connection, checking this box has a big effect and vice versa. Many consumers are not experts in the respective product domain and they therefore have little or no knowledge about what can be considered favorable or unfavorable attribute values. For example, somebody who has never bought a digital camera probably does not know how many megapixels the camera should have, and what is common for cameras that are on the market today. The direct feedback of IACA helps the users to get an idea of the composition of the choice set, for example, that today most cameras have at least seven or eight megapixels. Moreover, the sliders make this process more dynamic than just choosing a value from a drop-down menu or something similar. As a slider is moved from one end to the other, the users can discover the point at which products start to be eliminated and whether this happens slowly (i.e., one after the other) or suddenly (i.e., many at once). In the former case, this indicates that many products have different values, in the latter this means that many products have the same value on the respective attribute. Thus, the users easily and quickly get a feel for the distribution of values for a given attribute.

A potential weakness of the sliders is, however, that only single point cut-offs and no intervals can be determined. Moreover, the fact that all attributes can be used for eliminating has the consequence that the screen is relatively "full," which reduces the understandability of the system.

A basic problem of elimination features in general is that sequential eliminations are a hill climbing process that can lead to local optima, which may be significantly less desirable than the global optimum. For example, if the price cut-off is set to 200 CHF, all subsequent eliminations and comparisons happen in this reduced choice set. This prevents the user from detecting a possible alternative that is only slightly worse on the respective elimination criterion (e.g., 210 CHF) but more desirable overall. Edwards and Fasolo (2001) refer to this as "winnowed-out winners."

5.2.4 Strengths and weaknesses of the comparison phase

Strengths

The problem of information overload is biggest in the pre-selection phase. However, in the comparison phase I also tried to achieve the best possible way of information presentation to ensure that the users can focus exclusively on the information that really matters to them. That is why the ICM is empty at the outset of Phase 2. As opposed to most matrices that are nowadays available on the web, the ICM does not automatically display all available information at once. This is because my past research has shown that consumers are interested in only a small fraction of the available information (see Part I). With IACA, the users therefore have the ability to select the information they want to focus on and they can even arrange it in terms of its subjective importance. The latter matters mainly for the calculation of the overall value, but it also entails another advantage. Given the natural direction of reading in most of the world (i.e., from left to right and from the top to the bottom), it makes sense to display more important information towards the top and less important information towards the bottom of the screen. The fact that attributes of similar importance are closer together than attributes of varying importance further enhances this visual advantage. Another advantage of IACA is the possibility to display relative instead of absolute values. These different forms of information representation makes it easy to see how well an alternative fares in comparison to the others on a particular attribute. This feature is usually not present in a typical comparison matrix, but has been found in other contexts. For example, on one site (www.tigerdirect.com[16]) the ranges of some attributes are presented as lines (e.g, smallest to highest megapixels) and a cross on the line marks the relative position of the product in question.

The second goal, namely, to facilitate information integration by providing assistance for resource intensive calculations, is achieved with the introduction of the alternatives' overall value that is calculated by the system depending on the users' input. This calculation feature aims to help the consumer to resolve difficult tradeoffs that are not easy to mentally calculate. The interactivity of the system guarantees an overall value that is very specific to the users' preferences. A further advantage of the overall value is that it is immune to selective perception. In other words, it "sees" what the users sometimes overlook, especially if they want a particular alternative to come out first in the evaluation (for a reason that they might not be aware of). In addition, the process of selecting and arranging attributes often helps decision makers to get a better view of the decision problem and of what is actually important to them. Things that seemed to be very important at the beginning are often reevaluated when looking at the bigger picture. By re-setting the attribute weights, the users can perform a sensitivity analysis to see how their changes affect the overall value. In sum, this feature supports a more normatively oriented decision process by promoting the use of a compensatory strategy (i.e., WADD). This, in turn, is likely to lead to more accurate choices, or, in other words, to better decisions.

Weaknesses

One main problem with the present prototype of IACA concerns more the prototype itself than the system as such. Due to the fact that it is not a a fully functional website, its usability is clearly below what it could possibly be. However, this prototype was created for a first test that was intended to evaluate IACA's various features in terms of general acceptance and perceived overall utility rather than to run a rigorous usability test of an almost finished

[16]The feature described here has disappeared in the meantime.

application. Therefore, I will not go into more detail regarding usability issues. Further weaknesses are the following.

First, IACA's procedure for weight elicitation is very simple and attribute importance could probably be assessed more accurately with more sophisticated techniques (e.g., ratio method, swing weighting method, tradeoff method, or pricing out method), which have been tested with regard to their validity and consistency (for a discussion of different weight elicitation techniques, see Borcherding, Eppel, & von Winterfeldt, 1991; Borcherding, Schmeer, & Weber, 1995). Moreover, the AHP, which is widely used in organizational contexts, could be a further valid technique for the kind of choices studied here (Işıklar & Büyüközkan, 2007). However, I am skeptical that many consumers would be willing to engage in more complicated weight elicitation processes or even numerous repeated pairwise comparisons as required by the AHP, in particular when they do not really understand how the system works (see also Edwards & Fasolo, 2001). I therefore think that in the present case the benefits of simplicity outweigh the costs of a possibly reduced validity. With IACA, the users can see the whole picture throughout the decision process and the effects of their actions are immediate. Moreover, the way weights are elicited with IACA entails a further advantage. By moving one attribute at a time and holding all others constant, the users can make sensitivity analyses similar to the one used in AHP-based applications. This is in line with Edwards and Fasolo (2001) who suggested that a sensitivity analysis feature should be incorporated into decision aids. Thus, although IACA's method of weight elicitation might be less accurate than others, I am confident that this approach has a clear advantage in terms of user acceptance as well as ease of use.

Second, for the calculation of the overall value, the attribute values were normalized relative to all 45 alternatives by using a z-transformation. However, during the experiment reported below it became clear that this particular way of normalization was less than perfect, in particular because in some rare cases the resulting overall values did not make much sense. This problem disappears when the attribute values are normalized relative to only the alternatives that are being compared in the ICM. For the future development of IACA, this and other ways of normalization should therefore be considered.

Third and again related to the overall value, IACA uses values and not utilities. Given that it is reasonable to assume that many users' utility functions are not linear, this reduces the validity of the overall value. Another, similar problem is that there might be a lack of independence between the attributes. A high positive correlation between two attributes that are both rated as being important leads to double-counting and distorts the overall value. A decision maker might regard a certain combination of two attributes as being better or worse than their weighted sum (Hill & King, 1989, Garcia-Retamero et al., 2007).

However, I again point out that when creating IACA one of the main goals was to have a system that is very transparent and easy to understand. For a manager who makes risky decisions where mistakes are costly, a complex and highly accurate system may be desirable and useful, but the average consumer who buys a comparatively cheap product is probably not willing to invest the time that is needed for the execution of more sophisticated preference elicitation techniques. Moreover, there is evidence that even for professionals training and experience is necessary to provide complete and consistent answers (Tversky, 1974).[17]

In sum, IACA is a new decision aid, which incorporates ideas from various fields of research as well as from similar systems that are on the market today. While some of its features (e.g., pre-selection) already exist in a similar form, others, albeit inspired by existing technologies, are new, at least to my knowledge. As compared to other online decision aids, IACA has more functions that assist the users in their choices. The price for this increase in functionality is, however, a decrease in the ease of use of the system. Now, the question arises whether people are ready to pay this price or whether they prefer easier systems. I tried to answer this question empirically by conducting an experiment, which is described in the next chapter.

[17]For further ways of improving IACA, see the discussion of the experiment in Chapter 6 (p. 105).

Chapter 6

Experiment 3: Evaluating the InterActive Choice Aid

When a new product is created, it usually goes through repeated cycles of testing and refinement, a process often referred to as iterative design (Rubin, 1994). To avoid significant usability problems with the product that become apparent only after the product has been completely designed, prototypes of varying degrees of fidelity have become an integral part of the development process (Grady, 2000; Rudd, Stern, & Isensee, 1996).

Low fidelity prototypes are quickly constructed, very rough approximations of the product. They are limited in functionality and user interaction, and are mainly used to evaluate general concepts, design alternatives or screen layouts. These prototypes are hence used in particular at the beginning of the development process. They can be simple, sometimes even hand drawn paper representations of the product, whose use is mediated by a facilitator. For instance, if the user "presses a button" in the prototype, the facilitator provides a new sheet of paper representing the changed state of the system (i.e., after button press). The advantages of low fidelity prototypes are that their creation is quick and cost effective, with little or no programming skills being required. The cost of their simplicity is, however, that these prototypes can provide little error checking and important design decisions may be overlooked. A further problem is that during testing, they are demonstrated to the users rather than really employed (Rudd et al., 1996).

Although it has been argued that the use of low fidelity prototypes can lead to a reduced detection of usability problems (e.g., Nielsen, 1990), the current evidence suggests that they are equally suited for usability testing than high fidelity prototypes (for an overview, see Sauer & Sonderegger, in press). Another advantage of low fidelity prototypes is that because of their quick and effortless creation, designers are less inclined to defend their design and are therefore more receptive to user suggestions. Likewise, the users have the impression that the prototype is only a rough model, which encourages them to make critical recommendations (Grady, 2000).

In contrast, high fidelity prototypes are valid representations of the product in development. They are completely functional and interactive and can therefore be used and evaluated as if they were the final product. Evaluations based on high fidelity prototypes can be more thorough than those based on their low fidelity counterparts and they can be used to fine-tune the prototype. Moreover, realistic comparisons with other products can be made. However, these prototypes are much more costly in terms of time and money, and substantial programming knowledge is required. This can be problematic because often there are no funds for the development of such a prototype (Rudd et al., 1996).

The prototype of IACA that was created is strictly speaking not a high fidelity prototype, in particular due to the fact that it is not a real website. However, in terms of feel and functionality this prototype comes very close to high fidelity for most of the features. Yet, some features are much more cumbersome than they will be in the final application (e.g., the use of the ICM in Phase 2). Given the early stage of development and the cost that would be incurred by the creation of a truly high fidelity prototype, I decided that the current version would be the best solution. This prototype could be produced relatively quickly and without external help. Moreover, already during its creation many things were noticed that had to be different to how they were originally conceived, which could be implemented immediately. By consequence, the prototype used in the experiment had already gone through some cycles of development.

In the experiment described here, the prototype of IACA was evaluated in terms of perceived utility and user satisfaction. To see how this new decision aid fares in comparison to other applications, it was tested against two other (control) prototypes, which were adaptations of decision aids implemented in real world consumer websites. The three prototypes varied with respect to their functionality, that is, the number and type of tools they featured to facilitate choosing, or, in other words, the degree to which users had influence on how the information about the products was displayed. The tests were performed using two different product categories, mobile phones and digital cameras.

6.1 Hypotheses

Processing a large amount of information takes time. Consequently, a choice aid that allows for a substantial reduction of the information that has to be processed should equally have an effect on the time that is needed to make the decision. Decision aids typically contain several tools that allow for such a reduction of processing effort and time and positive effects have been demonstrated by Hostler, Yoon, and Guimaraes (2005), Pedersen (2000), and Vijayasarathy and Jones (2001). First, the number of alternatives that have to be inspected can be reduced quickly and efficiently with tools for eliminating undesirable alternatives from the consideration set. Second, tools for comparing several alternatives side-by-side make it unnecessary to switch between different websites containing detailed information about the products in question and should thus reduce the decision time further. Third, a tool that provides explicit assistance for resolving tradeoffs is also likely to speed up the process of choosing because the decision maker is required to make fewer calculations.

The three prototypes that were compared in this experiment contain a different number of such tools (see Section 6.2.1, p. 90). The first prototype features a tool for eliminations (low degree of functionality, *LowF*), the second tools for eliminations and side-by-side comparisons (medium degree of functionality, *MedF*), and the third tools for eliminations, side-by-side comparisons and resolving tradeoffs (high degree of functionality, *HighF*). Moreover, the high functionality prototype further reduces the amount of processing by enabling the users to limit the number of attributes that are displayed. Following from this, it is hypothesized that the decision time will decrease from the first prototype to the third.

Hypothesis 20a: *The decision time will be negatively correlated with the degree of prototype functionality, that is, it will be longest for LowF, shorter for MedF and shortest for HighF.*

However, it is also possible that the decision aid will prompt users to make a more careful and deliberate choice, which would be reflected in an increased decision time (E. L. Olson & Widing, 2002). Therefore, Hypothesis 20b reads as follows.

Hypothesis 20b: *The decision time will be positively correlated with the degree of prototype functionality, that is, it will be shortest for the LowF, longer for MedF and longest for HighF.*

These tools should not only have a positive influence on the time, but also on the effort of executing certain decision strategies such as elimination and additive strategies. Given that the three prototypes differ in regard to the number of alternatives the eliminations can be based on, it is expected that the ease of elimination will be rated better when this number is higher. Moreover, a side-by-side comparison tool should significantly increase the ease of comparison of alternatives, in particular when assistance for resolving tradeoffs is provided.

Hypothesis 21: *The ease of elimination of alternatives will be lowest for LowF, higher for MedF, and highest for HighF.*

Hypothesis 22: *The ease of comparison of alternatives will be lowest for LowF, higher for MedF, and highest for HighF.*

Finally, based on the assumption that such tools allow for a more straightforward and deliberate decision process, decision aids with a higher degree of functionality are thus likely to result in a higher confidence that a good choice was made than lower functionality decision aids.

HYPOTHESIS 23: *Decision confidence will be lowest for LowF, higher for MedF, and highest for HighF.*

A negative aspect of highly functional decision aids is, however, that they are perceived as more difficult to use and to understand, simply because the system gets more complex. Therefore:

HYPOTHESIS 24: *The understandability of the choice aid will be highest for LowF, lower for MedF, and lowest for HighF.*

HYPOTHESIS 25: *The ease of use of the choice aid will be highest for LowF, lower for MedF, and lowest for HighF.*

A more difficult to use system could equally lead to worse usability ratings, which are likely to be higher for more simple than for more complex systems. However, if the decision aid greatly facilitates choice in spite of its complexity then its usability should be higher than that for the more basic systems. Therefore, no clear predictions can be made in this case.

6.2 Method

6.2.1 Material

Prototypes

The three prototypes were created with the spreadsheet application Microsoft Excel.[1] They were set up so that they very closely resembled real websites in terms of content, appearance and functionality. Visual appearance and basic functionality were held constant across all three prototypes. Each of them contained a "help function," which was activated when the participants moved the mouse over a field labeled "Help" in the right or left top corner of the screen. Upon activation, a message appeared instructing the participants to refer to the experimenter for any questions. The experimenter answered their questions orally, but tried to mimic as closely as possible a real help function implemented on a web site.

The first prototype (LowF) had the least degree of functionality and was adapted from a major shopping website offering many different products (www.amazon.de). Here, the users could make a pre-selection of alternatives based on some attributes, either by choosing one

[1] Microsoft® Excel® 2004 for Mac, version 11.5.3

Table 6.1: Attributes available in the three prototypes

		Mobile Phones		
LowF (Amazon)	*MedF (Sunrise)*	*HighF (IACA)*	*cont.*	*cont.*
Dual-band	Brand	Price	Band support	Infrared
Mp3 player	Price	Size	SAR	Bluetooth
Camera	Weight	Weight	Brand	USB
Tri-band	Camera	Stand-by	Email	UMTS
UMTS	Bluetooth	Talk time	Active Sync	GPRS
Quad-band	LiveTV	Camera	Music Player	EDGE
Price	UMTS	Internal memory	LiveTV	HSDPA
		Display	WLAN	

		Digital Cameras		
LowF (Amazon)	*MedF (dpreview)*	*cont.*	*HighF (IACA)*	*cont.*
Resolution	Format	Storage types	Price	White balance override
Optical zoom	Price	Uncompressed format	Resolution	Image stabilization
Display size	Resolution	Optical viewfinder	Optical Zoom	Optical viewfinder
Image stabilization	Optical zoom	Display size	Digital Zoom	Manual exposure control
Optical viewfinder	Zoom wide	USB	Zoom Wide	Manual focus
Brand	Digital zoom	Firewire	Min. Aperture Size	Flip display
Price	Image stabilization	Battery	Max. Aperture Size	Self-timer
	Manual focus	Weight	Min. Shutter Speed	Continuous drive
	White balance override	Brand	Max. Shutter Speed	Built-in flash
	Min. shutter speed		Flash Distance	External flash
	Max. shutter speed		Video Resolution	USB
	Built-in flash		Min. ISO Rating	Firewire
	External flash		Max. ISO Rating	Orientation sensor
	Manual exposure control		Size	Uncompressed format
	Movie clips		Weight	Battery type
	Orientation sensor		Brand	Storage types

of several ranges of attribute values (e.g., $>$ 10 megapixels) or by indicating whether the presence of a certain feature was desired (e.g., mp3 player; yes/no). For the price attribute, the users could write their lower and upper limits in separate boxes. The attributes on which this pre-selection could be based are shown in Table 6.1. For the settings to be effective, the page had to be "refreshed," just as in the original web version. This was realized in Excel with a macro that was launched when a button labeled "Show Selection" was clicked. The only other influence users had on the presentation of information was the possibility to sort the products by price and customer rating. The field where the products were displayed, which was almost identical in all three prototypes, showed the name of the product and a link to another webpage containing more details. This webpage, which was visually identical

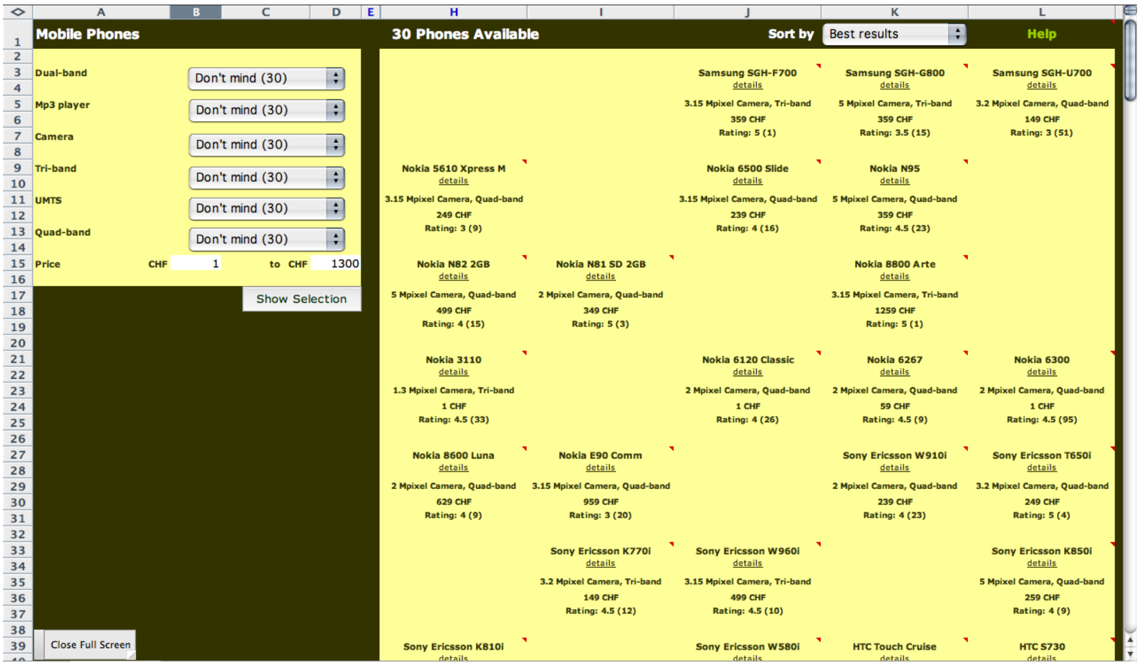

Figure 6.1: The LowF prototype (mobile phone condition)

to the interface of the prototype, was opened in an internet browser (Safari).[2] In contrast to the other two prototypes, underneath each alternative, additional information on two or three attributes was given. These attributes were camera and band support in the mobile phone condition and resolution, optical zoom and image stabilization in the digital camera condition. In addition, price and customer rating were shown (as on Amazon). Figure 6.1 contains a screenshot of the LowF prototype.

The second prototype (MedF), medium degree of functionality) was adapted from another real-world website, which was specific to only one product (www.sunrise.ch for mobile phones and www.dpreview.com for digital cameras). Again, the users could make a pre-selection based on attribute values (see Table 6.1). As in the LowF prototype, the page had to be "refreshed" to see the reduced choice set. To select a particular product for comparison, the users had to check a box to the left of the product name. The comparison could be done using a simple attribute-by-alternative matrix, which was displayed on another screen upon request. This matrix contained all the information about the products but could not be manipulated in any way. To make a decision, participants could either inform the experimenter about their choice or click on "Buy Phone" (which was located underneath each alternative in the comparison matrix) upon which a webpage opened on which it was written "Thank you for your participation" (see Figures 6.2 and 6.3).

The third prototype (HighF) had the highest degree of functionality. Given that this

[2]In fact, participants often did not even notice that they changed the application, even though they were made aware of this and instructed how to switch between applications before the experiment started.

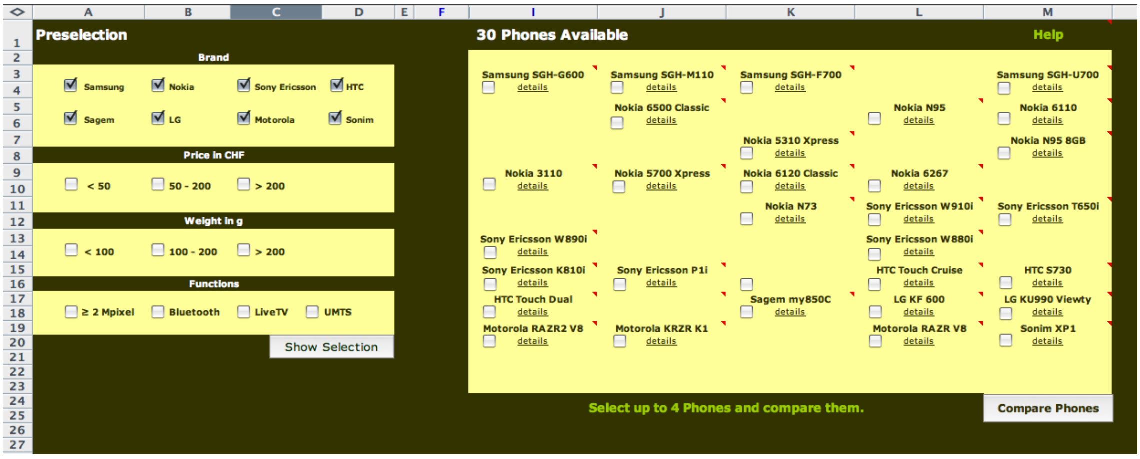

Figure 6.2: The MedF prototype: Pre-selection phase (mobile phone condition)

Figure 6.3: The MedF prototype: Comparison phase (mobile phone condition)

prototype is described in some detail in Section 5.2 (p. 76), here only some details that are important for the present experiment are added. In Phase 1, when the button leading to the second phase was clicked, the following message appeared right underneath it: "Please ask the experimenter to guide you to the next step." Upon that, a facilitator switched to the next screen and prepared the ICM by including the phones that remained after the first phase and the experiment proceeded. In the comparison phase, when the participants clicked on "Hide Phone" or "Show All Available Phones", the facilitator executed the requested action manually by deleting the respective phones from or including them in the matrix. When a participant clicked the button labeled "typical" in the mobile phone condition, the six attributes that were selected most often in Experiments 1 and 2 of Part I were included in the matrix, equally distributed from the top to the bottom. These attributes were: price, size, weight, stand-by time, talk time, camera, and music player. In the digital camera condition no past data could be used. Therefore, attributes that were prominent on many shopping websites were selected, namely, price, resolution, optical zoom, digital zoom, display size, video resolution, and size. A decision could be made in the same way as in MedF.

Task and stimuli

Participants were asked to make three choices, one with each of the three prototypes. Each choice was made out of a set of 30 products, which was drawn randomly from a larger set of 45 products and which was different for each participant and each prototype. Specifically, for each participant, the entire set was randomly divided into three equal parts A, B, and C, of which two were used for each prototype. That is, prototypes 1 to 3 received choice sets AB, AC, and BC, respectively (in terms of the order in the experiment). Two sets of stimuli were used, mobile phones and digital cameras (varied between-participants).

Questionnaires

Two questionnaires were used for the evaluation of the prototypes. To assess the usability of the three prototypes, a usability questionnaire was administered once for each prototype. This questionnaire was based on the *Post-Study System Usability Questionnaire* (PSSUQ) by Lewis (1995), which is a 19-item instrument that measures the users' satisfaction with system usability. Some small changes were made to adapt the questionnaire to my task and research questions. In particular, some minor changes were made to the wording, three items were changed, four removed and two added. The final questionnaire contained 17 items. Participants had to indicate the degree to which they agreed with each statement on a Likert scale ranging from 1 ("strongly disagree") to 7 ("strongly agree"). The only exception was Item 2, where the scale ranged from –3 ("too simple") to +3 ("too complex"). There was also a "not applicable" (N/A) point outside the scale. The questionnaire can be found in the appendix (p. 127).

The second questionnaire directly compared the three prototypes. It was administered after the participants had made a choice with each prototype and contained five dimensions: understandability, ease of use, ease of elimination of alternatives, ease of comparison of alternatives, and choice confidence (see appendix, p. 129). To indicate how well the prototypes fared in comparison to each other on each of these dimensions, participants were asked to locate the three prototypes on a line ranging from "low" (on the left) to "high" (on the right).

Semi-structured interview

In the semi-structured interview, I intended to give the participants room to express their thoughts about my new decision aid. I sought feedback in particular regarding the general acceptance and the perceived overall utility of the system. The interview also served as a check to see whether participants fully understood the system and its functions. To add a bit of structure to the interview, some questions were prepared, which served as a guideline. However, the experimenter could freely choose the questions he asked (and in which order). The questions can be seen in the appendix (p. 130).

6.2.2 Participants and payment

Participants were 24 students (10 female and 14 male) from different faculties of the University of Lausanne and of the Swiss Federal Institute of Technology of Lausanne (EPFL). The mean age was 24.67 years (SD = 3.7 years). They were recruited personally by both the present author and the experimenter, who had generally no prior knowledge regarding the participants' experience with online shopping websites and the internet in general. A flat fee of 20 SFR was paid for participation.

6.2.3 Design

Participants were assigned randomly to one of the two conditions of the between-participants variable *product category* (i.e., mobile phones or digital cameras). Each of them experienced all three conditions of the within-participants variable *degree of functionality* (i.e., low, medium, or high degree of influence). The order was counterbalanced, which resulted in six different orderings.

6.2.4 Procedure

Participants were informed that the goal of the present research was to create a website, which actively aids the decision maker during the process of choosing. In addition, they were told that three prototypes of such a website had been created for this study, which they were now going to evaluate. Besides that, no further information regarding the prototypes and research questions was given.

The experiment consisted of three steps: (1) choices and usability questionnaires, (2) comparison questionnaire, and (3) semi-structured interview. In the first step, participants chose a product with each of the three prototypes and filled out the usability questionnaire after each choice. They were informed that the prototypes were supposed to mimic real webpages in terms of appearance and functionality and that they thus could be used almost as if they were real. Particularities of the prototype that were due to technical constraints (i.e., the hidden pictures of the products or the inclusion of attributes into the ICM) were explained at the beginning of each of the two phases. In addition, given the many functions of IACA two further explanations were introduced in the HighF prototype. First, at the beginning of the pre-selection phase, a window with the following text appeared:

> *Welcome to Choice Advisor!*
>
> *Choice Advisor helps you finding the phone that best corresponds to your preferences. It consists of two phases, a pre-selection and a comparison phase.*
>
> *You are now in the pre-selection phase. Use the rulers and buttons on the left to display only the phones that you would like to examine in further detail.*
>
> *Once you are done, proceed to the comparison phase, where our side-by-side comparison system helps you evaluating and comparing the preselected phones.*

The window was meant to resemble a pop-up window in an internet browser. The participants were told that they could close the window whenever they liked and that this could be done simply by moving the mouse. Second, at the beginning of the comparison phase another pop-up window appeared, which contained a series of screenshots along with some explanations of how to use this phase. Participants could move from one screenshot to another using the arrow keys. Again, they were told that they could close this window whenever they liked. When participants had made a decision, the experimenter noted down the chosen product and proceeded with the experiment. The time participants needed to make a decision was also recorded, thereby excluding the time needed to prepare the screen for the comparison phase for the MedF and HighF Prototypes.

In the second step, participants compared the three prototypes to each other using the comparison questionnaire. Finally, in the third step the semi-structured interview was conducted. At the beginning of this interview, participants were debriefed about the motivation for the study and the meaning of the two control prototypes.

The whole experimented lasted for about one hour. Note that the developer of the prototypes (i.e., the present author) was present as a facilitator throughout the experiment. To avoid desirability effects, the experiment was performed by an experimenter who was blind to the research questions and the origin of the three prototypes. The semi-structured interview was conducted by the developer alone.

6.3 Results

On average, participants compared 4.22 alternatives with MedF (between 2 and 10, $Mdn = 3.5$) and 5.56 alternatives with HighF (between 2 and 27, $Mdn = 4$). This difference is not significant ($t(10) = -1.04, p = .32$). The overall mean number of attributes participants included in the ICM was 7.13 (between 4 and 13, $Mdn = 7$), with 6.7 (between 4 and 10, $Mdn = 7$) in the mobile phone condition and 7.7 (between 5 and 13, $Mdn = 7$) in the digital camera condition ($t(14) = -.933, p = .37$). The attributes that were included most often in the mobile phone condition were price (by 89% of the participants), standby, talk time, USB (67% each), and music player (56%). Eleven attributes were included by between 11% and 44% of the participants and six attributes were never used (i.e., active sync, live TV, infrared, EDGE, HSDPA, SAR). In the digital camera condition, all participants included the attribute price, followed by resolution (86%), optical zoom, and USB (57% each). Each of a further 19 attributes were included by between 14% and 43% and eight attributes were never used (min. shutter, flip display, self-timer, continuous drive, built-in flash, firewire, orientation sensor, and uncompressed format) (cf. Table 6.1).

To have an indication of whether IACA's comparison features led to "better" choices in terms of the overall value of the chosen alternative, the ranks of the alternatives chosen with HighF were compared to the ranks that would have been obtained with MedF if this system had the same comparison features. To obtain these ranks, the participants' individual comparison matrices of the MedF condition were reconstructed with IACA by using the participants' attributes and weights of the HighF condition. Indeed, the mean rank of the chosen alternative was higher (i.e., lower overall value) with MedF (3.09) than with HighF (2.19). However, this difference was only marginally significant ($t(10) = -1.79, p = .10$).[3]

In the following, I first briefly summarize the findings regarding the decision time and then present the results of the two questionnaires and the semi-structured interview. If not stated otherwise, all analyses regarding differences between the three prototypes were done using a mixed design ANOVA including the within-participants variable of prototype and the between-participants variables product category and order.

6.3.1 Decision time

The time needed to select an alternative differed significantly between the three prototypes ($F(2, 24) = 13.7, p < .001, MS_e = 21920$). Overall, decision times were longest for HighF (480 sec.), and significantly shorter for MedF and LowF (314 and 268 sec., respectively). The difference between MedF and HighF was significant ($t(23) = 3.96, p < .001$), but the difference between LowF and MedF was not ($t(23) = .882, p = .39$). There was also a significant effect

[3]Only the 11 participants who used the comparison phase in both the MedF and the HighF prototype were included in these analyses.

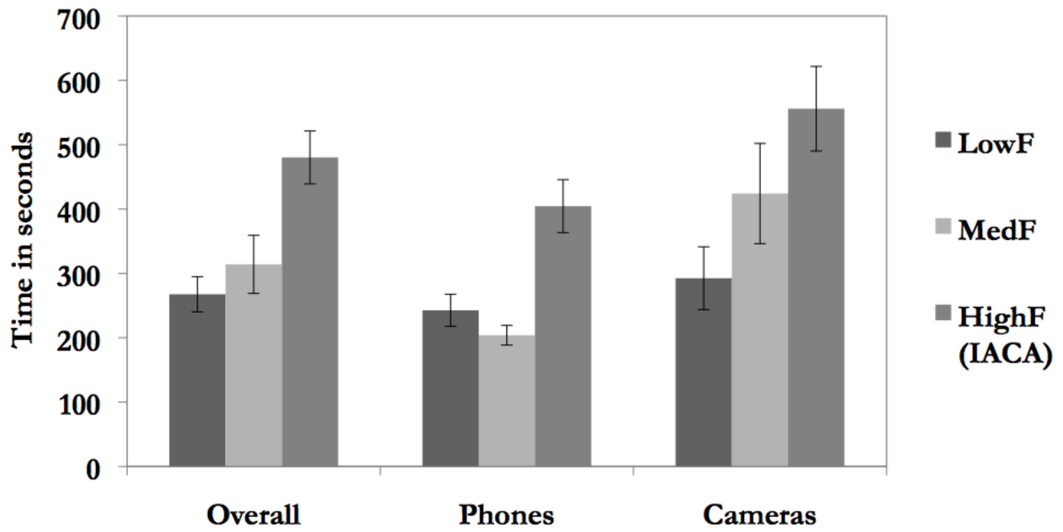

Figure 6.4: Mean choice times for the three prototypes

Note—The vertical bars denote the standard errors.

of product category ($F(1, 12) = 5.48, p = .04, MS_e = 64819$), but the interaction between the conditions prototype and product category was not significant ($F(2, 24) = 2.01, p = .16$). These results provide some evidence for Hypothesis 20b, and, consequently, do not support Hypothesis 20a.

6.3.2 Usability questionnaire

For the following analyses, two items (7 and 8) were omitted because they contained many missing values and participants were generally uncertain how to respond to them. In addition, Item 2 was excluded from the overall ratings because the response scale was different for this item (i.e., from –3 to +3), and the responses could hence not be combined with those of the other items. The reliability calculation of this questionnaire resulted in Cronbach's alphas of .95 (LowF), .97 (MedF), and .95 (HighF).[4] Gender did not have a significant effect on the ratings on any of the items (all p's$>$ .129). This variable was therefore excluded from all further analyses.

Overall, the three prototypes received mean usability ratings of 4.90 (LowF), 5.07 (MedF), and 4.62 (HighF). These values were significantly different ($F(2, 52) = 12.3, p < .001, MS_e = .119$). When looking at the contrasts, it was found that the ratings differed significantly between LowF and MedF ($t(27) = 3.7, p = .001$), but there was no difference between

[4]The inter-item correlations of items 12 and 13 were above .90 for LowF and MedF, which indicates that these two items are redundant. Averaging their values and re-doing the reliability analysis resulted in very similar Cronbach's alphas, which is why the analyses were calculated with the original items.

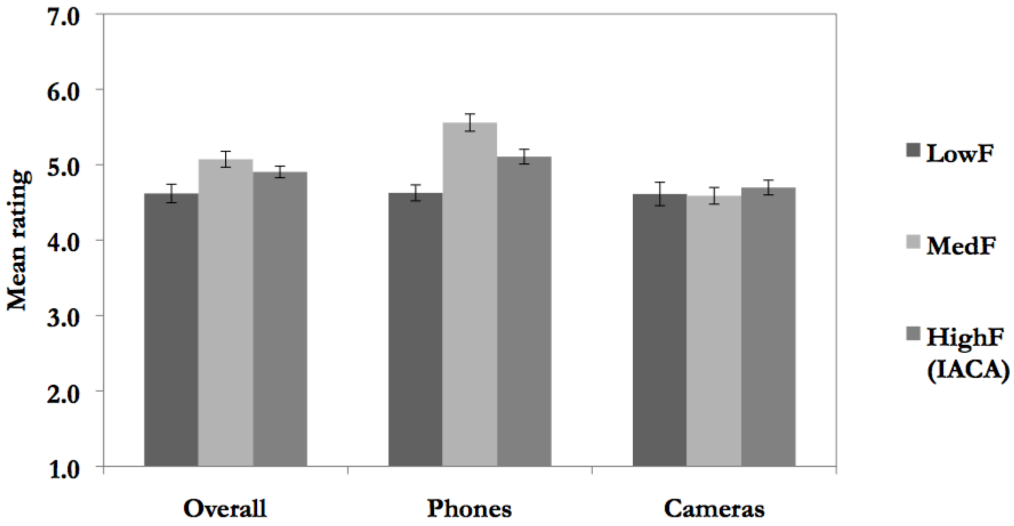

Figure 6.5: Mean overall usability ratings of the three prototypes (items 2, 7 and 8 were excluded).

Note—The vertical bars denote the standard errors.

MedF and HighF ($t(27) = -1.68, p = .11$). Moreover, there was a significant interaction between prototype and product category ($F(2, 52) = 13.6, p < .001$). The participants who chose mobile phones gave significantly different ratings to all three prototypes ($F(2, 52) = 25.6, p < .001$, contrasts: LowF–MedF: $t(13) = 10.0, p < .001$; MedF–HighF: $t(13) = -3.18, p = .007$). However, there was no difference in the ratings in the digital camera condition. Not surprisingly, the product category also had a significant effect on the usability ratings ($F(1, 26) = .4.54, p = .002, MS_e = .372$), which were generally higher in the mobile phone condition.[5] The overall ratings are shown in Figure 6.5.

However, when looking at the items individually, the differences in the ratings assigned to the three prototypes were generally quite small and significant only for Item 2 ($F(2, 22) = 8.16, p = .002, MS_e = .879$).[6] Here, participants rated LowF and MedF as rather too simple (-0.57 and -0.4, respectively, and HighF as rather too complex (0.54). The difference between MedF and HighF was significant ($t(23) = 3.32, p = .003$), but the ratings of LowF and MedF did not differ significantly ($t(22) = .385, p = .70$). However, in all three cases the mean ratings were close to the neutral mid-point (i.e., 0).

[5]When using only the first condition of each participant to simulate a between-participants design, no significant differences in the ratings of the three prototypes could be found neither overall nor for any of the items (all ps $> .05$).

[6]Note, that the significance test of Item 14 resulted in a p-value of .04 ($F(2, 24) = 3.58$). However, to maintain the familywise error rate over all items (except for items 7 and 8), the Bonferroni correction was used. The resulting statistical significance level against which the obtained p-values were compared was $p = \frac{1}{15} = .003$.

6.3.3 Comparison questionnaire

Given that I wanted the participants to use the prototypes as if they were in a real online shopping situation, they were not pointed to any of the features of the three systems. As a consequence, not every participant used the comparison phase in MedF and HighF.[7] Specifically, 18 participants (75%) used the comparison phase in MedF and 16 (67%) in HighF, but only 11 participants (46%) used it in both prototypes. The following analyses will be done for all participants as well as for the subset of 11 participants who used the comparison phase in the prototypes that featured one.

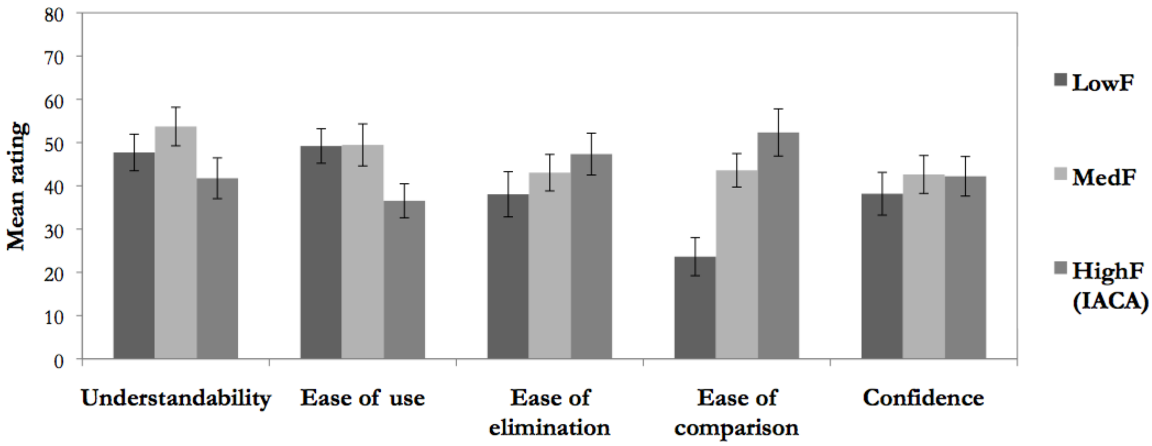

Figure 6.6: Mean participant ratings of the three prototypes on each of five dimensions (all participants)

Note—The vertical bars denote the standard errors.

To obtain the values for the analyses, the positions of the three points on each line were measured in mm, starting from the left end of the line (i.e., low). The values so obtained ranged from 0 to 78.

The average overall rating (AR) was calculated for each prototype across all five dimensions. Overall, participants preferred MedF (AR = 232) over HighF (AR = 220) and LowF (AR = 197). When looking at the two product categories separately, the same picture was found for the mobile phones (ARs = 267, 229, and 200, respectively), but a preference for HighF (AR = 211) over MedF (AR = 198) and LowF (AR = 193) in the digital cameras condition.

Gender of the participants did not have a significant effect on the ratings, neither overall

[7]When I asked the participants in the interview why they did not use the comparison phase, most of them admitted that they simply had not seen the respective button or that they got so confused during the pre-selection phase that they forgot about it. One participant thought the comparison phase of HighF would be identical to the one of MedF, which he had seen before.

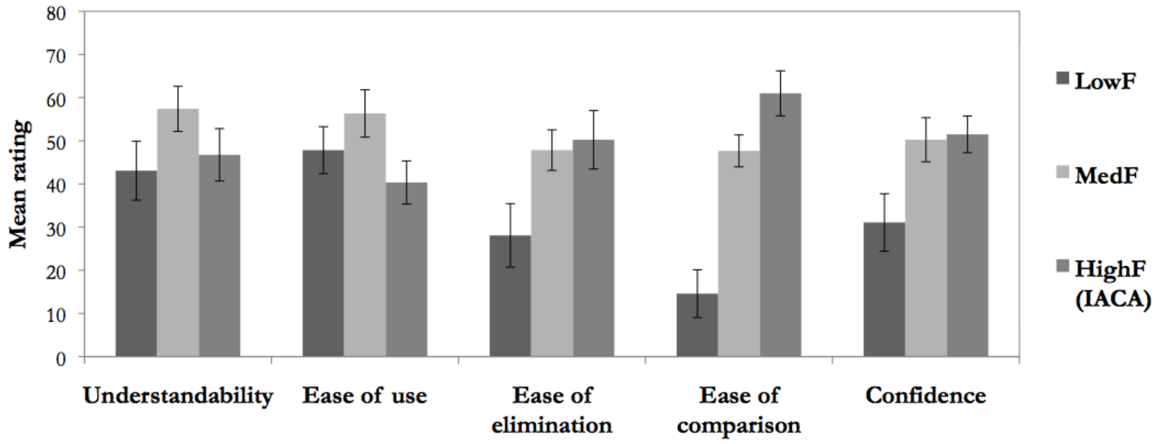

Figure 6.7: Mean participant ratings of the three prototypes on each of five dimensions (participants who used the comparison feature in both prototypes)

Note—The vertical bars denote the standard errors.

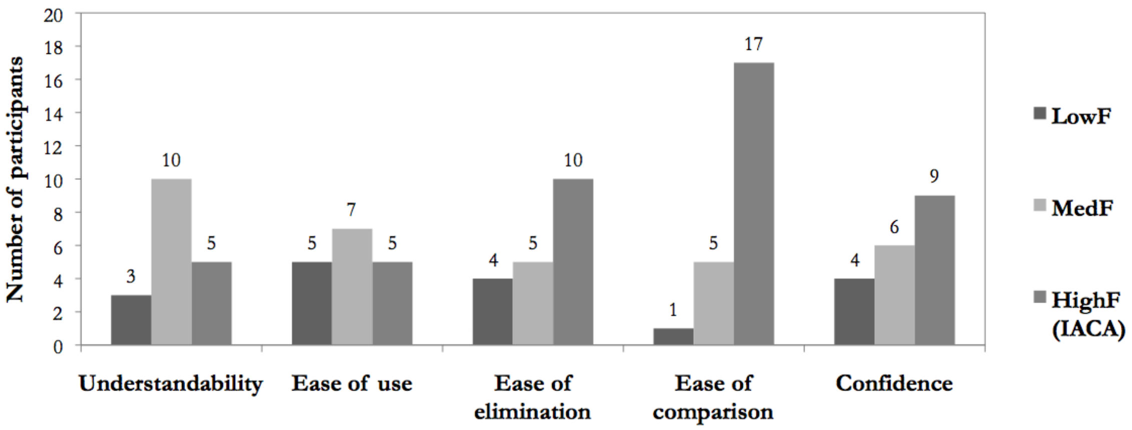

Figure 6.8: Number of times each prototype received the highest rating (all participants)

nor on any of the five dimensions (all p's $> .14$). This variable was therefore excluded from all further analyses.

To test whether these differences between the three prototypes were significant a mixed design MANOVA was run, which included the within-participants variable of degree of influence, the between-participants variables of product category and order, and the five dependent variables understandability, ease of use, ease of elimination, ease of comparison, and choice confidence. Overall, there were no significant differences between the three prototypes ($F(10,3) = 3.682, p = .16$). When including only LowF and HighF in the analysis, the overall effect was significant ($F(5,8) = 8.68, p = .004$). However, the difference was only significant for one dimension, namely, ease of comparison ($F(1,12) = 11.3, p = .006, MS_e = 872.5$).

A look at the contrasts revealed that the ratings did not differ significantly between MedF and HighF ($F(1, 12) = 1.33, p = .28, MS_e = 692.6$), but they did between LowF and MedF ($F(1, 12) = 16.1, p = .002, MS_e = 298.4$). Product category and order of the prototypes did not have a significant effect on the ratings (see Figure 6.6). Including only the 11 participants who used the comparison features in both MedF and HighF did not change the results. Here, six participants (54%) gave the highest ratings to MedF and five participants (46%) to HighF. In this subset, no participant rated LowF best (see Figure 6.7).

In a final analysis, the number of times each prototype received the highest rating was counted (ignoring ties) to see which prototype fared best on each dimension. As can be seen in Figure 6.8, most participants rated MedF to be the most understandable and the most easy to use, but preferred HighF regarding the ease of elimination, the ease of comparison and the confidence in their choice. However, these differences are only significant for ease of comparison ($\chi^2 = 18.1, p < .001$). Overall, 11 of 24 participants (46%) gave the highest ratings to MedF, 9 (37%) to HighF, and 4 (17%) to LowF. These differences are not significant ($\chi^2 = 3.25, p = .197$), and even if MedF and HighF are combined, the difference is still not significant ($\chi^2 = 3, p = .08$). Hence, there is evidence in favor of Hypothesis 22, and, albeit not being significant, these results also provide some support for Hypotheses 21 and 23. However, the data do not support Hypotheses 24 and 25. Not only are the results not significant, the means are contrary to what was expected, too.

6.3.4 Semi-structured interview

In this section, I first briefly summarize the participants' answers to the general Questions 1 to 9 and then describe the reported problems and possible improvements of IAPT (Questions 10 and 11). After that, a short summary of the answers to the specific questions is provided.

General questions

All but two participants (92%) indicated that they would use IACA (HighF) if it existed as a real website and all but one (22 of 23, 96%)[8] would use it at least once when searching for information about a product they want to purchase. However, only 6 of 22 participants (27%) would be willing to pay to use the system. The products participants could imagine to choose with IACA were mostly from the category of consumer electronics (e.g., mobile phones, digital cameras, computers, and mp3 players), but also other things like mobile phone plans, insurances, cars, holidays (e.g., flights and hotels), apartments, sports equipment (e.g., bikes and sports clothing), furniture, vacuum cleaners and music instruments were mentioned. About two thirds of the participants (15 of 23) would use the possibility of creating their own attributes and 74% (17 of 23) would like to have an attribute representing the product rating

[8]The reason why the number of total participants differs from question to question is that these questions were not always asked. See the Materials section (p. 95).

by a known review site or journal. All participants (16) stated that the interface helped them to select the way they like to choose and only one (of 15, 7%) participant thought that he changed his habitual way of choosing due to the architecture of the system. When asked whether they thought that they detected and explored all possibilities of the system or whether they would benefit from the instruction of an expert, the participants gave a mean rating of 4.7 (on a scale from 1 to 7, where 1 is a lot profit from the instruction and 7 is little profit from the instruction). Overall, IACA was rated to be of medium complexity (3.8 on a scale from 1 [not at all complex] to 7 [very complex]) and all participants would recommend the system to others, mostly to friends and family. Lastly, all of them thought that most people would learn to use the system very quickly.

Problems of IACA

Most usability problems were reported to be in the pre-selection phase. A frequent comment was that there was too much information on the screen (mentioned by at least five participants). Some participants had problems with the sliders in the pre-selection phase because it was not clear to them that they did not have to move all of them. When acting on this assumption, the threshold setting usually was a very cumbersome process because the elimination criteria became quickly too severe and no alternatives remained in the choice set. To go back, that is, to set the thresholds so that at least some products were in the choice set was often even more difficult. For some participants, the sliders were too sensitive (i.e., very small movements lead to many eliminations). In general, it seemed that the solution with the sliders was not optimal and even some of the participants who perfectly understood how to use them said that they would prefer other methods for setting thresholds (see below). Regarding the available attributes and in line with what was said above, some participants thought that there were too many unimportant attributes but that at the same time some important attributes were missing. A criticism regarding the comparison phase was that two attributes could not be given the same weight.[9]

Suggested improvements

The participants were asked to give us some ideas how a possible fourth prototype should look like. This question was used to elicit suggestions for the improvement of the actual prototype of IACA. Most recommendations concerned HighF, but this might be due to the general setting of the interview. Many participants recommended a combination of MedF and HighF, with the following improvements. To avoid information overload, some participants proposed to hide most of the sliders and check boxes but to give the possibility to unhide them when needed. As mentioned above, pull-down menus and intervals instead of single

[9]Note that this will be possible in the final application.

points as thresholds were suggested by several participants. It was also regarded as a helpful improvement to provide explanations of the attributes (and the sliders). Moreover, some participants would add the possibility to sort the products by criteria such as price and brand.

Some more fundamental changes to the structure of the system were also considered. One participant had the idea to start directly with the comparison phase. In his prototype, users would be asked to specify acceptance thresholds as soon as they move an attribute into the matrix. Another participant suggested a sequential search where the users start with browsing through very detailed descriptions of the products with one product per page. While browsing, the users can retain some of them for later comparison. A third idea was to start the pre-selection phase with some prototypical examples of products, such as different types of digital cameras (i.e., point and shoot or single-lens reflex). Depending on the selected category, the system interface should then be adapted to the complexity of the selected example (e.g., low functionality for products that are described on few attributes and high functionality for more complex products).

Specific questions

All but one participant (out of 24) understood the meaning of the sliders, check boxes, and buttons, and all of them detected the pictures of the products and the links to the complete descriptions. Each participant who used the comparison phase understood the meaning of the importance factor and the overall values and ranks. Almost all (14, 93%) of the 15 participants who were asked reported to have understood that they could inverse the polarization of the attributes, for example, if a heavy phone was desired instead of a light one. Twelve of 15 participants (80%) would find it helpful to have relative attribute values, which indicate the position of the respective product relative to the other products. This could be realized as a bar underneath each value, which is similar to the battery symbol of a mobile device. For instance, a bar that is filled to a quarter for the weight attribute means that the product's weight is 25% of the maximum weight in the choice set. No participant used this information in the prototype but this was probably because they did not detect this feature. Almost all participants (13 of 15, 87%) saw the three buttons labeled "all," "clear," and "typical", but only 2 of 16 (13%) used them. Some participants stated that they did not need them and some others thought that the buttons would not work (although they have been informed before the experiment that all buttons work). The options "Hide Phone" and "Show All Available Phones" were never used.

6.4 Discussion

The aim of this study was to evaluate the prototype of the InterActive Decision Aid in terms of perceived utility and user satisfaction. To see how well IAPT fared in comparison to already existing decision aids, two more prototypes were created, which were based on real-world consumer websites. The three prototypes differed in the number and type of tools they featured to facilitate choosing, ranging from low (LowF) to medium (MedF) to high functionality (HighF).

Unfortunately, the usability questionnaire did not yield very useful results. The ratings were very similar on all three prototypes and the differences were significant on only one item (i.e., Item 2, degree of complexity). In addition, the Cronbach's alpha values were very high, which suggests that people responded almost identically to all questions and might have ignored the subtle differences between them. Likewise, only on one dimension of the comparison questionnaire (i.e., ease of elimination) significant differences between the three prototypes could be found. Nevertheless, some insights can be gained from the present data, which are discussed now.

Overall, participants slightly preferred the prototype of medium functionality and this prototype was also rated best on the dimensions of understandability and ease of use. HighF (IACA) was rated best regarding the two dimensions of ease of elimination and ease of comparison of alternatives. The low functionality prototype was the least preferred overall. It seems that for my participants, MedF offered the best trade-off between ease of use and functionality.

However, there is evidence that many of the goals that I had in mind when I created IACA were achieved. First, the tools provided for elimination and comparison served their ends, that is, they facilitated these tasks. Although the differences in the ratings were generally quite small and significant only for one dimension (i.e., comparison of alternatives), the means pointed in the expected direction. Second, the alternatives chosen with HighF had a higher overall value than those chosen with MedF. The use of IACA hence resulted in choices that were more in line with the normatively oriented WADD strategy (i.e., the "gold standard," see p. 33). In other words and similar to what was found by Häubl and Trifts (2000), people made better decisions with the new decision aid. Third, almost half of the participants indicated their confidence to be higher when choosing with this prototype than with the other two. This is an indicator that the choice aid increased the quality of the decision, at least subjectively (cf. Häubl & Trifts, 2000; Xiao & Benbasat, 2007). Fourth, all participants stated that the interface helped them to choose the way they like to choose and only one thought that the system altered his habitual way of choosing. The objective to create a decision aid that aims to assist people in their natural way of making a choice has thus been accomplished. Fifth, the vast majority of the participants stated that they would find it helpful to have relative values in addition to the absolute ones. This therefore seemed to be a desired feature and the fact that it was not used is very likely due to usability problems. Sixth, the mean number of attributes that were used for the comparison of alternatives (i.e., 7) was very similar to what was found in Experiments 1 and 2 (i.e., 6 and 7, respectively). This is further evidence for the assumption that people base their choice on only a subset of the available information (cf. Jacoby, Szybillo, & Busato-Schach, 1977) and hence demonstrates that IACA's feature to give the users the ability to decide which attributes to include in the comparison matrix is reasonable. This is also in line with Ariely (2000) who found that people preferred to have a higher level of control over an information system. Moreover, a higher degree of control has been found to increase trust, satisfaction, and the perception of usefulness of the system (McNee, Lam, Konstan, & Riedl, 2003; Pereira, 2000; Wang, 2005). Seventh and finally, during the interview, almost all participants stated that they would use the prototype with the highest functionality if it was a real website. They felt that once they understood how the system works, its features could be very helpful to facilitate the process of choosing. This is particularly interesting given that the decision times were significantly longer with HighF than with the other two prototypes.

On the negative side, however, the use of the HighF prototype was rather cumbersome not only due to its suboptimal usability but also due to its high level of functionality, which might have overshadowed the positive effects of the new tools (cf. Peintner et al., 2008). For instance, the possibility to make a sensitivity analysis has been used by only one participant. Moreover, it has to be acknowledged that the presence of the developer of IACA during the experiment might have biased the results. Although it was attempted to minimize such effects by using a double blind design, the possibility cannot be excluded that participants wanted to please the developer and gave more favorable ratings than they would have without him being present. Likewise, a Pygmalion effect (Rosenthal & Jacobson, 1992) may have occurred, that is, the developer unknowingly influenced the participants during their choices and evaluations in a way that biased the results in the desired direction. However, the fact that three prototypes were evaluated in comparison and without any information about their origin renders such an effect rather improbable. Finally, the way the participants have been recruited could have led to a biased sample that contained in particular people who have much experience with the kind of technology that was being tested, resulting in more favorable results for the higher functionality prototypes. However, during the experiment it became clear that there were quite large inter-individual differences regarding people's ease with the tested technology. It is hence unlikely that the recruitment process led to an unrepresentative sample.

In sum, it appears that consumers appreciate the possibilities provided by the InterActive Choice Aid—at least when making choices between products that require many trade-offs between attributes (e.g., multimedia products). However, it also became clear that a crucial point for the design of a decision aid is its usability. Frequently, systems that offer a high degree of functionality are also more difficult to understand, which, in turn, may discourage potential users. I hence believe that consumers can and will benefit from highly functional decision aids like IACA only when these systems are easy to understand and to use.

Further studies should pay attention to these factors and take into account the improvement suggestions given in the following. If this is done carefully, subsequent evaluations of IACA are likely to result in more significant and also more favorable data.

6.4.1 Suggested improvements of IACA

Leading on from the assumption that the tools featured in IACA are, in principle, desirable, now some suggestions are given here of how both phases of IACA could be improved to enhance the usability of the system.

Improvements of the pre-selection phase

Albeit intuitively appealing, the sliders turned out to be the major usability problem of the pre-selection phase. Even though they are implemented in a very similar form on some real-world consumer websites, they were not appreciated much by the participants. As suggested by some participants, a better solution might be to replace the sliders with pull-down menus where the user can choose one of several intervals (e.g., a price between 100 SFR and 150 SFR). Acceptance intervals could equally be determined in the way they have been implemented in the LowF prototype for the price attribute, where the users wrote their lower and upper limits in separate boxes (see Figure 6.1, p. 92). Moreover, the number of sliders and check boxes had the tendency to confuse the users because they felt that there was too much information on the screen. Therefore, it could be beneficial to display elimination options only for some attributes, similar to how this was done in the MedF prototype. Users who want to base their elimination on more or other attributes can access these via a "more" button or something similar. A more advanced system could even use machine learning algorithms to establish which attributes are shown initially, either generally or individually for each user, depending on his or her past behavior.

A related problem was that many participants had the tendency to set too severe thresholds, which often resulted in an elimination of all alternatives. This could be avoided in part by interpreting the thresholds in a less strict way, in the sense of the Just-Noticeable-Differences mentioned in Part I (see p. 34). In other words, alternatives with an attribute value very close to the threshold set by the user (e.g., within 10%) would be saved from elimination. To visualize these alternatives, they could be highlighted with a color (e.g., red). These JNDs should be adaptable by the user. However, to avoid making the system even more complicated, this adaptation should not be mandatory but only be made upon request from the users. In addition, the possibility to exclude certain alternatives from further elimination should be given to the users. In this way, consumers could retain alternatives they find appealing and they could also compare very dissimilar alternatives, for example, some "should" options with a more fancy "want" option (see Bazerman, Tenbrunsel, & Wade-Benzoni, 1998).

Visually indicating how well each alternative corresponds to the consumer's preferences is another possible improvement. For example, the color of the attribute value could change according to how far away (and in which direction) it is from the threshold set by the user.

A different but probably worthwhile method for threshold setting is to use relative in addition to absolute values (be it with sliders or with other methods). Here, the users would not indicate a maximum or minimum acceptable value but rather the desired relative position of the product, for example, to eliminate all products that are not within the top 25% on the respective attribute.

Further possible improvements are the following:

- Some additional information about the alternatives could be displayed together with product name and picture, similar to how this was done in the LowF prototype. This could be information about attributes that are often requested but also on the attributes that are currently being used for elimination.
- Add boxes next to each product that can be checked to select the respective product for comparison (cf. MedF). This would probably increase understandability because this is a standard component of many current websites.
- Add a reset button to set all thresholds back to their original position.
- Add the possibility of sorting the displayed alternatives by one of several criteria (e.g., price).
- Add links to external review sites.
- To prevent dominated alternatives from being chosen, they could be highlighted (cf. Edwards & Fasolo, 2001). However, for several attributes such as design, there is no general agreement as to what attribute values are desirable because their evaluation is highly subjective. Therefore, dominance in the strict sense cannot be determined. Moreover, the vendors would probably very quickly remove such dominated alternatives.

Although it is very likely that the above suggestions would improve the current system, given the observed problems, it is probably worthwhile to more drastically re-conceptualize the pre-selection phase. One possible approach is the *Teaching Salesman* technique (Stolze & Ströbel, 2004) where consumers are asked about the tasks or purposes they need the product for (e.g., a manager who needs smart phone functions such as mobile internet and push e-mail). The system then proposes products that correspond to the described needs. Thereafter, users can communicate their preferred values on certain attributes and are then provided with three recommendations, a top-score option, an "upgrade" option (more expensive but more features) and an "alternative" option (different features). A similar idea is to first ask the users to assign themselves to one of several categories, such as businessman, mother, or adolescent. Depending on the chosen category, the system could then propose products that are typically used by the people pertaining to the respective category. Given that these systems translate the consumers' needs to attribute specifications, they are particularly useful for users who have a low lever of expertise in the respective product category. Stolze and Nart (2004) show that the combination of such a need-based system with the possibility to specify desirable attribute levels (i.e., feature-based) is preferred over an exclusively feature-based system. Finally, a more applied idea for the further development of IACA is to adapt any advertisements that are displayed on the same page to the thresholds set by the users (e.g., expensive cameras when the maximum acceptable price is high etc.).

Improvements of the comparison phase

A general problem with two-stage strategies where alternatives are eliminated in the first stage is that the eliminated alternatives could have come out as the best alternative if they had not been rejected earlier in the process (Edwards & Fasolo, 2001). A possibility to overcome this problem would be to propose additional alternatives in the comparison phase that are similar to the ones currently being compared but that were eliminated in the first phase.

As already briefly mentioned in Section 5.2.4 (p. 83), another area of improvement concerns the calculation of the overall value. In the present prototype, the normalization of the attribute values is relatively simple (i.e., z-scores). A more sophisticated method that could be used in this context is the vector normalization employed in many MADM methods (e.g., TOPSIS; Hwang & Yoon, 1981), with different normalization procedures for *benefit* attributes (e.g., megapixels of a digital camera) and *cost* attributes (e.g., price). Moreover, the system could elicit which attributes are *nonmonotonic* attributes, that is, attributes where the optimal value does not lie at one of the extremes but somewhere in the middle of the attribute range. For example, for some people the optimal mobile phone may be of "average size." In other words, their utility curve resembles an inverse u-shaped function. On these attributes, still another normalization technique would need to be employed (for details, see Yoon & Hwang, 1995). In addition, the attribute weights could be obtained directly from the ranks (e.g., reciprocal weights or rank sum weights, Edwards & Fasolo, 2001; Yoon & Hwang, 1995). Although with this procedure some information might be lost, in the experiment many participants expressed their attribute weights only on an ordinal scale (i.e., rank order) and therefore did not make use of the possibility to indicate how much more or less important an attribute was in comparison to another attribute by placing it closer or further away from it (i.e., interval scale). It cannot be known based on the present results, however, whether this finding echoes the participants' preference or whether it is simply due to a lack of understanding. Before making such a change to the system, this question should thus be answered empirically.

Further improvements could be:

- Instead of having an empty ICM at the outset, the attributes that have been used for elimination could already be included in the matrix, in the order that the thresholds were set in Phase 1.
- The ability to sort the alternatives in the ICM by attribute values and the overall value could be added.
- The system could highlight differences between the alternatives with colors, for example, green for small differences and red for large differences.
- The absolute attribute values could be highlighted in various colors, which correspond to the relative position of the alternative (e.g., green if the alternative is among the top 25% and red if it is among the bottom 25%).
- The relative attribute values could be represented as a bar underneath the absolute values, as suggested in Section 6.3.4 (p. 104).

Chapter 7

General conclusions

In this dissertation, two new approaches for studying and aiding decision strategies have been proposed and evaluated. The first approach is the method of *InterActive Process Tracing*, which is described in Part I. This method focuses on the description of consumers' choice strategies and is a novel form of combining and thereby complementing several established process tracing techniques. In two experiments it has been shown that the new interactive process tracing method is a valid technique for identifying human decision processes. Various findings in the related literature were replicated and a detailed description of the strategies people used when making a purchase decision was achieved. Similar to Bettman (1970), Larcker and Lessig (1983), Einhorn et al. (1979), and Li et al. (2000), it was shown that models constructed based on verbal reports described the participants' behavior quite well. This also demonstrates that retrospective verbal protocols are a valid measure for the detection of cognitive processes, at least in the context of consumer choices.

A more critical finding that was observed in both experiments is that people's search for information often deviated from what would be expected given the described strategy. 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 is commonly assumed. A possible reason for this is that Mouselab alters the way that information is searched and processed. For instance, Glöckner and Betsch (2008) found that under time pressure participants switched from compensatory to noncompensatory 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 sec on average). The authors suspect that the well-documented switch from compensatory to noncompensatory 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 the search patterns did not differ much between Mouselab and eye tracking makes this explanation a bit less likely, at least for the present experiments. Moreover, in these experiments, the 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 no time pressure was imposed. Thus, I believe that it is sensible to use verbal 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 Phases 1 and 3 are particularly crucial for the detection of cognitive decision processes. However, I 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.

Moreover, an attempt was made to improve IAPT by using eye tracking instead of Mouselab. The rationale for this was that eye tracking is expected to result in more valid data due to the fact that it is less prone to reactivity than Mouselab. However, a direct comparison of these two information search measures showed that the former method did not provide more informative data than the latter, and the use of eye tracking did also not result in a higher validity of the described strategies. In general, the data obtained with both methods was very similar, but the use of eye tracking was more cumbersome and resulted in more noisy data that was more difficult to interpret. I therefore conclude that for the kind of task studied here, Mouselab is a valid and convenient method and the use of eye tracking does not result in any improvements to the method of InterActive Process Tracing.

Despite the fact that so far it has been tested only in one domain (i.e., consumer choice) and with only one product (i.e., mobile phones), I am confident that IAPT is a valuable research tool that is not limited to consumer choice situations and that can be used to answer many different research questions. The attempt to improve the method by integrating eye tracking technology was not successful, but more ways how IAPT could be developed further can be conceived, for instance, by using other process tracing tools such as *Flashlight* (Schulte-Mecklenbeck et al., 2008) (see p. 63).

In conclusion, these 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 that has been described in Part I, the proposed method can also be used for applied purposes. For instance, the findings obtained with IAPT 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 the consumer). Moreover, IAPT can be used for the development of decision aids, such as interactive choice aids that are implemented in consumer websites (e.g., Edwards & Fasolo, 2001; Häubl & Trifts, 2000). 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), and thereby significantly reduce the amount of information that has to be processed. Given the often very large number of products offered by today's online shops, the need of such choice aids is constantly growing, and, not surprisingly, many shopping sites have already implemented a decision aid.

In Part II, such a decision aid was developed, driven by the results found in the two experiments described in Part I and based on already existing real-world decision aids. This decision aid, the *InterActive Choice Aid*, aims to assist the natural process of choosing by providing tools that facilitate the execution of resource-intensive tasks. In particular its comparison feature is a novel and promising development, with which I tried to integrate both considerations from prescriptive decision theory (i.e., how a rational decision maker should choose) as well as insights from descriptive decision research and practical considerations.

The experiment conducted to evaluate a first prototype of IACA showed that these tools, for example for eliminating undesirable products and making in-depth comparisons of alternatives, served their ends, that is, they aided the decision maker during the process of choosing. However, the encountered usability problems make clear that attention has to be devoted not only to the tools and their theoretical basis but also to more practical aspects such as their implementation and the overall usability of the system. Overall, the first test of the new choice aid gave us some promising results and many ideas of how to continue this line of research. Given that many of today's purchases are made on the internet, further developments of IACA are not only of interest to researchers but also to retailers and, in particular, to customers who have to find their way through a plethora of similar products and without the help of a sales assistant. The InterActive Choice Aid might be implemented on shopping websites or on websites that provide guidance to the customer, for instance by featuring reviews and buying guides (e.g., www.consumerreports.org).

However, so far I have never asked the question of whether using a choice aid is actually in the interest or to the benefit of the consumer. The following caveats show that this is actually not a trivial question.

First, there is evidence that although a rational and thorough approach may be deemed desirable by decision makers, people are in fact less satisfied after using a decision aid. That is, although the decision aid enhances objective decision quality, it reduces the confidence in the decision (Abualsamh, Carlin, & McDaniel, 1990; Cats-Baril & Huber, 1987). A potential explanation for this is the fact that the decision aid increases the awareness of decisional conflict by making the trade-offs between attributes more explicit. Thereby, the downsides of the chosen alternative become more salient (cf. Wilson et al., 1993). Moreover, in Ariely's (2000) study a higher control over the displayed information had detrimental effects on the participants' ability to use the information because the demands on the cognitive resources were higher. However, although IACA is not immune to such effects, the calculation feature was introduced to provide (computational) support for resolving such trade-offs. I am therefore confident that this feature attenuates potential negative effects. In fact, in the experiment described here (Experiment 3), nearly half of the participants indicated that they had the highest confidence in their choice when they made their choice with the HighF prototype. Moreover, there is also evidence that decision aids increase the users' satisfaction with their choices (Kmett, Arkes, & Jones, 1999).

A second caveat is the finding that decision makers seem to dislike using models that are based on complex procedures (Zachary, 1988) and often prefer relatively unsophisticated methods (Kottemann & Davis, 1991). This finding should be taken seriously for the design of decision aids, especially when these are targeted at non-professionals. As mentioned above, in the present case the benefits of simplicity may outweigh the costs of a less valid preference elicitation. Moreover, to avoid the possibility that consumers mistrust the decision aid, care should be taken to ensure a high degree of transparency (cf. Section 5.1.3, p. 75).

Third, there is evidence that managers exhibit a certain reluctance to use decision support systems (Finlay, 1994). It is therefore reasonable to ask whether consumers would use similar systems, especially given that in consumer choice the stakes are much lower. However, one reason for the managers' behavior is that they are very often under severe time constraints, which is usually not the case for the consumer who makes only a small number of non-routine purchase decisions per year.

Finally, users might often not have well-defined preferences when they search for a product but rather construct them on the spot as they search for information about the products (cf. Payne et al., 1993). This is particularly likely to happen when consumers are not familiar with the product category in question or when the product category has changed much since their last purchase (e.g., due to technological advances). Similarly, Pu and Chen (2008) argue that people often become aware of their preferences only when they come across unacceptable attribute values. For instance, a certain customer may decide to restrict the consideration set to mobile phones that have a bluetooth connection only when he or she sees an otherwise acceptable mobile phone that does not have this feature. This circumstance could cause problems for both phases of IACA in that consumers are not certain which attributes to use for the elimination and the comparison of alternatives. However, the fact that in the first experiment of this dissertation it was found that consumers were well able to select the attributes they wanted to base their choice on—no matter whether they were given a list containing all available attributes or not—provides some evidence that at least in the domain of mobile phones people are well aware of the attributes they find important. Moreover, the rationale for providing sliders to set the cut-offs was to allow the users to experience attribute ranges in a playful way. By setting and re-setting cut-offs, users can build up a representation of the environment, which ultimately leads to a construction of their preferences. Likewise, Kramer (2007) found that people are more likely to accept top-ranked recommendations when they were based on an explicit and transparent method of preference elicitation as opposed to an implicit preference elicitation method.

Thus, to be accepted and used by a multitude of consumers, a decision aid needs to have benefits that outweigh the perceived costs (cf. Peintner et al., 2008), and it needs to be simple, transparent, and easy to use. In particular, care should be taken to assist the consumers in their natural way of choosing instead of forcing them to follow a supposedly rational or ideal procedure. If these points are respected, I agree with Häubl and Murray (2006) who conclude that "[w]ell-designed electronic product recommendation agents can and should play a more prominent role in improving the overall value of online shopping."

To conclude, both approaches proposed in this dissertation are promising novel developments, one that has a more descriptive purpose and the other that has a more applied purpose. Both make contributions to the existing literature by building on it, extending it, and by pointing to interesting new avenues for research on the topic of decision making.

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.

Abualsamh, R. A., Carlin, B., & McDaniel, R. R. J. (1990). Problem structuring heuristics in strategic decision making. *Organizational Behavior and Human Decision Processes*, *45*, 159–174.

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17*, 734–749.

Aksoy, L., & Bloom, P. N. (2001). *Impact of ordered alternative lists on decision quality: The role of perceived similarity.* Paper presented at th Society for Consumer Research Winter Conference, Scottsdale, AZ, February 15–17, 2001.

Alba, J., Lynch, J., Weitz, B., Janiszewski, R. L., Sawyer, A., & Wood, S. (1997). Interactive home shopping: consumer, retailer, and manufacturer incentives to participate in electronic marketplaces. *Journal of Marketing*, *61*, 38–53.

Allais, P. M. (1953). The behavior of rational man in risk situations—A critique of the axioms and postulates of the American School. *Econometrica*, *21*, 503-546.

Al-Natour, S., Benbasat, I., & Cenfetelli, R. T. (2008). The effect of process and outcome similarity on user's evaluations of decision aids. *Decision Sciences*, *39*, 175–211.

Arch, D. C., Bettman, J. R., & Pakkar, P. (1978). Subjects' information processing in information display board studies. *Advances in Consumer Research*, *5*, 555–560.

Ariely, D. (2000). Controlling the information flow: Effects on consumers' decision making and preferences. *Journal of Consumer Research*, *27*, 233–248.

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.

Baron, J. (2000). *Thinking and deciding* (3rd ed.). Cambridge: Cambridge University Press.

Bazerman, M. H., Tenbrunsel, A. E., & Wade-Benzoni, K. (1998). Negotiating with yourself and losing: Understanding and managing competing internal preferences. *Academy of Management Review*, *23*, 225–241.

Bettman, J. R. (1970). Information processing models of consumer behavior. *Journal of Marketing Research*, *7*, 370–376.

Bettman, J. R. (1979). *An information processing theory of consumer choice.* Reading, MA: Addison Wesley.

Biggs, S. F., Bedard, J. C., Gaber, B. G., & Linsmeier, T. J. (1985). The effects of task size and similarity on the decision behavior of bank loan officers. *Management Science*, *31*, 970–987.

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.

Borcherding, K., Eppel, T., & von Winterfeldt, D. (1991). Comparison of weighting judgments in multiattribute utility measurement. *Management Science*, *37*, 1603–1619.

Borcherding, K., Schmeer, S., & Weber, M. (1995). Biases in multiattribute weight elicitation. In J.-P. Caverni, M. Bar-Hillel, F. H. Barron, & H. Jungermann (Eds.), *Contributions to decision-making I* (pp. 1–28). Amsterdam: North-Holland.

Bröder, A. (2000). A methodological comment on behavioral decision research. *Psychologische Beiträge*, *42*, 645–662.

Brucks, M. (1988). Search Monitor: An approach for computer-controlled experiments involving consumer information search. *Journal of Consumer Research*, *15*, 117–121.

Cats-Baril, W. L., & Huber, G. P. (1987). Decision support systems for ill-structured problems: An empirical study. *Decision Sciences*, *18*, 350–372.

Chu, P. C., & Spires, E. S. (2000). The joint effects of effort and quality on decision strategy choice with computerized decision aids. *Decision Sciences*, *31*, 259–284.

Cook, G. J., & Swain, M. R. (1993). A computerized approach to decision process tracing for decision support system design. *Decision Sciences*, *24*, 931–952.

Darwin, C. R. (1958). *The autobiography of Charles Darwin 1809-1882. With the original omissions restored. Edited and with appendix and notes by his grand-daughter Nora Barlow* (N. Barlow, Ed.). London: Collins.

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.

Dhar, R. (1997). Consumer preference for a no-choice option. *Journal of Consumer Research*, *24*, 215–231.

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.

Eisenführ, F., & Weber, M. (1994). *Rationales Entscheiden* (2nd ed.). Berlin: Springer.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, *75*, 643-669.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Fasolo, B., & McClelland, G. H. (1999). *Tracing decision processes on the web.* Poster presented at the annual meeting of the Society for Judgment and Decision Making in Los Angeles, U.S.A.

Favier, J., & Bouquet, M. (2006). *Europe's eCommerce Forecast: 2006 to 2011.* Technographics Research. Retrieved May 26th, 2009, from: http://www.forrester.com/Research/Document/Excerpt/0,7211,38297,00.html.

Finlay, P. (1994). *Introducing decision support systems.* Cambridge, MA: Blackwell.

Ford, J. K., Schmitt, N., Schlechtman, S. L., Hults, 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 one cue is not enough: Combining fast and frugal heuristics with compound cue processing. *Quarterly Journal of Experimental Psychology*, *60*, 1197-1215.

Gensch, D. H. (1987). A two-stage disaggregate attribute choice model. *Marketing Science*, *6*, 223-231.

Gigerenzer, G. (2007). Why heuristics work. *Perspectives on Psychological Science*, *3*, 20–29.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The Take The Best heuristic. In G. Gigerenzer & P. M. Todd (Eds.), *Simple heuristics that make us smart* (p. 75-95). New York: Oxford University Press.

Gigerenzer, G., Hoffrage, U., & Goldstein, D. G. (2008). Fast and frugal heuristics are plausible models of cognition: Reply to Dougherty, Franco-Watkins, and Thomas (2008). *Psychological Review*, *115*, 230–239.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Glöckner, A., & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1055-1075.

Grady, H. M. (2000). Web site design: A case study in usability testing using paper prototypes. In *Proceedings of 2000 Joint IEEE International and 18th Annual Conference on Computer Documentation (IPCC/SIGDOC 2000)* (pp. 39–45). Piscataway, NJ: IEEE Educational Activities Department.

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., & Murray, K. B. (2006). Double agents: Assessing the role of electronic product-recommendation systems. *MIT Sloan Management Review*, *47*, 8-12.

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.

Hill, D. J., & King, M. F. (1989). Preserving consumer autonomy in an interactive informational environment: Toward developmentof a consumer decision aid model. *Advances in Consumer Research*, *16*, 144–151.

Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, *114*, 733–758.

Holsapple, C. W. (2008). DSS architecture and types. In F. Burstein & C. W. Holsapple (Eds.), *Handbook on decision support systems 1* (p. 163-189). Berlin/Heidelberg: Springer Verlag.

Hostler, R. E., Yoon, V. Y., & Guimaraes, T. (2005). Assessing the impact of internet agent on end users' performance. *Decision Support Systems*, *41*, 313–325.

Huber, O. (1982). *Entscheiden als Problemlösen.* Bern: Huber.

Huber, O., & Huber, O. W. (2003). *What makes gambles special? - Active search for probability and risk defusing information in gambles and quasi realistic risky tasks.* Unpublished manuscript, University of Fribourg, Switzerland.

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.

Hwang, C. L., & Yoon, K. P. (1981). *Multiple attribute decision making: Methods and applications.* Berlin/Heidelberg/New York: Springer Verlag.

Işıklar, G., & Büyüközkan, G. (2007). Using a multi-criteria decision maing approach to evaluate moblie phone alternatives. *Computer Standards and Interfaces*, *29*, 265–274.

Iyengar, S. S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*, 995–1006.

Jacoby, J., Szybillo, G. J., & Busato-Schach, J. (1977). Information acquisistion behavior in brand choice situations. *Journal of Consumer Research*, *11*, 708–718.

Janis, I., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment.* New York: Free Press.

Jedetski, J., Adelman, L., & Yeo, C. (2002). How web site decision technology affects consumers. *IEEE Internet Computing*, *6*, 72–79.

Johnson, R. M. (1985). Adaptive conjoint analysis. In *Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis and Computer Interviewing* (pp. 253–265). Ketchum, ID: Sawtooth Software Inc.

Johnson, R. M. (1991). Adaptive conjoint analsysis: Some caveats and suggestions. *Journal of Marketing Research*, *28*, 223–225.

Jungermann, H., Pfister, H., & Fischer, K. (1998). *Die Psychologie der Entscheidung.* Berlin: Spektrum.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263-291.

Karelaia, N. (2006). Thirst for confirmation in multi-attribute choice: Does search for consistency impair decision performance? *Organizational Behavior and Human Decision Processes*, *100*, 128–143.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preference and value tradeoffs.* New York: Wiley.

Kmett, C. M., Arkes, H. R., & Jones, S. K. (1999). The influence of decision aids on high school students' satisfaction with their college choice decision. *Personality and Social Psychology Bulletin*, *25*, 1293–1301.

Kottemann, J. E., & Davis, F. D. (1991). Decisional conflict and user acceptance of multicriteria decision-making aids. *Decision Sciences*, *22*, 918–926.

Kramer, T. (2007). The effect of measurement task transparency on preference construction and evaluations of personalized recommendations. *Journal of Marketing Research*, *44*, 224–233.

Kühberger, A. (1994). Risiko und Unsicherheit: Zum Nutzen des Subjective Expected Utility-Modells. *Psychologische Rundschau*, *45*, 3-23.

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.

Lewis, J. R. (1995). IBM computer satisfaction usability questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, *7*, 57–78.

Li, S., Shue, L., & Shiue, W. (2000). The development of a decision model for liquidity analysis. *Expert Systems with Applications*, *19*, 271–278.

Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, *7*, 76–80.

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.

Lopes, L. L. (1990). Discussion on a chapter by Mano and Payne, Bettman, and Johnson. In R. M. Hogarth (Ed.), *Insights in decision-making: A tribute to Hillel J. Einhorn.* Chicago: University of Chicago Press.

Martin, J. M., & Norton, M. I. (2008). *Shaping product attribute valuation and consumer choice by partitioning the web* (Tech. Rep.). Boston, MA: Harvard Business School.

McNee, S., Lam, S. K., Konstan, J. A., & Riedl, J. (2003). Interfaces for eliciting new user preferences in recommender systems. In *Proceedings of the 9th International Conference on User Modeling* (pp. 178–187). Johnston, PA, USA.

Miles, G. E., Howes, A., & Davies, A. (2000). A framework for understanding human factors in web-based electronic commerce. *International Journal of Human-Computer Studies*, *52*, 131–163.

Montgomery, A. L., Hosanagar, K., Krishnan, R., & Clay, K. B. (2004). Designing a better shopbot. *Management Science*, *50*, 189–206.

Montgomery, H., & Svenson, O. (1976). On decision rules and information processing strategies in multiattribute decision making. *Scandinavian Journal of Psychology*, *17*, 283-291.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, N.J.: Prentice-Hall.

Nielsen, J. (1990). Paper versus computer implementations as mockup scenarios for heuristic evaluation. In *Proceedings of the IFIP TC13 Third International Conference on Human-Computer Interaction, August 27-31, 1990, Cambridge, UK* (pp. 315–320). Amsterdam, The Netherlands: North-Holland.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than you can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Nixon, H. K. (1924). Attnetion and interest in advertising. *Archives of Psychology*, *72*, 5–67.

O'Keefe, R. M., & McEachern, T. (1998). Web-based customer decision support systems. *Communications of the ACM*, *41*, 71–78.

Olshavsky, R. W. (1979). Task complexity and contingent processing in decision making: A replication and extension. *Organizational Behavior and Human Performance*, *24*, 300–316.

Olson, D. L. (2008). Multi-criteria decision support. In F. Burstein & C. W. Holsapple (Eds.), *Handbook on decision support systems 1* (pp. 299–313). Berlin/Heidelberg: Springer Verlag.

Olson, E. L., & Widing, R. E. (2002). Are interactive decision aids better than passive decision aids? A comparison with implications for information providers on the internet. *Journal of Interactive Marketing*, *16*, 22–33.

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.

Pedersen, P. (2000). Behavioral effects of using software agents for product and merchant brokering: An experimental study of consumer desicion making. *International Journal of Electronic Commerce*, *5*, 125–141.

Peintner, B., Viappiani, P., & Yorke-Smith, N. (2008). Preferences in interactive systems: Technical challenges and case studies. *AI Magazine*, *29*, 13–24.

Pereira, R. E. (2000). Optimizing human-computer interaction for the electronic commerce environment. *Journal of Electronic Commerce Research*, *1*, 23–44.

Pinker, S. (1997). *How the mind works.* New York: Norton.

Plous, S. (1993). *The psychology of judgment and decision making.* New York: Mc-Graw Hill.

Pu, P., & Chen, L. (2008). User-involved preference elicitation for product search and recommender systems. *AI Magazine*, *29*, 93–103.

Ramaprasad, A. (1987). Cognitive process as a basis for MIS and DSS design. *Management Science*, *33*, 139–148.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Reisen, N., Hoffrage, U., & Mast, F. W. (2008). Identifying decision strategies in a consumer choice situation. *Judgment and Decision Making*, *3*(8), 641–658.

Riedl, R., Brandstätter, E., & Roithmayr, F. (2008). Identifying decision strategies: A process- and outcome-based classification 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, & the ABC Research Group (Eds.), *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.

Rosenthal, R., & Jacobson, L. (1992). *Pygmalion in the classroom* (expanded edition). New York: Irvington.

Rubin, J. (1994). *The handbook of usability testing.* New York: Wiley.

Rudd, J., Stern, K., & Isensee, S. (1996). Low vs. high-fidelity prototyping debate. *Interactions*, *3*, 76–85.

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., & Dosher, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 676–696.

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.

Saaty, T. L. (1980). *The analytical hierarchial process.* New York: Wiley.

Sauer, J., & Sonderegger, A. (in press). The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion. *Applied Ergonomics.*

Schulte-Mecklenbeck, M., Murphy, R. O., & Hutzler, F. (2008). Flashlight. *Annual conference of the Society for Judgment and Decision Making, November 14–17, Chicago, IL*, 14.

Seidl, C. (2002). Preference reversal. *Journal of Economic Surveys*, *16*(5), 621–655.

Senecal, S., & Nantel, J. (2004). The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, *80*, 159–169.

Seref, M. M. H., & Ahuja, R. K. (2008). Spreadsheet-based decision support systems. In F. Burstein & C. W. Holsapple (Eds.), *Handbook on decision support systems 1* (pp. 277–298). Berlin/Heidelberg: Springer Verlag.

Shah, K. S., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*, 207–222.

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

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.

Simon, H. A. (1987). Rational decision making in business organizations. In L. Green & J. H. Kagel (Eds.), *Advances in behavioral economics* (Vol. 1, pp. 18–47). Norwood, NJ: Ablex.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*, 1–19.

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.

Stolze, M., & Nart, F. (2004). Well-integrated needs-oriented recommender components regarded as helpful. In *CHI 2004 Extended Abstracts on Human Factors in Computing Systems* (p. 1571). Vienna, Austria, April 24–29.

Stolze, M., & Ströbel, M. (2004). Recommending as personalized teaching. In *Designing personalized user experiences in eCommerce* (pp. 293–313). Norwell, MA: Kluwer Academic Publishers.

Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance*, *23*, 86–112.

Todd, P. M., & Benbasat, I. (1991). An experimental investigation of the impact of computer based decision aids on decision making strategies. *Information Systems Research*, *2*, 87–115.

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

Todd, P. M., & Benbasat, I. (1994a). The influence of decision aids on choice strategies: An experimental analysis of the role of cognitive effort. *Organizational Behavior and Human Decision Processes*, *60*, 36–74.

Todd, P. M., & Benbasat, I. (1994b). The influence of decision aids on choice strategies under conditions of high cognitive load. *IEEE Transactions on Systems, Man, and Cybernetics*, *24*, 537–547.

Todd, P. M., & Benbasat, I. (2000). Inducing compensatory information processing through decision aids that facilitate effort reduction: An experimental assessment. *Journal of Behavioral Decision Making*, *13*, 91–106.

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

Tversky, A. (1974). Judgment und uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-458.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *57*, 297-323.

van Gog, T., Paas, F., van Merriënboer, J. J. G., & 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.

Vijayasarathy, L. R., & Jones, J. M. (2001). Do internet shopping aids make a difference? An empirical investigation. *Electronic Markets*, *11*, 75–83.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Wang, W. (2005). *Design of trustworthy online recommendation agents: Explanation facilities and* *decision strategy support.* Unpublished doctoral dissertation, University of British Columia, Vancouver, Canada.

Weber, E. U., Baron, J., & Loomes, G. (2001). *Conflict and tradeoffs in decision making.* Cambridge: Cambridge University Press.

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.

Westenberg, M. R. M., & Koele, P. (1994). Multi-attribute evaluation processes: Methodological and conceptual issues. *Acta Psychologica*, *87*, 65–84.

White, C. M., & Hoffrage, U. (2009). Testing the tyranny of too much choice against the allure of more choice. *Psychology and Marketing*, *26*, 280–298.

White, C. M., Reisen, N., & Hoffrage, U. (2009). Choice deferral arising from absolute evaluation or relative comparison. *Submitted for publication.*

Willemsen, M. C., & Johnson, E. J. (2006). *Mouselabweb: Monitoring information acquisition processes on the web.* Retrieved October 14, 2008, from http://www.mouselabweg.org.

Williamson, J., Ranyard, R., & Cuthbert, L. (2000). A conversation-based process tracing method foruse with naturalistic decisions: An evaluation study. *British Journal of Psychology*, *91*, 203–221.

Wilson, T. D., Lisle, D. J., Schooler, J., Hodges, S. D., Klaaren, K. J., & LaFleur, S. J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin*, *60*, 181–192.

Wright, P., & Barbour, F. (1977). *Phased decision strategies: Sequels to an initial screening.* (Research Paper No. 353). Stanford, CA: Standford University, Graduate School of Business.

Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly*, *31*, 137–209.

Yoon, K. P., & Hwang, C. L. (1995). *Multiple attribute decision making: An introduction.* Sage University Paper series on Quantitative Applications in the Social Sciences, 07-104. Thousand Oaks, CA: Sage.

Zachary, W. W. (1988). Decision support systems: Designing to extend the cognitive limits. In M. Helander (Ed.), *Handbook of human-computer interaction.* Amsterdam, The Netherlands: Elsevier.

Zack, M. H. (1993). Interactivity and communication mode choice in ongoing management groups. *Information Systems Research*, *4*, 207–239.

Appendix A

Questionnaire items

A.1 Usability questionnaire

The questionnaire's instructions and items are as follows:

This questionnaire gives you an opportunity to tell us our reactions to the system you used. Your responses will help us understand what aspects of the system you are particularly concerned about and the aspects that satisfy you. To as great degree as possible, think about all the tasks that you have done with the system while you answer these questions. Please read each statement and indicate how strongly you agree or disagree with the statement by circling a number on the scale. If a statement does not apply to you, circle N/A. Please write comments to elaborate on your answers. Thank you!

The items were presented in the following format:

1. Overall, I am satisfied with how easy it is to use this system.

strongly disagree							strongly agree	
←	– – –	– – –	– – –	– – –	– – –	– – –	→	
	1	2	3	4	5	6	7	N/A

Comments:

Items:

1. Overall, I am satisfied with how easy it is to use this system.
2. The system was too simple (i.e., too rigid) vs. the system was too complex (i.e., too many functions/steps).*‡
3. I was able to make my choice quickly using this system.
4. I was able to efficiently choose a phone/camera using this system.
5. I felt comfortable using this system.
6. It was easy to learn to use this system.
7. I would have appreciated getting more error messages when I ran into problems.*
8. Whenever I made a mistake using the system, I could recover easily and quickly.
9. The information (such as on-line help, on-screen messages, and other documentation) provided with this system was enough.
10. It was easy to find the information about the phones/cameras that I needed.
11. The organization of information on the system screens was clear.*
12. The interface of this system was pleasant.

Note. The interface includes those items that you use to interact with the system. For example, some components of the interface are the keyboard, the mouse, and the screens (including their use of graphics and their language).

13. I liked using the interface of this system.
14. This system has all the functions and capabilities I expect to have.
15. Overall, I am satisfied with the system.
16. I am confident that I made a good choice, that is, I chose a phone/camera that will satisfy me.†
17. It was easy to find the product I want.†

---

*Item adapted to the task

† Additional item

‡ Scale from –3 (“too simple) to +3 (“too complex)

A.2 Comparison questionnaire

The questionnaire's instructions and items are as follows:

This questionnaire gives you an opportunity to tell us how you rate the three prototypes you just used in comparison to each other. For each of the five dimensions, locate each prototpye (with the corresponding number) on the line between "low" and "high." You can also add comments, if you wish.

The items were presented in the following format:

1. Understandability (i.e., it was easy to understand how the system works)

   low ← − − − − − − − − − − − − − − − − − − → high

   Comments:

Items:

1. Understandability (i.e., it was easy to understand how the system works)
2. Ease of use (the system was easy to use)
3. Ease of reducing the full set of products to a subset of particularly interesting products (i.e., I could quickly exclude unattractive phones/cameras from my choice set)
4. Ease of comparison of products (i.e., it was easy to compare the different phones/cameras to each other)
5. Confidence that I have made a good choice (i.e., that I chose a phone/camera that will satisfy me).

A.3 Semi-structured interview

Some or all of the following questions were asked, the order was not determined. The items were the following.

General questions:

1. Would you use Prototype X (say number) if it existed? (yes/no)
   (a) How frequently?
   (b) For which products?
   (c) Would you pay to use this service? (yes/no)
2. Would you make use of the possibility to create own attributes? (yes/no)
3. Would you like to have an attribute, which represents the rating by a known review site/journal? (yes/no)
4. Did the interface help you to choose the way you like to choose? (yes/no)
5. Did the way you normally choose change due to the architecture of the system? (yes/no)
6. Do you think you detected and explored all possibilites of the system, or would you benefit from the instruction of an expert? (On a scale from 1 – 7, where 1 is much profit from an instruction and 7 is few profit from an instruction)
7. How complex do you find the system on a scale from 1 – 7, where 1 is not complex at all and 7 is very complex?
8. Would you recommend Prototype X (say number) it to others? If yes, to whom? (yes/no)
9. Do you think that most people would learn to use this system very quickly? (yes/no)
10. What problems do you see with this system?
11. Based on what you have seen, how should a fourth prototype look like? In particular, if you were to construct a new prototype, would you use one of the three prototypes as a starting point or would you rather start from scratch?

Specific questions:

1. Did you understand the meaning of the rulers/buttons? (yes/no)

2. Did you detect the pictures of the products in the comments? (yes/no)

3. Did you detect the links to the complete descriptions of the products? (yes/no)

4. Did you understand the meaning of the importance factor? (yes/no)

5. Did you understand the meaning of the "negative attributes"? (yes/no)

6. Did you understand the meaning of the overall values/ranks? (yes/no)

7. Did you (1) use the relative information and did you (2) find it helpful? (yes/no)

8. Did you (1) see the buttons labeled all / clear / typical? Did you (2) use them?(yes/no)

Appendix B

Stimuli

B.1 Experiments 1 & 2

Table B.1: Stimuli used in Experiments 1 & 2

		Mobile Phones			
	Attribute	1	2	3	4
1a	Image*	see Figure B.3			
1b	Phone Name†	Samsung SGH-D357	Samsung SGH-Z500	Samsung SPH-A940	Samsung SGH-T309
2	Band Support	tri-band	tri-band	dual-band	tri-band
3	Bluetooth	YES	YES	YES	NO
4	Color	White	Black or silver	Silver	Black
5	Connectivity (USB)	NO	YES	NO	NO
6	Email Support (POP3+IMAP4)	YES	YES	YES	YES
7	External LCD Display	96x64, 4 colors	80x64, 65k colors	96x96, 65k colors	96x96, 4 colors
8	External Memory	NO	up to 256 MB	up to 256 MB	NO
9	GPRS	YES	YES	YES	YES
10	Infrared	NO	YES	NO	NO
11	Integrated Digital Camera	NO	1 Mpixel	1.3 Mpixel	0.3 Mpixel
12	Integrated Speakerphone	YES	YES	YES	YES
13	Internal Display	128x160, 65k colors	176x220, 262k colors	176x220, 262k colors	128x160, 65k colors
14	Internal Memory	NO	50 MB	60 MB	NO
15	Java Games	YES	YES	YES	YES
16	MMS (Photos+Text+Sound)	YES	YES	YES	YES
17	MP3 player	NO	YES	YES	YES
18	Office©	NO	NO	NO	NO
19	Price	165 €	300 €	330 €	150 €
20	Size	86x45x25 mm	89x44.5x24.5 mm	93x46x26 mm	94x48x25 mm
21	SMS	YES	YES	YES	YES
22	Speech Recording	NO	YES	NO	NO
23	Stand-by Time	200h	260h	293h	216h
24	Stereo FM	NO	NO	NO	NO
25	Talk Time	5h	3.4h	4h	4.5h
26	UMTS(3G) Compatibility	NO	YES	NO	NO
27	VibraCall Alert	YES	YES	YES	YES
28	Video Playback with Sound	NO	YES	YES	YES
29	Video Streaming	NO	NO	NO	NO
30	Voice Command	NO	NO	YES	YES
31	WAP	YES	YES	YES	YES
32	Weight	100g	95g	139g	85g
33	Windows© Mobile	NO	NO	NO	NO

* used only in Experiment 1

table continued on next page

† used only in Experiment 2

Table B.1—*Continued*

		Mobile Phones						
	Attribute	5	6	7	8	9	10	11
1a	Image*	see Figure B.3						
1b	Phone Name†	Samsung SGH-D307	Samsung SGH-E635	Samsung SGH-X495	Samsung SGH-T809	Samsung SCH-A850	Samsung SCH-I730	Samsung SCH-A630
2	Band Support	tri-band	tri-band	tri-band	quad-band	dual-band	dual-band	dual-band
3	Bluetooth	YES	NO	NO	YES	NO	YES	NO
4	Color	Silver	Black or silver	Silver	Black	Black or silver	Black	Black or silver
5	Connectivity (USB)	NO	NO	NO	NO	NO	YES	NO
6	Email Support (POP3+IMAP4)	YES	NO	YES	YES	YES	YES	YES
7	External LCD Display	96x96, 2 colors	NO	2 colors	NO	96x96, 2 colors	NO	NO
8	External Memory	NO	NO	NO	up to 512 MB	NO	up to 128 MB	NO
9	GPRS	YES	YES	YES	YES	YES	YES	YES
10	Infrared	NO	YES	NO	NO	NO	YES	NO
11	Integrated Digital Camera	NO	0.3 Mpixel	NO	1.3 Mpixel	0.3 Mpixel	NO	NO
12	Integrated Speakerphone	YES	YES	YES	YES	YES	YES	YES
13	Internal Display	176x220, 262k colors	128 x160, 65k colors	128x160, 65k colors	240x320, 262k colors	128x160, 65k colors	240x320, 65k colors	128x160, 65k colors
14	Internal Memory	NO	9 MB	NO	70 MB	NO	128 MB	NO
15	Java Games	YES	YES	YES	YES	NO	YES	NO
16	MMS (Photos+Text+Sound)	YES	YES	YES	YES	YES	YES	YES
17	MP3 player	NO	NO	NO	YES	NO	YES	NO
18	Office©	NO	NO	NO	NO	NO	YES	NO
19	Price	160 €	125 €	167 €	250 €	210 €	500 €	170 €
20	Size	83x45x20 mm	85x43x22 mm	90x47x24 mm	99x51x15 mm	84x46x25 mm	114x58x25 mm	90x47x26 mm
21	SMS	YES	YES	YES	YES	YES	YES	YES
22	Speech Recording	NO	NO	NO	NO	NO	NO	NO
23	Stand-by Time	216h	240h	200h	130h	170h	130h	200h
24	Stereo FM	NO	NO	NO	NO	NO	NO	NO
25	Talk Time	5h	4h	2.5h	4.3h	3.3h	2.2h	3.3h
26	UMTS(3G) Compatibility	NO	NO	NO	NO	NO	NO	NO
27	VibraCall Alert	YES	YES	YES	YES	YES	YES	YES
28	Video Playback with Sound	NO	NO	NO	NO	NO	YES	NO
29	Video Streaming	NO	NO	NO	NO	NO	NO	NO
30	Voice Command	YES	YES	YES	NO	YES	NO	YES
31	WAP	YES	YES	YES	YES	YES	YES	YES
32	Weight	122g	79g	86g	95g	108g	183g	102g
33	Windows© Mobile	NO	NO	NO	NO	NO	YES	NO

* used only in Experiment 1
† used only in Experiment 2

table continued on next page

Table B.1—*Continued*

		Mobile Phones						
	Attribute	12	13	14	15	16	17	18
1a	Image*	see Figure B.3						
1b	Phone Name†	Motorola MotoPebl	Motorola L6	Motorola MPX220	Motorola V3i	Motorola SLVR	Motorola V710	Motorola V600
2	Band Support	quad-band	tri-band	quad-band	quad-band	quad-band	quad-band	quad-band
3	Bluetooth	YES	YES	YES	YES	YES	YES	NO
4	Color	Black	Silver	Black or silver	Silver	Black	Chrome	Silver
5	Connectivity (USB)	YES	YES	YES	YES	YES	NO	YES
6	Email Support (POP3+IMAP4)	YES	NO	YES	YES	YES	NO	YES
7	External LCD Display	96x32, 2 colors	NO	98x64, 4k colors	96x80, 65k colors	NO	NO	NO
8	External Memory	NO	NO	up to 1 GB	up to 256 MB	up to 256 MB	up to 256 MB	NO
9	GPRS	YES	YES	YES	YES	YES	NO	YES
10	Infrared	NO	NO	YES	NO	NO	NO	NO
11	Integrated Digital Camera	0.3 Mpixel	0.3 Mpixel	1.2 Mpixel	1.23 Mpixel	0.3 Mpixel	1 Mpixel	1 Mpixel
12	Integrated Speakerphone	YES	NO	YES	YES	YES	YES	YES
13	Internal Display	176x220, 262k colors	128x160, 65k colors	176x220, 65k colors	176x220, 262k colors	176x220, 262k colors	176x220, 262k colors	176x220, 65k colors
14	Internal Memory	5 MB	NO	32 MB	12 MB	5 MB	NO	5 MB
15	Java Games	YES	YES	YES	YES	NO	NO	YES
16	MMS (Photos+Text+Sound)	YES	YES	YES	YES	YES	YES	YES
17	MP3 player	YES	YES	YES	YES	YES	YES	NO
18	Office©	NO	NO	NO	NO	NO	NO	NO
19	Price	250 €	155 €	207 €	360 €	166 €	130 €	207 €
20	Size	87x49x20 mm	113x49x10.9 mm	100x48x24.3 mm	98x53x13 mm	113.5x49x11.5mm	94x50x23 mm	94x50x23 mm
21	SMS	YES	YES	YES	YES	YES	YES	YES
22	Speech Recording	YES	NO	YES	NO	NO	YES	YES
23	Stand-by Time	250h	345h	260h	310h	350h	165h	185h
24	Stereo FM	NO	NO	NO	NO	NO	NO	NO
25	Talk Time	6.5h	6h	4.5h	7h	6h	3h	3h
26	UMTS(3G) Compatibility	NO	NO	NO	NO	NO	NO	NO
27	VibraCall Alert	YES	YES	YES	YES	YES	YES	YES
28	Video Playback with Sound	YES	YES	YES	YES	NO	YES	YES
29	Video Streaming	NO	NO	NO	NO	NO	NO	NO
30	Voice Command	YES	NO	YES	NO	YES	NO	NO
31	WAP	YES	YES	NO	YES	NO	YES	YES
32	Weight	110g	86g	110g	100g	96g	127g	116g
33	Windows© Mobile	NO	NO	YES	NO	NO	NO	NO

* used only in Experiment 1

† used only in Experiment 2

table continued on next page

Table B.1—*Continued*

		Mobile Phones						
	Attribute	19	20	21	22	23	24	25
1a	Image*	see Figure B.3						
1b	Phone Name†	Motorola ROKR	Motorola E815	Motorola E310	Motorola V65p	Motorola V171	Motorola V188	Motorola V400
2	Band Support	tri-band	quad-band	quad-band	quad-band	dual-band	quad-band	quad-band
3	Bluetooth	YES	YES	NO	NO	NO	NO	NO
4	Color	Silver	Silver	Black	Silver	Silver	Black	Silver
5	Connectivity (USB)	NO	NO	NO	NO	NO	NO	YES
6	Email Support (POP3+IMAP4)	YES	NO	NO	NO	NO	NO	YES
7	External LCD Display	NO	96x64, 4k colors	NO	96x32, 2 colors	NO	96x32, 2 colors	96x32, 2 colors
8	External Memory	up to 256 MB	up to 256 MB	NO	NO	NO	NO	NO
9	GPRS	YES	NO	NO	NO	YES	YES	YES
10	Infrared	NO	NO	NO	NO	NO	NO	NO
11	Integrated Digital Camera	1 Mpixel	1.3 Mpixel	NO	NO	NO	NO	0.3 Mpixel
12	Integrated Speakerphone	YES	YES	YES	NO	NO	YES	YES
13	Internal Display	176x220, 262k colors	176x220, 262k colors	128x128, 4k colors	128x128, 4k colors	96x64, 4k colors	128x128, 65k colors	176x220, 65k colors
14	Internal Memory	NO	NO	NO	NO	NO	NO	NO
15	Java Games	YES	YES	YES	YES	YES	NO	YES
16	MMS (Photos+Text+Sound)	YES	YES	YES	YES	YES	YES	YES
17	MP3 player	YES	NO	NO	YES	NO	NO	YES
18	Office©	NO	NO	NO	NO	NO	NO	NO
19	Price	330 €	160 €	50 €	166 €	90 €	108 €	130 €
20	Size	108x46x20.5 mm	94x49x24 mm	94x50x23 mm	87x45x32 mm	90x48x23 mm	86x45x23 mm	94x50x23 mm
21	SMS	YES	YES	YES	YES	YES	YES	YES
22	Speech Recording	NO	YES	NO	NO	NO	NO	YES
23	Stand-by Time	185h	120h	200h	78h	200h	300h	200h
24	Stereo FM	NO	NO	NO	NO	NO	NO	NO
25	Talk Time	5h	3h	2h	3h	5h	5h	3h
26	UMTS(3G) Compatibility	NO	NO	NO	NO	NO	NO	NO
27	VibraCall Alert	YES	YES	YES	YES	YES	YES	YES
28	Video Playback with Sound	YES	NO	NO	NO	NO	NO	NO
29	Video Streaming	NO	NO	NO	NO	NO	NO	NO
30	Voice Command	NO	NO	NO	YES	NO	NO	NO
31	WAP	YES	YES	NO	NO	YES	YES	YES
32	Weight	170g	115g	127g	138g	87g	80g	120g
33	Windows© Mobile	NO	NO	NO	NO	NO	NO	NO

* used only in Experiment 1

† used only in Experiment 2

table continued on next page

Table B.1—*Continued*

		Mobile Phones					
	Attribute	26	27	28	29	30	31
1a	Image*	see Figure B.3					
1b	Phone Name†	Sony S700i	Sony W600	Sony Z520a	Sony W800	Sony J300a	Sony T290a
2	Band Support	tri-band	quad-band	quad-band	tri-band	NO	dual-band
3	Bluetooth	YES	YES	YES	YES	NO	NO
4	Color	Black	Blue, white or orange	White, grey or blue	White	Pink, white or grey	White or black
5	Connectivity (USB)	YES	YES	YES	YES	YES	NO
6	Email Support (POP3+IMAP4)	YES	YES	YES	YES	YES	YES
7	External LCD Display	NO	NO	101x80, 4k colors	NO	NO	NO
8	External Memory	128 MB	NO	NO	up to 1 GB	NO	NO
9	GPRS	YES	YES	YES	YES	YES	YES
10	Infrared	YES	YES	YES	YES	NO	NO
11	Integrated Digital Camera	1.3 Mpixel	1.3 Mpixel	0.3 Mpixel	2 Mpixel	NO	NO
12	Integrated Speakerphone	YES	YES	YES	YES	YES	NO
13	Internal Display	240x320, 262k colors	176x220, 262k colors	128x160, 65k colors	176x220, 262k colors	128x128 pixel, 65k colors	101x80 pixel, 4k colors
14	Internal Memory	32 MB	256 MB	16 MB	34 MB	12 MB	NO
15	Java Games	YES	YES	YES	YES	YES	NO
16	MMS (Photos+Text+Sound)	YES	YES	YES	YES	YES	YES
17	MP3 player	YES	YES	YES	YES	YES	NO
18	Office©	NO	NO	NO	NO	NO	NO
19	Price	356 €	331 €	248 €	497 €	165 €	107 €
20	Size	107.5x49x24.5 mm	93x46.5x22.5 mm	83x46x24 mm	100x46x20.5 mm	99.1x42.6x18.2 mm	101.5x43.9x19 mm
21	SMS	YES	YES	YES	YES	YES	YES
22	Speech Recording	YES	YES	YES	YES	YES	YES
23	Stand-by Time	300h	400h	400h	400h	300h	300h
24	Stereo FM	NO	YES	NO	YES	NO	NO
25	Talk Time	7h	8.5h	9h	9h	7h	12h
26	UMTS(3G) Compatibility	NO	NO	NO	NO	NO	NO
27	VibraCall Alert	YES	YES	YES	YES	YES	YES
28	Video Playback with Sound	YES	YES	YES	YES	YES	NO
29	Video Streaming	YES	YES	YES	YES	YES	NO
30	Voice Command	YES	YES	YES	YES	YES	NO
31	WAP	YES	YES	YES	YES	YES	YES
32	Weight	137g	120g	94g	99g	78g	79g
33	Windows© Mobile	NO	NO	NO	NO	NO	NO

* used only in Experiment 1

† used only in Experiment 2

table continued on next page

Table B.1—*Continued*

		Mobile Phones						
	Attribute	32	33	34	35	36	37	38
1a	Image*	see Figure B.3						
1b	Phone Name†	Sony S710a	Sony K500i	Sony P900a	Nokia 8801	Nokia 6256i	Nokia 6236i	Nokia 6235i
2	Band Support	tri-band	tri-band	tri-band	tri-band	dual-band	dual-band	dual-band
3	Bluetooth	YES	NO	YES	YES	YES	NO	NO
4	Color	Black	Blue or silver	Silver	Silver	Silver	Black or silver	Silver
5	Connectivity (USB)	YES	NO	YES	NO	YES	NO	YES
6	Email Support (POP3+IMAP4)	YES	YES	YES	YES	YES	NO	NO
7	External LCD Display	NO	NO	NO	NO	96x65, 4k colors	NO	NO
8	External Memory	up to 128 MB	NO	up to 1 GB	NO	up to 65 MB	NO	NO
9	GPRS	YES	YES	YES	YES	YES	YES	YES
10	Infrared	YES	YES	YES	NO	YES	YES	YES
11	Integrated Digital Camera	1.3 Mpixel	0.3 Mpixel	0.3 Mpixel	0.48 Mpixel	0.3 Mpixel	0.3 Mpixel	0.3 Mpixel
12	Integrated Speakerphone	YES	YES	YES	YES	YES	YES	YES
13	Internal Display	240x320, 262k colors	128x160, 65k colors	208x320, 262k colors	208x208, 265k colors	128x160, 65k colors	128x128, 65k colors	128x128, 65k colors
14	Internal Memory	32 MB	12 MB	64 MB	64 MB	8 MB	10 MB	10 MB
15	Java Games	YES	YES	YES	YES	YES	YES	YES
16	MMS (Photos+Text+Sound)	YES	YES	YES	YES	YES	YES	YES
17	MP3 player	YES	YES	YES	YES	YES	YES	YES
18	Office©	NO	NO	YES	NO	NO	NO	NO
19	Price	356 €	207 €	664 €	664 €	249 €	191 €	216 €
20	Size	107.5x49x24.5 mm	102x46x14 mm	115x58x26 mm	103x44x18 mm	84x46x25 mm	102x41x17 mm	102x41x17 mm
21	SMS	YES	YES	YES	YES	YES	YES	YES
22	Speech Recording	NO	YES	YES	YES	YES	YES	YES
23	Stand-by Time	300h	300h	400h	192h	262h	286h	288h
24	Stereo FM	NO	NO	NO	YES	YES	YES	YES
25	Talk Time	7h	7h	13h	3h	3.2h	3.5h	3.5h
26	UMTS(3G) Compatibility	NO	NO	NO	NO	NO	NO	NO
27	VibraCall Alert	YES	YES	YES	YES	YES	YES	YES
28	Video Playback with Sound	YES	YES	YES	YES	YES	YES	YES
29	Video Streaming	YES	NO	YES	NO	YES	YES	YES
30	Voice Command	YES	YES	YES	YES	YES	YES	YES
31	WAP	YES	YES	YES	NO	YES	YES	YES
32	Weight	137g	88g	155g	134g	126g	98g	98g
33	Windows© Mobile	NO	NO	NO	NO	NO	NO	NO

* used only in Experiment 1
† used only in Experiment 2

table continued on next page

Table B.1—*Continued*

		Mobile Phones						
	Attribute	39	40	41	42	43	44	45
1a	Image*	see Figure B.3						
1b	Phone Name†	Nokia 6102	Nokia 6061	Nokia 6030	Nokia 7270	Nokia 6682	Nokia 7610	Nokia 7280
2	Band Support	tri-band	dual-band	dual-band	tri-band	tri-band	tri-band	tri-band
3	Bluetooth	NO	NO	NO	NO	YES	YES	YES
4	Color	Black or silver	Black or silver	Black	Black or silver	White, blue or silver	Black	Black
5	Connectivity (USB)	YES	NO	NO	NO	YES	YES	NO
6	Email Support (POP3+IMAP4)	YES	YES	NO	NO	YES	YES	NO
7	External LCD Display	96x65, 4k colors	NO	NO	96x65, 4k colors	NO	NO	NO
8	External Memory	NO	NO	NO	NO	up to 512 MB	up to 64 MB	NO
9	GPRS	YES	YES	YES	YES	YES	YES	YES
10	Infrared	YES	NO	NO	NO	NO	NO	YES
11	Integrated Digital Camera	0.3 Mpixel	NO	NO	0.3 Mpixel	1.3 Mpixel	1 Mpixel	0.3 Mpixel
12	Integrated Speakerphone	YES	YES	YES	NO	NO	YES	YES
13	Internal Display	128x160, 65k colors	128x160, 65k colors	128x128, 65k colors	128x160, 65k colors	176x208, 262k colors	176x208, 65k colors	104x208, 65k colors
14	Internal Memory	3.5 MB	3 MB	0.5 MB	32 MB	10 MB	8 MB	50 MB
15	Java Games	YES	NO	YES	NO	YES	NO	NO
16	MMS (Photos+Text+Sound)	YES	YES	YES	YES	YES	YES	YES
17	MP3 player	YES	YES	YES	YES	YES	NO	NO
18	Office©	NO	NO	NO	NO	YES	NO	NO
19	Price	124 €	83 €	124 €	255 €	559 €	141 €	332 €
20	Size	82x43x23 mm	82x43x23 mm	100x42x17 mm	88x46x22 mm	108x55x21 mm	105x51x18 mm	111x31x18 mm
21	SMS	YES	YES	YES	YES	YES	YES	YES
22	Speech Recording	YES	NO	NO	YES	NO	YES	YES
23	Stand-by Time	336h	384h	288h	270h	264h	240h	264h
24	Stereo FM	YES	NO	YES	YES	NO	NO	YES
25	Talk Time	4h	3.5h	3h	4h	4h	3h	4h
26	UMTS(3G) Compatibility	NO	NO	NO	NO	NO	NO	NO
27	VibraCall Alert	YES	YES	YES	YES	YES	YES	YES
28	Video Playback with Sound	YES	NO	NO	YES	YES	YES	NO
29	Video Streaming	NO	NO	NO	NO	NO	NO	NO
30	Voice Command	YES	NO	NO	YES	NO	YES	YES
31	WAP	YES	YES	YES	NO	YES	YES	YES
32	Weight	97g	93g	122g	121g	130g	117g	85g
33	Windows© Mobile	NO	NO	NO	NO	YES	NO	NO

* used only in Experiment 1

† used only in Experiment 2

table continued on next page

Table B.1—*Continued*

		Mobile Phones				
	Attribute	46	47	48	49	50‡
1a	Image*	see Figure B.3				
1b	Phone Name†	Nokia 6820	Nokia 6822	Nokia 9300	Audiovox SMT5600	Nokia 7830
2	Band Support	triband	triband	tri-band	tri-band	tri-band
3	Bluetooth	YES	YES	YES	YES	YES
4	Color	Silver or blue	Silver	Silver	Black	silver
5	Connectivity (USB)	NO	NO	YES	YES	NO
6	Email Support (POP3+IMAP4)	YES	YES	YES	YES	NO
7	External LCD Display	NO	NO	128x128, 65k colors	NO	NO
8	External Memory	NO	NO	64 MB	up to 1 GB	NO
9	GPRS	YES	YES	YES	YES	yes
10	Infrared	YES	YES	NO	NO	NO
11	Integrated Digital Camera	1 Mpixel	0.3 Mpixel	NO	0.3 Mpixel	2 Mpixel
12	Integrated Speakerphone	YES	YES	YES	YES	YES
13	Internal Display	128x128, 4k colors	128x128, 65k colors	640x200, 65k colors	176x220, 65k colors	104 x 208, 65k colors
14	Internal Memory	3.5 MB	5 MB	80 MB	64 MB	52 MB
15	Java Games	YES	YES	YES	YES	YES
16	MMS (Photos+Text+Sound)	YES	YES	YES	YES	YES
17	MP3 player	NO	NO	YES	YES	YES
18	Office©	NO	NO	NO	NO	NO
19	Price	208 €	332 €	332 €	124 €	415 €
20	Size	103x44x20 mm	103x45x19 mm	135x49x20 mm	104x44x15 mm	110x29x20 mm
21	SMS	YES	YES	YES	YES	YES
22	Speech Recording	YES	NO	YES	YES	YES
23	Stand-by Time	288h	288h	200h	288h	240 h
24	Stereo FM	NO	NO	NO	NO	YES
25	Talk Time	3h	3h	4h	3h	3 h
26	UMTS(3G) Compatibility	NO	NO	NO	NO	NO
27	VibraCall Alert	YES	YES	YES	YES	YES
28	Video Playback with Sound	YES	YES	YES	YES	YES
29	Video Streaming	NO	NO	NO	YES	NO
30	Voice Command	NO	YES	NO	NO	YES
31	WAP	YES	YES	YES	YES	YES
32	Weight	100g	100g	167g	104g	80g
33	Windows© Mobile	NO	NO	NO	YES	YES

* used only in Experiment 1
† used only in Experiment 2
‡ used only in Experiment 1

Figure B.1: Images of the mobile phones used in Experiment 1 (in alphabetical order)

B.2 Experiment 3

Table B.2: Stimuli used in Experiment 3: a) Mobile phones

		Mobile Phones				
	Attribute	Samsung SGH-G600	Samsung SGH-M110	Samsung SGH-F700	Samsung SGH-G800	Samsung SGH-U700
1	Price	249 CHF	1 CHF	359 CHF	359 CHF	149 CHF
2	Size	101 x 47.8 x 14.9 mm	109 x 48 x 17.9 mm	112 x 56 x 15.9 mm	103 x 51 x 17 mm	103 x 50 x 12.1 mm
3	Weight	104 g	95 g	139 g	139 g	86 g
4	Standby	300 h	400 h	300 h	370 h	270 h
5	Talk time	3.5 h	8 h	4.5 h	8.7 h	5 h
6	Camera	5 Mpixel	0.3 Mpixel	3.15 Mpixel	5 Mpixel	3.2 Mpixel
7	Music Player	Yes	No	Yes	Yes	Yes
8	Internal Memory	55 MB	2 MB	100 MB	160 MB	42 MB
9	Display	240 x 320 Pixel	128 x 128 Pixel	240 x 440 Pixel	240 x 320 Pixel	240 x 320 Pixel
10	Band Support	Quad-band	Dual-band	Tri-band	Tri-band	Quad-band
11	Email	Yes	No	Yes	Yes	No
12	Active Sync	No	No	No	No	No
13	LiveTV	Yes	No	No	Yes	No
14	WLAN	No	No	No	No	No
15	Infrared	No	No	No	No	No
16	Bluetooth	Yes	Yes	Yes	Yes	Yes
17	USB	Yes	No	Yes	Yes	Yes
18	UMTS	No	No	No	Yes	Yes
19	GPRS	Yes	Yes	Yes	Yes	Yes
20	EDGE	Yes	No	Yes	No	Yes
21	HSDPA	No	No	Yes	No	Yes
22	SAR	0.56 W/kg	0.29 W/kg	0.07 W/kg	0.221 W/kg	0.82 W/kg
23	Customer Rating*	4.5 (10)	2.5 (2)	5 (1)	3.5 (15)	3 (51)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high); the numbers in brackets indicate the number of customers who rated this product

table continued on next page

Table B.2—*Continued*

		Mobile Phones						
	Attribute	Nokia 5610 Xpress M	Nokia 6500 Classic	Nokia 6500 Slide	Nokia N95	Nokia 6110	Nokia N82 2GB	Nokia N81 SD 2GB
1	Price	249 CHF	239 CHF	239 CHF	359 CHF	249 CHF	499 CHF	349 CHF
2	Size	98.5 x 48.5 x 17 mm	96.5 x 46.5 x 16.4 mm	96.5 x 46.5 x 16.4 mm	99 x 53 x 21 mm	101 x 49 x 20 mm	112 x 50.2 x 17.3 mm	102 x 50 x 17.9 mm
3	Weight	111 g	125 g	125 g	128 g	125 g	114 g	140 g
4	Standby	320 h	310 h	310 h	280 h	265 h	225 h	410 h
5	Talk time	6 h	6 h	6 h	6 h	3.5 h	4.3 h	4 h
6	Camera	3.15 Mpixel	2 Mpixel	3.15 Mpixel	5 Mpixel	2 Mpixel	5 Mpixel	2 Mpixel
7	Music Player	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8	Internal Memory	20 MB	20 MB	No	160 MB	No	100 MB	No
9	Display	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel
10	Band Support	Quad-band	Quad-band	Quad-band	Quad-band	Quad-band	Quad-band	Quad-band
11	Email	Yes	Yes	Yes	Yes	Yes	Yes	Yes
12	Active Sync	No	No	No	No	No	No	No
13	LiveTV	No	Yes	No	No	No	No	Yes
14	WLAN	No	No	No	Yes	No	No	Yes
15	Infrared	No	No	No	Yes	No	No	No
16	Bluetooth	Yes	Yes	Yes	Yes	Yes	Yes	Yes
17	USB	Yes	Yes	Yes	Yes	Yes	Yes	Yes
18	UMTS	No	Yes	Yes	Yes	No	No	No
19	GPRS	Yes	Yes	Yes	Yes	Yes	Yes	Yes
20	EDGE	Yes	Yes	Yes	Yes	Yes	No	Yes
21	HSDPA	No	No	No	Yes	Yes	Yes	No
22	SAR	1.14 W/kg	1.10 W/kg	1.31 W/kg	0.58 W/kg	1.06 W/kg	0.62 W/kg	0.96 W/kg
23	Customer Rating*	3 (9)	3.5 (12)	4 (16)	4.5 (23)	4 (21)	4 (15)	5 (3)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high);
the numbers in brackets indicate the number of customers who rated this product

table continued on next page

Table B.2—*Continued*

		Mobile Phones						
	Attribute	Nokia 5310 Xpress	Nokia 8800 Arte	Nokia N95 8GB	Nokia 3110	Nokia 5700 Xpress	Nokia 6120 Classic	Nokia 6267
1	Price	139 CHF	1259 CHF	759 CHF	1 CHF	259 CHF	1 CHF	59 CHF
2	Size	103.8 x 44.7 x 9.9 mm	109 x 45.6 x 14.6 mm	99 x 53 x 21 mm	108.5 x 45.7 x 15.6 mm	108 x 50 x 17 mm	105 x 46 x 15 mm	93.9 x 46.9 x 21.5 mm
3	Weight	70.2 g	150 g	128 g	87 g	115 g	89 g	105 g
4	Standby	300 h	300 h	280 h	370 h	290 h	250 h	290 h
5	Talk time	5.3 h	3 h	6 h	4 h	3.5 h	3 h	6.5 h
6	Camera	2 Mpixel	3.15 Mpixel	5 Mpixel	1.3 Mpixel	2 Mpixel	2 Mpixel	2 Mpixel
7	Music Player	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8	Internal Memory	30 MB	1024 MB	8192 MB	9 MB	No	No	No
9	Display	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	128 x 160 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel
10	Band Support	Tri-band	Tri-band	Quad-band	Tri-band	Quad-band	Quad-band	Quad-band
11	Email	Yes	Yes	Yes	Yes	No	Yes	Yes
12	Active Sync	No	No	No	No	No	No	No
13	LiveTV	Yes	No	No	No	No	No	No
14	WLAN	No	No	Yes	No	No	No	No
15	Infrared	No	No	Yes	Yes	Yes	No	No
16	Bluetooth	Yes	Yes	Yes	Yes	Yes	Yes	Yes
17	USB	Yes	Yes	Yes	Yes	Yes	Yes	Yes
18	UMTS	No	Yes	No	No	Yes	Yes	Yes
19	GPRS	Yes	Yes	Yes	Yes	Yes	Yes	Yes
20	EDGE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
21	HSDPA	No	No	Yes	No	No	Yes	No
22	SAR	1.07 W/kg	0.36 W/kg	0.58 W/kg	0.78 W/kg	1.24 W/kg	1.18 W/kg	0.31 W/kg
23	Customer Rating*	4 (5)	5 (1)	4.5 (23)	4.5 (33)	4.5 (2)	4 (26)	4.5 (9)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high); the numbers in brackets indicate the number of customers who rated this product

table continued on next page

Table B.2—*Continued*

		Mobile Phones						
	Attribute	Nokia 6300	Nokia 8600 Luna	Nokia E90 Comm	Nokia N73	Sony Ericsson W910i	Sony Ericsson T650i	Sony Ericsson W890i
1	Price	1 CHF	629 CHF	959 CHF	139 CHF	239 CHF	249 CHF	349 CHF
2	Size	106.4 x 43.6 x 11.7 mm	107 x 45 x 15.9 mm	132 x 57 x 20 mm	110 x 49 x 19 mm	99.5 x 50 x 12.5 mm	104 x 46 x 12.5 mm	104 x 46.5 x 9.9 mm
3	Weight	91 g	143 g	210 g	116 g	86 g	95 g	78 g
4	Standby	348 h	240 h	330 h	350 h	400 h	300 h	360 h
5	Talk time	3.5 h	3.5 h	5 h	6 h	9 h	7 h	9.5 h
6	Camera	2 Mpixel	2 Mpixel	3.15 Mpixel	3.15 Mpixel	2 Mpixel	3.2 Mpixel	3.15 Mpixel
7	Music Player	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8	Internal Memory	No	128 MB	128 MB	42 MB	40 MB	16 MB	26 MB
9	Display	240 x 320 Pixel	240 x 320 Pixel	800 x 352 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel
10	Band Support	Quad-band	Quad-band	Quad-band	Quad-band	Quad-band	Quad-band	Quad-band
11	Email	Yes	Yes	Yes	Yes	Yes	Yes	Yes
12	Active Sync	No	No	Yes	No	No	No	No
13	LiveTV	No	No	No	No	No	No	No
14	WLAN	No	No	Yes	No	No	No	No
15	Infrared	No	No	Yes	Yes	No	No	No
16	Bluetooth	Yes	Yes	Yes	Yes	Yes	Yes	Yes
17	USB	Yes	Yes	Yes	Yes	No	Yes	Yes
18	UMTS	No	No	No	Yes	Yes	Yes	No
19	GPRS	Yes	Yes	Yes	Yes	No	Yes	Yes
20	EDGE	Yes	Yes	Yes	Yes	No	No	Yes
21	HSDPA	No	No	Yes	No	Yes	No	Yes
22	SAR	0.57 W/kg	0.88 W/kg	0.58 W/kg	1.13 W/kg	0.9 W/kg	1.80 W/kg	1.38 W/kg
23	Customer Rating*	4.5 (95)	4 (9)	3 (20)	3.5 (31)	4 (23)	5 (4)	4 (5)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high);
the numbers in brackets indicate the number of customers who rated this product

table continued on next page

Table B.2—*Continued*

		Mobile Phones						
	Attribute	Sony Ericsson K770i	Sony Ericsson W960i	Sony Ericsson W880i	Sony Ericsson K850i	Sony Ericsson K810i	Sony Ericsson P1i	Sony Ericsson W580i
1	Price	149 CHF	499 CHF	149 CHF	259 CHF	139 CHF	359 CHF	99 CHF
2	Size	105 x 47 x 14.5 mm	109 x 55 x 16 mm	103 x 46.5 x 9.5 mm	102 x 48 x 17 mm	106 x 48 x 17 mm	106 x 55 x 17 mm	99 x 47 x 14 mm
3	Weight	95 g	119 g	71 g	118 g	103 g	124 g	94 g
4	Standby	400 h	370 h	425 h	400 h	400 h	440 h	370 h
5	Talk time	10 h	9 h	6.5 h	9 h	10 h	10 h	9 h
6	Camera	3.2 Mpixel	3.15 Mpixel	2 Mpixel	5 Mpixel	3.2 Mpixel	3.2 Mpixel	2 Mpixel
7	Music Player	Yes	Yes	Yes	Yes	Yes	Yes	Yes
8	Internal Memory	16 MB	8 GB	16 MB	40 MB	64 MB	160 MB	12 MB
9	Display	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel
10	Band Support	Tri-band	Tri-band	Tri-band	Quad-band	Tri-band	Tri-band	Yes
11	Email	Yes	Yes	Yes	Yes	Yes	Yes	Yes
12	Active Sync	No	Yes	No	No	No	Yes	No
13	LiveTV	No	Yes	No	Yes	No	Yes	No
14	WLAN	No	Yes	No	No	No	Yes	No
15	Infrared	No	No	No	No	Yes	Yes	No
16	Bluetooth	Yes	Yes	Yes	Yes	Yes	Yes	Yes
17	USB	Yes	Yes	Yes	Yes	Yes	Yes	Yes
18	UMTS	Yes	Yes	Yes	No	Yes	Yes	No
19	GPRS	Yes	Yes	Yes	Yes	Yes	Yes	Yes
20	EDGE	No	No	No	Yes	No	No	Yes
21	HSDPA	No	No	No	Yes	No	No	No
22	SAR	1.69 W/kg	0.96 W/kg	1.45 W/kg	0.67 W/kg	1.31 W/kg	0.56 W/kg	0.74 W/kg
23	Customer Rating*	4.5 (12)	4.5 (10)	4 (34)	4 (9)	4.5 (37)	4 (64)	3 (29)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high);
the numbers in brackets indicate the number of customers who rated this product

table continued on next page

Table B.2—*Continued*

		Mobile Phones						
	Attribute	HTC Touch Cruise	HTC S730	HTC Touch Dual	HTC TyTN II	Sagem my850C	LG KF 600	LG KU990 Viewty
1	Price	559 CHF	359 CHF	469 CHF	659 CHF	1 CHF	259 CHF	359 CHF
2	Size	110 x 58 x 15.5 mm	100 x 50 x 19 mm	107 x 55 x 15.8 mm	112 x 59 x 19 mm	96 x 49 x 16 mm	101.2 x 50.7 x 14.1 mm	103.5 x 54.4 x 14.8 mm
3	Weight	130 g	120 g	120 g	190 g	95 g	107 g	112 g
4	Standby	450 h	290 h	180 h	400 h	300 h	480 h	340 h
5	Talk time	7 h	318 h	5 h	6 h	3.5 h	4 h	4 h
6	Camera	3.15 Mpixel	2 Mpixel	2 Mpixel	3.0 Mpixel	2 Mpixel	3.15 Mpixel	5 Mpixel
7	Music Player	Yes	Yes	No	Yes	Yes	Yes	Yes
8	Internal Memory	128 MB	64 MB	No	128 MB	64 MB	No	100 MB
9	Display	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 320 Pixel	240 x 400 Pixel
10	Band Support	Quad-band	Quad-band	Tri-band	Quad-band	Tri-band	Tri-band	Tri-band
11	Email	Yes	Yes	Yes	Yes	Yes	Yes	Yes
12	Active Sync	Yes	Yes	Yes	Yes	No	No	No
13	LiveTV	No	Yes	Yes	No	No	No	Yes
14	WLAN	Yes	Yes	No	Yes	No	No	No
15	Infrared	No	No	No	No	No	No	No
16	Bluetooth	Yes	Yes	Yes	Yes	Yes	Yes	Yes
17	USB	Yes	Yes	Yes	Yes	Yes	Yes	Yes
18	UMTS	No	No	Yes	No	Yes	No	No
19	GPRS	Yes	Yes	Yes	Yes	Yes	Yes	Yes
20	EDGE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
21	HSDPA	Yes	Yes	Yes	Yes	No	No	Yes
22	SAR	0.89 W/kg	0.82 W/kg	1.59 W/kg	0.39 W/kg	0.70 W/kg	0.743 W/kg	0.826 W/kg
23	Customer Rating*	4.5 (7)	4.5 (10)	3.5 (2)	3.5 (17)	4 (5)	4 (18)	4 (34)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high);
the numbers in brackets indicate the number of customers who rated this product

table continued on next page

Table B.2—*Continued*

		Mobile Phones				
	Attribute	Motorola RAZR2 V8	Motorola KRZR K1	Motorola MOTO Q9h	Motorola RAZR V8	Sonim XP1
1	Price	549 CHF	59 CHF	329 CHF	149 CHF	149 CHF
2	Size	103 x 53 x 11.9 mm	103 x 42 x 16 mm	118 x 67 x 11.8 mm	103 x 53 x 11.9 mm	133 x 50 x 22 mm
3	Weight	117 g	102 g	134 g	117 g	130 g
4	Standby	330 h	300 h	480 h	280 h	8 h
5	Talk time	7.8 h	6 h	6.5 h	6.5 h	4 h
6	Camera	2 Mpixel	2 Mpixel	2.0 Mpixel	2.0 Mpixel	No
7	Music Player	Yes	Yes	Yes	Yes	No
8	Internal Memory	420 MB	20 MB	100 MB	512 MB	9 MB
9	Display	240 x 320 Pixel	176 x 220 Pixel	320 x 240 Pixel	240 x 320 Pixel	128 x 160 Pixel
10	Band Support	Quad-band	Quad-band	Quad-band	Quad-band	Tri-band
11	Email	Yes	Yes	Yes	Yes	Yes
12	Active Sync	No	No	Yes	No	No
13	LiveTV	No	No	No	No	No
14	WLAN	No	No	No	No	No
15	Infrared	No	No	No	No	No
16	Bluetooth	Yes	Yes	Yes	Yes	No
17	USB	Yes	Yes	Yes	Yes	Yes
18	UMTS	No	No	No	No	No
19	GPRS	Yes	Yes	Yes	Yes	No
20	EDGE	Yes	Yes	Yes	Yes	No
21	HSDPA	No	No	Yes	No	No
22	SAR	0.34 W/kg	0.59 W/kg	0.47 W/kg	0.34 W/kg	0.68 W/kg
23	Customer Rating*	3.5 (17)	3.5 (26)	2 (1)	5 (3)	4 (3)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high); the numbers in brackets indicate the number of customers who rated this product

Table B.3: Stimuli used in Experiment 3: b) Digital cameras

		Digital Cameras						
	Attribute	Canon A 720 IS	Canon A560	Canon IXUS 70	Canon IXUS 75	Canon IXUS 860 IS	Canon S5 IS	Canon SX 100 IS
1	Price	293 CHF	188 CHF	268 CHF	301 CHF	443 CHF	515 CHF	350 CHF
2	Resolution (Mpixel)	8.0 Mpixel	7.1 Mpixel	7.1 Mpixel	7.1 Mpixel	8.0 Mpixel	8.0 Mpixel	8.3 Mpixel
3	Optical zoom	6x	4x	3x	3x	3.8x	12x	10x
4	Digital zoom	4x	4x	4x	4x	4x	4x	4x
5	Display size	2.5 ”	2.5 ”	2.5”	3.0 ”	3.0 ”	2.5 ”	2.5 ”
6	Storage included	16 MB	16 MB	32 MB	32 MB	32 MB	32 MB	32 MB
7	Video resolution	640 x 480	640 x 480	640 x 480	640 x 480	640 x 480	640 x 480	640 x 480
8	Size	97 x 67 x 49 mm	91 x 64 x 43 mm	86 x 54 x 19 mm	92 x 57 x 20 mm	93 x 59 x 26 mm	117 x 80 x 78 mm	109 x 71 x 47 mm
9	Weight	250 g	205 g	175 g	170 g	165 g	550 g	316 g
10	Minimum aperture size	F2.8	F2.6	F2.8	F2.9	F2.8	F2.7	F2.8
11	Maximum aperture size	F4.8	F5.5	F4.9	F4.9	F5.8	F8	F4.3
12	Minimum shutter speed	15 sec	15 sec	15 sec	15 sec	15 sec	15 sec	15 sec
13	Maximum shutter speed	1/2000 sec	1/2000 sec	1/1500 sec	1/1500 sec	1/1600 sec	1/3200 sec	1/2500 sec
14	Lowest ISO rating	80	80	80	80	80	80	80
15	Highest ISO rating	1600	1600	1600	1600	1600	1600	1600
16	White balance override	6 positions + manual	5 positions + manual	5 positions + manual	5 positions + manual	5 positions + manual	6 positions + manual	6 positions + manual
17	Image stabilization	yes	no	no	no	yes	yes	yes
18	Optical viewfinder	yes	yes	yes	no	no	no	no
19	Manual exposure control	yes	no	no	no	no	yes	yes
20	Zoom wide	35 mm	35 mm	35 mm	35 mm	28 mm	36 mm	36 mm
21	Manual focus	yes	yes	no	no	no	yes	yes
22	Flip display	no	no	no	no	no	yes	yes
23	Flash distance	3.5 m	2.2 m	3.5 m	3.5 m	4.0 m	5.2 m	3 m
24	Self-timer	2 or 10 sec	2 or 10 sec	2 or 10 sec	2 or 10 sec	2 or 10 sec	2 or 10 sec or custom	2 or 10 sec or custom
25	Continuous drive	1.3 fps	1.7 fps	1.7 fps	1.7 fps	1.3 fps	1.5 fps	1.3 fps
26	Built-in flash	yes	yes	yes	yes	yes	yes	yes
27	External flash	no	no	no	no	no	yes	no
28	USB	yes	yes	yes	yes	yes	yes	yes
29	Firewire (IEEE 1394)	no	no	no	no	no	no	no
30	Orientation sensor	yes	yes	yes	yes	yes	yes	yes
31	Uncompressed format	no	no	no	no	no	no	no
32	Format	Compact	Compact	Ultra Compact	Ultra Compact	Ultra Compact	Compact	Compact
33	Battery	AA	AA	specific	specific	specific	AA	AA
34	Sensor type	CCD	CCD	CCD	CCD	CCD	CCD	CCD
35	Brand	1	1	1	1	1	1	1
36	Storage types	SD/MMC card	SD/SDHC card	SD/SDHC card	SD/SDHC card	SD/SDHC/MMC card	SD/SDHC/MMC card	SD/SDHC/MMC card
37	Customer Rating*	4 (25)	4 (13)	4.5 (72)	4 (53)	4 (33)	4.5 (42)	4.5 (14)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high);
the numbers in brackets indicate the number of customers who rated this product

table continued on next page

Table B.3—*Continued*

		Digital Cameras				
	Attribute	Casio Exilim EX-S880	Casio EXILIM EX-V7	Casio Exilim EX-Z1080	Casio Exilim EX-Z75	Casio Exilim EX-Z80
1	Price	253 CHF	238 CHF	247 CHF	234 CHF	253 CHF
2	Resolution (Mpixel)	8.1 Mpixel	7.1 Mpixel	10.1 Mpixel	7.1 Mpixel	8.1 Mpixel
3	Optical zoom	3x	7x	3x	3x	3x
4	Digital zoom	3x	4x	4x	4x	4x
5	Display size	2.8 ”	2.5 ”	2.6”	2.6 ”	2.6 ”
6	Storage included	10.8 MB	11 MB	11.4 MB	8 MB	11.4 MB
7	Video resolution	848 x 480	848 x 480	848 x 480	640 x 480	848 x 480
8	Size	95 x 60 x 17 mm	96 x 60 x 25 mm	91 x 57 x 24 mm	95 x 61 x 20 mm	90 x 52 x 19 mm
9	Weight	138 g	190 g	165 g	156 g	100 g
10	Minimum aperture size	F2.7	F3.4	F2.8	F3.1	F3.1
11	Maximum aperture size	F5.2	F5.3	F5.1	F5.9	F5.3
12	Minimum shutter speed	4 sec	60 sec	4 sec	4 sec	4 sec
13	Maximum shutter speed	1/2000 sec	1/800 sec	1/1000 sec	1/2000 sec	1/2000 sec
14	Lowest ISO rating	64	64	80	50	64
15	Highest ISO rating	800	1600	3200	400	1600
16	White balance override	6 positions + manual	6 positions + manual	6 positions + manual	6 positions + manual	6 positions + manual
17	Image stabilization	no	yes	yes	no	no
18	Optical viewfinder	no	no	no	no	no
19	Manual exposure control	no	yes	yes	no	no
20	Zoom wide	38 mm	38 mm	38 mm	38 mm	38 mm
21	Manual focus	yes	yes	yes	yes	yes
22	Flip display	no	no	no	no	no
23	Flash distance	2.8 m	2.8 m	3.3 m	3.5 m	2.6 m
24	Self-timer	2 or 10 sec	2 or 10 sec	2 or 10 sec	2 or 10 sec	2 or 10 sec
25	Continuous drive	yes	yes	yes	1.2 sec per image	yes
26	Built-in flash	yes	yes	yes	yes	yes
27	External flash	no	no	no	no	no
28	USB	yes	yes	yes	yes	yes
29	Firewire (IEEE 1394)	no	no	no	no	no
30	Orientation sensor	no	no	no	no	no
31	Uncompressed format	no	no	no	no	no
32	Format	Ultra Compact	Compact	Compact	Ultra Compact	Compact
33	Battery	specific	specific	specific	specific	specific
34	Sensor type	CCD	CCD	CCD	CCD	CCD
35	Brand	2	2	2	2	2
36	Storage types	SD/MMC card + Internal	SD/MMC card + Internal	SC/SDHC/MMC/MMCplus + internal	SC/SDHC/MMC/MMCplus + internal	SD/MMC/SDHC card + Internal
37	Customer Rating*	4 (43)	3 (24)	4.5 (19)	4 (101)	3 (1)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high); the numbers in brackets indicate the number of customers who rated this product

table continued on next page

Table B.3—*Continued*

		Digital Cameras				
	Attribute	FujiFilm F480 Zoom	FujiFilm F50fd	FujiFilm S5700	FujiFilm S8000fd	Kodak C613
1	Price	222 CHF	317 CHF	268 CHF	394 CHF	109 CHF
2	Resolution (Mpixel)	8.0 Mpixel	12.0 Mpixel	7.1 Mpixel	8.0 Mpixel	6.16 Mpixel
3	Optical zoom	4x	3x	10x	18x	3x
4	Digital zoom	5.1x	8.2x	4.8x	5.1x	5x
5	Display size	2.7 "	2.7 "	2.5 "	2.4 "	2.4 "
6	Storage included	12 MB	25 MB	27 MB	58 MB	16 MB
7	Video resolution	320 x 240	640 x 480	640 x 480	640 x 480	640 x 480
8	Size	96 x 55 x 23 mm	93 x 59 x 23 mm	106 x 76 x 81 mm	111 x 78 x 79 mm	91 x 66 x 35 mm
9	Weight	170 g	205 g	406 g	510 g	187 g
10	Minimum aperture size	F2.7	F2.8	F3.5	F2.8	F2.7
11	Maximum aperture size	F5.4	F8	F13.6	F4.5	F4.8
12	Minimum shutter speed	4 sec	8 sec	4 sec	4 sec	1 sec
13	Maximum shutter speed	1/1500 sec	1/2000 sec	1/1000 sec	1/2000 sec	1/1400 sec
14	Lowest ISO rating	64	100	64	100	80
15	Highest ISO rating	800	6400	1600	6400	1250
16	White balance override	5 positions	6 positions + manual	7 positions	6 positions + manual	4 positions
17	Image stabilization	no	yes	no	yes	no
18	Optical viewfinder	no	no	no	no	no
19	Manual exposure control	no	yes	no	yes	no
20	Zoom wide	28 mm	35 mm	38 mm	27 mm	36 mm
21	Manual focus	no	no	yes	yes	no
22	Flip display	no	no	no	no	no
23	Flash distance	4.0 m	6.0 m	6.2 m	8.8 m	2 m
24	Self-timer	2 or 10 sec	2 or 10 sec	2 or 10 sec	2 or 10 sec	2 or 10 sec
25	Continuous drive	yes	yes	yes	1.3 fps	no
26	Built-in flash	yes	yes	yes	yes	yes
27	External flash	no	no	no	no	no
28	USB	yes	yes	yes	yes	yes
29	Firewire (IEEE 1394)	no	no	no	no	no
30	Orientation sensor	no	no	no	no	no
31	Uncompressed format	no	no	no	RAW	no
32	Format	Compact	Compact	SLR-like	SLR-like	Compact
33	Battery	specific	specific	AA	AA	AA
34	Sensor type	CCD	CCD	CCD	CCD	CCD
35	Brand	5	5	5	5	3
36	Storage types	xD Picture Card + Internal+SD/SDHC	xD Picture Card, SD card + Internal	xD Picture Card	xD Picture Card + Internal	SD/MMC card + Internal
37	Customer Rating*	4 (2)	4 (12)	4.5 (43)	4 (10)	4 (30)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high);
the numbers in brackets indicate the number of customers who rated this product

table continued on next page

Table B.3—*Continued*

		Digital Cameras				
	Attribute	Kodak V1273	Kodak Z 710	Kodak Z712	Nikon Coolpix L16	Nikon Coolpix P50
1	Price	378 CHF	296 CHF	375 CHF	188 CHF	250 CHF
2	Resolution (Mpixel)	12 Mpixel	7.1 Mpixel	7.1 Mpixel	7.1 Mpixel	8.1 Mpixel
3	Optical zoom	3x	10x	11.8x	3x	3.6x
4	Digital zoom	5x	5x	4.2x	4x	3x
5	Display size	3.0 "	2.0 "	2.5 "	2.8 "	2.4 "
6	Storage included	32 MB	32 MB	32 MB	21 MB	52 MB
7	Video resolution	1280 x 720	640 x 480	640 x 480	640 x 480	640 x 480
8	Size	93 x 58 x 21 mm	98 x 78 x 73 mm	104 x 74 x 70 mm	95 x 61 x 30 mm	95 x 66 x 44 mm
9	Weight	188 g	360 g	350 g	175 g	210 g
10	Minimum aperture size	F3.1	F2.8	F2.8	F2.8	F2.8
11	Maximum aperture size	F5.7	F3.7	F4.8	F4.7	F5.6
12	Minimum shutter speed	8 sec	8 sec	0.5 sec	4 sec	4 sec
13	Maximum shutter speed	1/1164 sec	1/1000 sec	1/1000 sec	1/1500 sec	1/1000 sec
14	Lowest ISO rating	64	64	80	64	64
15	Highest ISO rating	6400	800	3200	1600	2000
16	White balance override	4 positions	4 positions	4 positions	5 positions + manual	5 positions + manual
17	Image stabilization	yes	no	no	no	yes
18	Optical viewfinder	no	no	no	no	no
19	Manual exposure control	no	yes	yes	no	no
20	Zoom wide	37 mm	38 mm	36 mm	35 mm	28 mm
21	Manual focus	no	no	no	no	no
22	Flip display	no	no	no	no	no
23	Flash distance	4.1 m	4.9 m	4.7 m	3.6 m	6.9 m
24	Self-timer	2 or 10 sec	2 or 10 sec	2 or 10 sec	2 or 10 sec	3 or 10 sec
25	Continuous drive	3 fps	1.7 fps	yes	yes	yes
26	Built-in flash	yes	yes	yes	yes	yes
27	External flash	no	no	no	no	no
28	USB	yes	yes	yes	yes	yes
29	Firewire (IEEE 1394)	no	no	no	no	no
30	Orientation sensor	yes	no	no	no	no
31	Uncompressed format	no	no	no	no	no
32	Format	Ultra Compact	SLR-like	SLR-like	Compact	Compact
33	Battery	specific	AA	specific	AA	specific
34	Sensor type	CCD	CCD	CCD	CCD	CCD
35	Brand	3	3	3	9	9
36	Storage types	SD/SDHC/MMC card + Internal	SD/MMC card + Internal	SD/MMC card + Internal	SD/MMC/SDHC card + Internal	SD/MMC/SDHC card + Internal
37	Customer Rating*	3 (1)	4 (9)	4 (7)	4 (5)	4 (5)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high);
the numbers in brackets indicate the number of customers who rated this product

table continued on next page

Table B.3—*Continued*

		Digital Cameras				
	Attribute	Nikon Coolpix S200	Nikon Coolpix S51	Nikon Coolpix S550	Olympus SP-510UZ	Olympus SP-560UZ
1	Price	203 CHF	315 CHF	405 CHF	253 CHF	475 CHF
2	Resolution (Mpixel)	7.1 Mpixel	8.9 Mpixel	10.0 Mpixel	7.1 Mpixel	8.0 Mpixel
3	Optical zoom	3x	3x	5x	10x	18x
4	Digital zoom	4x	4x	4x	5x	5.6x
5	Display size	2.5 ”	3.0 ”	2.5 ”	2.5 ”	2.5 ”
6	Storage included	20 MB	13 MB	50 MB	21 MB	46 MB
7	Video resolution	640 x 480	640 x 480	640 x 480	640 x 480	640 x 480
8	Size	92 x 57 x 19 mm	93 x 59 x 21 mm	90 x 53.5 x 22 mm	106 x 75 x 70 mm	116 x 79 x 78 mm
9	Weight	165 g	175 g	150 g	410 g	365 g
10	Minimum aperture size	F3.1	F3.3	F2.7	F2.8	F2.8
11	Maximum aperture size	F5.9	F4.2	F5.8	F3.7	F4.5
12	Minimum shutter speed	4 sec	4 sec	4 sec	15 sec	4 sec
13	Maximum shutter speed	1/1000 sec	1/1500 sec	1/1500 sec	1/1000 sec	1/2000 sec
14	Lowest ISO rating	50	100	64	80	50
15	Highest ISO rating	1000	1600	2000	4000	6400
16	White balance override	5 positions + manual	5 positions + manual	5 positions + manual	5 positions + manual	6 positions + manual
17	Image stabilization	yes	yes	no	no	yes
18	Optical viewfinder	no	no	no	no	no
19	Manual exposure control	no	no	no	yes	yes
20	Zoom wide	38 mm	38 mm	35 mm	38 mm	27 mm
21	Manual focus	yes	no	no	yes	yes
22	Flip display	no	no	no	no	no
23	Flash distance	5.4 m	6.0 m	3.5 m	4.5 m	4.5 m
24	Self-timer	10 sec	3 or 10 sec	3 or 10 sec	2 or 12 sec	2 or 12 sec
25	Continuous drive	yes	yes	yes	1.7 fps	15 fps
26	Built-in flash	yes	yes	yes	yes	yes
27	External flash	no	no	no	no	no
28	USB	yes	yes	yes	yes	yes
29	Firewire (IEEE 1394)	no	no	no	no	no
30	Orientation sensor	no	no	no	no	no
31	Uncompressed format	no	no	no	RAW	RAW
32	Format	Ultra Compact	Ultra Compact	Ultra Compact	SLR-like	SLR-like
33	Battery	specific	specific	specific	AA	AA
34	Sensor type	CCD	CCD	CCD	CCD	CCD
35	Brand	9	9	9	8	8
36	Storage types	SD/SDHC card + Internal	SD/SDHC card + Internal	SD/MMC/SDHC card + Internal	xD Picture Card + Internal	xD Picture Card + Internal
37	Customer Rating*	3.5 (7)	4 (65)	4 (12)	3.5 (10)	5.5 (6)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high);
the numbers in brackets indicate the number of customers who rated this product

table continued on next page

Table B.3—*Continued*

		Digital Cameras				
	Attribute	Olympus μ 1030 SW	Olympus μ 770 SW	Panasonic DMC-FX55	Panasonic DMC-FZ18	Panasonic DMC-FZ50
1	Price	582 CHF	475 CHF	364 CHF	553 CHF	729 CHF
2	Resolution (Mpixel)	10.1 Mpixel	7.1 Mpixel	8.1 Mpixel	8.0 Mpixel	10.0 Mpixel
3	Optical zoom	3.6x	3x	3.6x	18x	12x
4	Digital zoom	5x	5x	4x	4x	4x
5	Display size	2.7 "	2.5 "	3.0 "	2.5 "	2.0 "
6	Storage included	14.6 MB	18 MB	27 MB	27 MB	32 MB
7	Video resolution	640 x 480	640 x 480	848 x 480	848 x 480	848 x 480
8	Size	93 x 61 x 21 mm	92 x 59 x 21 mm	95 x 57 x 23 mm	118 x 75 x 88 mm	141 x 86 x 142 mm
9	Weight	203 g	195 g	165 g	407 g	734 g
10	Minimum aperture size	F3.5	F3.5	F2.8	F2.8	F2.8
11	Maximum aperture size	F5.1	F5.0	F5.6	F8	F11
12	Minimum shutter speed	0.5 sec	4 sec	8 sec	60 sec	60 sec
13	Maximum shutter speed	1/2000 sec	1/2000 sec	1/2000 sec	1/2000 sec	1/2000 sec
14	Lowest ISO rating	80	80	100	100	100
15	Highest ISO rating	1600	1600	6400	6400	3200
16	White balance override	6 positions	5 positions	4 positions + manual	4 positions + manual	4 positions + 2 manual
17	Image stabilization	no	yes	yes	yes	yes
18	Optical viewfinder	no	no	no	no	no
19	Manual exposure control	no	no	no	yes	yes
20	Zoom wide	28 mm	38 mm	28 mm	28 mm	35 mm
21	Manual focus	no	no	no	yes	yes
22	Flip display	no	no	no	no	yes
23	Flash distance	3.8 m	3.8 m	6.3 m	6.0 m	7 m
24	Self-timer	12 sec	12 sec	2 or 10 sec	2 or 10 sec	2 or 10 sec
25	Continuous drive	yes	1.1 fps	3 fps	3.2 fps	2 fps
26	Built-in flash	yes	yes	yes	yes	yes
27	External flash	no	no	no	no	yes
28	USB	yes	yes	yes	yes	yes
29	Firewire (IEEE 1394)	no	no	no	no	no
30	Orientation sensor	no	no	yes	no	yes
31	Uncompressed format	no	no	no	RAW	RAW
32	Format	Ultra Compact	Ultra Compact	Ultra Compact	SLR-like	SLR-like
33	Battery	specific	specific	specific	specific	specific
34	Sensor type	CCD	CCD	CCD	CCD	CCD
35	Brand	8	8	6	6	6
36	Storage types	xD Picture Card + Internal	xD Picture Card + Internal	SD/MMC/SDHC card + Internal	SD/MMC/SDHC card + Internal	SD/MMC/SDHC card
37	Customer Rating*	3.5 (10)	5 (6)	4 (4)	4 (18)	4 (28)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high);
the numbers in brackets indicate the number of customers who rated this product

table continued on next page

Table B.3—*Continued*

		Digital Cameras				
	Attribute	Panasonic DMC-TZ2	Panasonic DMC-TZ3	Panasonic DMC-TZ5	Samsung L730	Samsung L830
1	Price	315 CHF	380 CHF	554 CHF	176 CHF	190 CHF
2	Resolution (Mpixel)	6.0 Mpixel	7.1 Mpixel	9.1 Mpixel	7.2 Mpixel	8.1 Mpixel
3	Optical zoom	10x	10x	10x	3x	3x
4	Digital zoom	4x	4x	4x	5x	5x
5	Display size	2.5 ”	3.0 ”	3.0 ”	2.5 ”	2.5 ”
6	Storage included	13 MB	13 MB	50 MB	16 MB	10 MB
7	Video resolution	848x 480	848x 480	1280 x 720	800 x 592	800 x 592
8	Size	105 x 59 x 37 mm	105 x 59 x 37 mm	103 x 59 x 36 mm	95 x 59 x 21 mm	95 x 59 x 21 mm
9	Weight	260 g	280 g	245 g	183 g	183 g
10	Minimum aperture size	F3.3	F3.3	F3.3	F3.0	F3.0
11	Maximum aperture size	F4.9	F4.9	F4.9	F5.6	F5.6
12	Minimum shutter speed	8 sec	8 sec	60 sec	8 sec	8 sec
13	Maximum shutter speed	1/2000 sec	1/2000 sec	1/2000 sec	1/1500 sec	1/1500 sec
14	Lowest ISO rating	100	100	100	50	50
15	Highest ISO rating	3200	3200	1600	1600	1600
16	White balance override	4 positions + 2 manual	4 positions + 2 manual	4 positions + manual	5 positions + manual	5 positions + manual
17	Image stabilization	yes	yes	yes	no	no
18	Optical viewfinder	no	no	no	no	no
19	Manual exposure control	no	no	no	no	no
20	Zoom wide	28 mm	28 mm	28 mm	38 mm	38 mm
21	Manual focus	no	no	no	no	no
22	Flip display	no	no	no	no	no
23	Flash distance	4.2 m	4.2 m	5.3 m	4.6 m	4.6 m
24	Self-timer	2 or 10 sec	2 or 10 sec	2 or 10 sec	2 or 10 sec	2 or 10 sec
25	Continuous drive	no	3 fps	2.5 fps	yes	yes
26	Built-in flash	yes	yes	yes	yes	yes
27	External flash	no	no	no	no	no
28	USB	yes	yes	yes	yes	yes
29	Firewire (IEEE 1394)	no	no	no	no	no
30	Orientation sensor	yes	yes	yes	no	no
31	Uncompressed format	no	no	no	no	no
32	Format	Compact	Compact	Compact	Compact	Compact
33	Battery	specific	specific	specific	specific	specific
34	Sensor type	CCD	CCD	CCD	CCD	CCD
35	Brand	6	6	6	4	4
36	Storage types	SD/MMC/SDHC card + Internal	SD/MMC/SDHC card + Int.	SD/MMC/SDHC card + Internal	SD/MMC/SDHC card + Internal	SD/MMC/SDHC card + Internal
37	Customer Rating*	4 (11)	4 (14)	4.5 (14)	4.5 (15)	5 (7)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high);
the numbers in brackets indicate the number of customers who rated this product

Table B.3—*Continued*

		Digital Cameras				
	Attribute	Samsung NV15	Samsung S85	Samsung S860	Sony DSC-H3	Sony DSC-H9
1	Price	208 CHF	157 CHF	143 CHF	309 CHF	508 CHF
2	Resolution (Mpixel)	10.1 Mpixel	8.2 Mpixel	8.1 Mpixel	8.1 Mpixel	8.0 Mpixel
3	Optical zoom	3x	5x	3x	10x	15x
4	Digital zoom	5x	5x	3x	20x	2x
5	Display size	2.5 "	2.5 "	2.4 "	2.5 "	3.0 "
6	Storage included	20 MB	20 MB	11 BM	31 MB	31 MB
7	Video resolution	640 x 480	640 x 480	640 x 480	640 x 480	640 x 480
8	Size	97 x 60 x 19 mm	100 x 63 x 23 mm	93 x 62 x 27 mm	106 x 69 x 48 mm	110 x 83 x 86 mm
9	Weight	152 g	204 g	173 g	264 g	440 g
10	Minimum aperture size	F2.8	F2.8	F2.8	F3.5	F2.7
11	Maximum aperture size	F5.1	F4.6	F5.2	F4.4	F4.5
12	Minimum shutter speed	14 sec	15 sec	8 sec	30 sec	30 sec
13	Maximum shutter speed	1/1500 sec	1/2000 sec	1/2000 sec	1/2000 sec	1/4000 sec
14	Lowest ISO rating	80	50	80	100	80
15	Highest ISO rating	3200	1600	1000	3200	3200
16	White balance override	6 positions + manual	4 positions	5 positions + manual	7 positions	5 positions
17	Image stabilization	yes	no	no	yes	yes
18	Optical viewfinder	no	no	no	no	no
19	Manual exposure control	no	no	no	no	yes
20	Zoom wide	34 mm	38 mm	38 mm	38 mm	31 mm
21	Manual focus	no	no	no	yes	yes
22	Flip display	no	no	no	no	yes
23	Flash distance	2.4 m	2.5 m	4.0 m	7 m	9.8 m
24	Self-timer	2 or 10 sec	2 or 10 sec	2 or 10 sec	2 or 10 sec	10 sec
25	Continuous drive	yes	yes	yes	2 fps	2.2 fps
26	Built-in flash	yes	yes	yes	yes	yes
27	External flash	no	no	no	no	no
28	USB	yes	yes	yes	yes	yes
29	Firewire (IEEE 1394)	no	no	no	no	no
30	Orientation sensor	no	no	no	no	no
31	Uncompressed format	no	no	no	no	no
32	Format	Compact	Compact	Compact	Compact	SLR-like
33	Battery	specific	specific	AA	specific	specific
34	Sensor type	CCD	CCD	CCD	CCD	CCD
35	Brand	4	4	4	7	7
36	Storage types	SD/MMC/SDHC card + Int.	SD/MMC/SDHC card	SD/MMC/SDHC card + Int.	Memory Stick Duo / Pro Duo + Int.	Memory Stick Duo / Pro Duo + Int.
37	Customer Rating*	5 (4)	4 (13)	4 (7)	4.5 (12)	4 (22)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high);
the numbers in brackets indicate the number of customers who rated this product

Table B.3—*Continued*

		Digital Cameras		
	Attribute	Sony DSC-T2	Sony DSC-T50	Sony DSC-W55
1	Price	458 CHF	457 CHF	217 CHF
2	Resolution (Mpixel)	8.1 Mpixel	7.1 Mpixel	7.2 Mpixel
3	Optical zoom	3x	3x	3x
4	Digital zoom	6x	6x	6x
5	Display size	2.7 ”	3.0 ”	2.5 ”
6	Storage included	4 GB	58 MB	56 MB
7	Video resolution	640 x 480	640 x 480	640 x 480
8	Size	87 x 57 x 20 mm	95 x 57 x 23 mm	90 x 59 x 23 mm
9	Weight	156 g	160 g	147 g
10	Minimum aperture size	F3.5	F3.5	F2.8
11	Maximum aperture size	F4.5	F10	F5.2
12	Minimum shutter speed	1 sec	1 sec	1 sec
13	Maximum shutter speed	1/1000 sec	1/1000 sec	1/2000 sec
14	Lowest ISO rating	80	80	100
15	Highest ISO rating	3200	1000	1000
16	White balance override	7 positions	5 positions	5 positions
17	Image stabilization	yes	yes	no
18	Optical viewfinder	no	no	yes
19	Manual exposure control	no	no	no
20	Zoom wide	38 mm	38 mm	38 mm
21	Manual focus	no	yes	no
22	Flip display	no	no	no
23	Flash distance	3.2 m	3.4 m	3.9 m
24	Self-timer	2 or 10 sec	2 or 10 sec	10 sec
25	Continuous drive	2 fps	yes	1.1 fps
26	Built-in flash	yes	yes	yes
27	External flash	no	no	no
28	USB	yes	yes	yes
29	Firewire (IEEE 1394)	no	no	no
30	Orientation sensor	no	no	no
31	Uncompressed format	no	no	no
32	Format	Ultra Compact	Ultra Compact	Compact
33	Battery	specific	specific	specific
34	Sensor type	CCD	CCD	CCD
35	Brand	7	7	7
36	Storage types	Memory Stick Duo / Pro Duo + Internal	Memory Stick Duo / Pro Duo + Internal	Memory Stick Duo / Pro Duo + Internal
37	Customer Rating*	3.5 (3)	4.5 (3)	4.5 (10)

*available only in the LowF prototype; the scale ranges from 1 (low) to 5 (high);
the numbers in brackets indicate the number of customers who rated this product

Figure B.2: Images of the mobile phones used in Experiment 3 (in alphabetical order)

Figure B.3: Images of the digital cameras used in Experiment 3 (in alphabetical order)