Journal of Behavioral Decision Making, *J. Behav. Dec. Making* (2014)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/bdm.1847

An Optimized Design of Choice Experiments: A New Approach for Studying Decision Behavior in Choice Task Experiments

JELLA PFEIFFER[1]*, DEJAN DUZEVIK[2], FRANZ ROTHLAUF[3], ERIC BONABEAU[4,5] and KOICHI YAMAMOTO[6]
[1]*Institute of Information Systems and Marketing, Karlsruhe Institute of Technology, Karlsruhe, Germany*
[2]*Concentric, Cambridge, MA, USA*
[3]*Department of Information Systems and Business Administration, University of Mainz, Mainz, Germany*
[4]*Icosystem, Cambridge, MA, USA*
[5]*Minerva Schools, Keck Graduate Institute, San Francisco, CA, USA*
[6]*Dentsu, Tokyo, Japan*

ABSTRACT

In this paper, we present a new approach for the optimal experimental design problem of generating diagnostic choice tasks, where the respondent's decision strategy can be unambiguously deduced from the observed choice. In this new approach, we applied a genetic algorithm that creates a one-to-one correspondence between a set of predefined decision strategies and the alternatives of the choice task; it also manipulates the characteristics of the choice tasks. In addition, this new approach takes into account the measurement errors that can occur when the preferences of the decision makers are being measured. The proposed genetic algorithm is capable of generating diagnostic choice tasks even when the search space of possible choice tasks is very large. As proof-of-concept, we used this novel approach to generate respondent-specific choice tasks with either low or high context-based complexity that we operationalize by the similarity of alternatives and the conflict between alternatives. We find in an experiment that an increase in the similarity of the alternatives and an increase in the number of conflicts within the choice task lead to an increased use of non-compensatory strategies and a decreased use of compensatory decision strategies. In contrast, the size of the choice tasks, measured by the number of attributes and alternatives, only weakly influences the strategy selection. Copyright © 2014 John Wiley & Sons, Ltd.

KEY WORDS decision strategies; genetic algorithms; choice tasks; experimental design; context-based complexity

INTRODUCTION

When shopping online and choosing between products, customers usually cannot interact with employees and instead rely on the online shop to offer appropriate decision support. Thus, the success of online stores heavily depends on understanding how customers decide when choosing between alternative products. Offering the appropriate decision support tools is a challenge for designers of online stores because not all decision makers use the same decision strategy (Einhorn & Hogarth, 1981; Gigerenzer & Selten, 2001; Payne, Bettman, & Johnson, 1993). Indeed, the particular decision strategy used by a customer can vary according to the customer's personal characteristics and social context, as well as according to the characteristics of the choice task (Payne et al., 1993). Therefore, the designer of an online store does not usually know which particular decision strategy is being used by the customer.

This problem can be overcome by gaining a better understanding of people's decision behaviors and by improving the models of decision behavior. Not only can better models of decision behavior be used to develop decision support tools that facilitate the purchase process, they can also help predict customers' buying decisions (Feit, Beltramo, & Feinberg, 2010; Guo, 2010; Wright, 1975). Consequently, a number of researchers in behavioral economics, consumer psychology, and cognitive and marketing science focus on understanding adaptive decision-making behavior (Montgomery & Svenson, 1976; Hauser & Wernerfelt, 1990; Luce, Payne, & Bettman, 1999; Payne, 1976; Shugan, 1980; Gigerenzer & Selten, 2001; Bettman, Johnson, Luce, & Payne, 1993; Swait & Adamowicz, 2001).

Up until now, two standard methods have been used to study decision making: *procedural methods* and *outcome-based* methods. Procedural methods infer the decision strategy used from the sequence of actions performed by the decision makers during the decision process. However, these procedural methods often fail at determining what particular strategy was being used (Bröder & Schiffer, 2003; Ford, Schmitt, Schechtman, Hults, & Doherty, 1989; Ball, 1997; Rieskamp & Hoffrage, 1999; Reisen, Hoffrage, & Mast, 2008). Recently, some authors have offered an approach to overcome this problem (see Johnson, Schulte-Mecklenbeck, and Willemsen, 2008, for an example). In contrast to procedural methods, outcome-based methods define a few decision strategies beforehand and search for the decision strategy that best explains the final decision. However, these methods do not differentiate well between strategies because several different strategies can often lead to the same choice (Bröder, 2000; Bröder & Schiffer, 2003; Rieskamp & Hoffrage, 2008; Glöckner, 2009; Rieskamp & Hoffrage, 1999). Creating choice tasks where the outcomes of the decision strategies sufficiently differ from one another is known as the diagnostic task selection problem (Jekel, Fiedler, & Glöckner, 2011). This problem has gained increasing interest in the past years.

The diagnostic task selection problem is an optimal experimental design problem. In optimal experimental design

*Correspondence to: Jella Pfeiffer, Institute of Information Systems and Marketing, Karlsruhe Institute of Technology, Karlsruhe, Germany. E-mail: jella.pfeiffer@kit.edu

Copyright © 2014 John Wiley & Sons, Ltd.

problems, different goals can be pursued, such as choosing optimal treatment levels and/or improving the efficiency by generating stimuli that minimize the variances and covariances of the parameter estimates (Atkinson & Federov, 1992; Kuhfeld, Tobias, & Garratt, 1994). Improving the efficiency of experimental designs has been of great interest to researchers working on linear and generalized linear choice models (e.g., Kuhfeld et al., 1994; Goos, Vermeulen, & Vandebroeak, 2010), or on nonlinear choice models (Myung & Pitt, 2009; Cavagnaro, Myung, Pitt, & Kujala, 2010), for example, when discriminating generalized expected utility models and probability weight functions in the case of risky choices (Cavagnaro, Gonzalez, Myung, & Pitt, 2013; Cavagnaro, Pitt, Gonzalez, & Myung, 2013). However, when the number of possible stimuli is very high and the experimental design problem cannot be solved analytically, solving the task selection problem becomes computationally very challenging. Myung and Pitt (2009) addressed this problem and proposed a sequential Monte Carlo search for generating random samples of the potential stimuli using simulated annealing. Their method searches for an experimental design that best differentiates between the choice models under consideration. The drawback of their method is that it can only differentiate between two models and it requires high-dimensional integration. Furthermore, the effects of the design variable must be expressed in explicit functional form, which typically does not allow the model to be applied to problems where the relationship between design and user behavior is unclear or inferred from experiments. Consequently, it is difficult to systematically test the influences of, for example, choice task characteristics because there is no functional relationship between choice task characteristic and outcome; furthermore, many observations per subject are needed.

In this paper, we propose an optimal experimental design approach to solve the diagnostic task selection problem. The properties of our approach are the following:

1. Assuming a predefined set of possible decision strategies, the approach creates efficient designs by constructing choice tasks in such a way that each available alternative is consistent with only one possible decision strategy. It is thus possible to identify the strategy used by a decision maker because the decision strategy can be unambiguously deduced from a single choice decision.
2. The choice tasks are designed in such a way that the unambiguous assignment of alternatives to decision strategies is robust with respect to small modifications of the choice tasks.
3. In addition to ensuring an efficient design, our optimal experimental design method allows optimal treatment levels to be chosen: It manipulates the characteristics of the choice tasks (e.g., the complexity of the choice task). This makes it possible to systematically create respondent-specific choice tasks that are maximally different in terms of predefined characteristics chosen by the researcher.
4. The approach is fast enough to be used for scenarios where choice tasks with high diagnosticity are individually generated for each respondent.

The closest work to ours is the Euclidean Diagnostic Task Selection (EDTS) approach by Jekel et al. (2011). In addition to the observation of the outcome (choice decision), the authors include two behavioral measurements that allow them to differentiate between different decision strategies even when several strategies explain the same choice. They generated choice tasks with maximal diagnosticity by creating all possible choice tasks and choosing the ones that had the lowest overlap of strategies and had large differences in the two following behavioral measurements: predicted times of the competing strategies and perceived confidence levels of these strategies. Because Jekel et al.'s approach relies on the generation and evaluation of all possible choice tasks, the computational costs are high for scenarios where the search space of possible choice tasks is large. Large search spaces are not uncommon, given that already in scenarios with just four alternatives, five attributes, and four attribute levels, there are about $4^{20} \approx 10^{12}$ potential choice tasks (including the symmetric ones). The problem of high computational costs further compounds when respondent-specific choice tasks are needed and/or additional treatment level characteristics of the choice tasks are manipulated—two scenarios that were not considered by Jekel et al. (2011) in their exact algorithm. Therefore, we suggest a genetic algorithm (GA) that is fast and produces choice tasks with a high diagnosticity, although it does not guarantee finding the optimal experimental design. Indeed, it can generate as many alternatives per choice task as there are strategies to be distinguished from one another and does not rely on additional behavioral measurements. Nevertheless, if needed, our approach can easily take into account behavioral measurements as well.

We will demonstrate the applicability and benefits of the new approach in an experimental study where the search space of potential choice tasks is large and where respondent-specific choice tasks with high and low complexity are generated, respectively. The complexity of a choice task is described by two different aspects: the attribute variance (which measures the similarity of alternatives) and the pairwise inter-attribute correlation of the utilities (which measures the number of conflicts in a choice task) (Bettman et al., 1993; Luce, Bettman, & Payne, 1997). These two context-based complexity measures are highly relevant in real-world buying decisions where a large product variety often forces the user to choose between similar high-quality products whose features have to be traded off against each other (Fasolo, Carmeci, & Misuraca, 2009). Despite the practical relevance of these complexity measures, few studies analyze the influence of these measures on decision-making behavior (Scheibenhenne, Greifeneder, & Todd, 2010; Pfeiffer et al., 2014). In order to fill this gap in the research, we formulated two hypotheses stating that both an increase in similarity and an increase in number of conflicts lead to an increase in the use of non-compensatory, low-effort strategies and a decrease in the use of compensatory, high-effort strategies. In our experimental design, we considered two strategies that fall into each of these two groups and tested the hypotheses in an experimental study using choice tasks generated by the new approach. The empirical findings confirmed the two hypotheses, and the results showed that the two context-based complexity measures influence strategy selection more

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

than the number of alternatives or attributes does. This latter finding is a new contribution to the decision-making literature.

The paper is structured as follows. In the next section, we formally describe decision strategies, discuss several approaches to measure decision behavior, and elaborate on the problem of task diagnosticity. Then, we introduce our optimal experimental design approach and point out how it differs from related approaches. In the next section, we evaluate our approach by showing how it generates choice tasks with high diagnosticity and also with optimal treatment levels. Then, we present a proof-of-concept: We show that our approach generates an experimental design that yields results that can contribute to literature about the influence of complexity on choice behavior. We end the paper with a conclusion and an outline of limitations and avenues for future research.

DECISION BEHAVIOR

Decision strategies

In a choice task c, decision makers select product p from a set of n product alternatives $P = \{p_1, \ldots, p_n\}$. All alternatives are characterized by the same set of m attributes $A = \{a_1, \ldots, a_m\}$. Given the set L_i of possible attribute levels, each attribute a_i can take a value $\ell \in L_i$. For all attributes, the function $\alpha_i(p)$, where $\alpha_i : P \to L_i$, assigns an attribute level to each product $p \in P$. As an example, we can characterize cell phones by a set of different attributes, such as price, brand, and battery runtime. For all n different cell phones, $\alpha_{\text{brand}}(p)$ assigns an attribute level (e.g., Nokia, Siemens, and Motorola) to the "brand" attribute of phone p. Likewise, $\alpha_{\text{price}}(p)$ and $\alpha_{\text{batteryruntime}}(p)$ assign an attribute level to the "price" and "battery runtime" attributes of phone p.

A decision maker's preferences are characterized by several properties such as the importance $w(a_i)$ of an attribute a_i, the utility $u(l)$ of an attribute level $l \in L_i$, or the aspiration function $s(a_i)$ of an attribute. The importance $w(a_i) : A \to \mathbb{R}$ determines how important an attribute a_i is for the choice decision. The utility $u(l) : L_i \to \mathbb{R}$ determines the utility of the different attribute levels $l \in L_i$ for the decision maker. The aspiration function $s(a_i) : A \to \cup L_i$ assigns the decision maker's unacceptable attribute levels to each attribute a_i. Decision strategies describe the behavior of a decision maker when choosing an alternative from a set P of product alternatives. Common examples of decision strategies are as follows (Glöckner, 2009; Riedl, Brandstätter, & Roithmayr, 2008):

1. Weighted Additive Difference (WAD) (Fishburn, 1970);
2. Simple Majority Decision Rule (MAJ) (Arrow, 1951);
3. Lexicographic Heuristic (LEX) (Tversky, 1969); and
4. Elimination by Aspect Strategy (EBA) (Tversky, 1972).

Weighted Additive Difference is the classic utility-maximizing strategy (Anderson, 1974; Payne et al., 1993). Given the utility U of an alternative p as

$$U(p) = \sum_{i=1}^{m} w(a_i) \cdot u(\alpha_i(p)), \quad (1)$$

a decision maker applying WAD selects the alternative with the highest utility $max_{p \in P}(U(p))$. The utility of an alternative is the sum of the weighted utilities corresponding to its attribute levels.

Simple Majority Decision Rule is a majority decision rule. For each alternative $p \in P$, a decision maker counts the number M of attributes where the attribute level utilities are equal to or higher than the attribute level utilities of the other alternatives as

$$M(p) = |\{a_i \in A | u(\alpha_i(p)) \geq u(\alpha_i(q)), \forall q \in P, q \neq p\}| \quad (2)$$

A decision maker applying MAJ chooses the alternative with the highest number of best attribute levels $max_{p \in P}(M(p))$.

LEX considers attributes in decreasing order of importance and chooses the alternative whose attribute level utility is highest for the most important attributes. Formally, we order the m different attributes in a vector $A^w = (a_1, \ldots, a_m)$ such that $w(a_i) \geq w(a_{i+1})$. Given the utility vector $v_i = (u(\alpha_1(p_i)), u(\alpha_2(p_i)), \ldots, u(\alpha_m(p_i)))$ of an alternative $p_i \in P$, we order $V = (v_1, v_2, \ldots, v_n)$ lexicographically in decreasing order. Then, LEX selects the first element of the lexicographically ordered vector V.

The deterministic version of EBA considers attributes in decreasing order of importance and removes alternatives from the set of product alternatives if their attribute level is unacceptable.[1] The strategy stops if there is only one alternative left or if all the attributes have been considered (Tversky, 1972; Montgomery & Svenson, 1976; Thorngate, 1980; Yee, Dahan, Hauser, & Orlin, 2007). With A^w ordering the m attributes earlier ($w(a_i) \geq w(a_{i+1})$) and $P_i = \{p \in P | \alpha_i(p) \in s(a_i)\}$, we define $P'_i = \{p \in P | p \in P_j, \forall j \leq i\}$. P_i contains all products that are unacceptable on the ith most important attribute; P'_i contains all products that are unacceptable on at least one of the i most important attributes. Let $\max \in \{1, \ldots, m\}$ be the largest index such that $P'_{\max} \neq P$; EBA selects the product(s) $P - P'_{\max}$.

All four strategies have a long history in decision-making research. WAD (Tversky, 1969; Fishburn, 1970; Montgomery & Svenson, 1976) and MAJ (Inada, 1964; Arrow, 1951) are two examples of compensatory decision strategies where the low level of a product attribute can be compensated by the high level of another attribute. WAD is a weighted utility maximization model that is often the assumed normative model of decision making (Anderson, 1974; Dawes, 1979; Payne et al., 1993), while MAJ is a strategy that maximizes the number of dominant attribute levels (Riedl et al., 2008). Both strategies use all the information presented in a choice task. Thus, the computational effort of WAD and MAJ is high, because a decision maker must consider all the attribute levels. For WAD, she must in addition weight the attributes and assign utilities to the attribute levels. By contrast, LEX and EBA are non-compensatory, simplifying, low-effort strategies that use only part of the information in a choice task (Tversky, 1969, 1972; Fishburn, 1974). Such strategies often iteratively focus on one attribute and reduce the number of available alternatives without taking into account the other attributes. For example, a low attribute level utility can immediately exclude an

[1] The original version of the EBA is stochastic and considers attributes in a probabilistic order that depends on the importance of the attributes.

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

alternative from a choice task, independently of the attribute levels of its other attributes. Thus, non-compensatory strategies are selective in their information search and often require little cognitive effort (Rieskamp & Hoffrage, 2008). It has been shown that simple decision strategies can be very effective and yield decisions of high quality (Hogarth & Karelaia, 2005; Åstebro & Elhedhli, 2006).

Measuring decision behavior

To explain and determine the strategies applied by decision makers, researchers use two common approaches in empirical studies of adaptive decision making (Harte & Koele, 2001). First, *procedural* methods infer the decision strategies used by individuals by observing the sequence of actions they follow before stating their final choice (Ford et al., 1989; Einhorn & Hogarth, 1981). To this end, studies use verbal protocols(Biggs, Bedard, Gaber, & Linsmeier, 1985; Bettman & Park, 1980; Ball, Langholtz, Auble, & Sopchak, 1998; Crow, Olshavsky, & Summers, 1980), eye tracking (Russo, 1978; Russo & Dosher, 1983; Russo & Leclerc, 1994; Reisen et al., 2008; Pfeiffer et al., 2014), and computerized process-tracing tools (Cook & Swain, 1993) such as MouseLab (Bettman et al., 1990) and extensions of MouseLab (Jasper & Shapiro, 2002; Schulte-Mecklenbeck, Kuehberger, & Ranyard, 2011). For an overview of process-tracing methods, the reader can refer to Schulte-Mecklenbeck, Murphy, and Hutzler (2011).

Second, *outcome-based* methods can be divided into *structural approaches* and *comparative model fitting* (Bröder & Schiffer, 2003). Structural approaches use formal models that represent the relationship between the attribute level utilities of the alternatives in a choice task and the final choice made. These approaches search for a single, parsimonious choice model that maximizes the likelihood of correctly predicting the respondent's final choice (Harte & Koele, 2001; Feit et al., 2010; Tsetlin & Winkler, 2009; Ding, Park, & Bradlow, 2009). These structural approaches often use a linear additive utility function (equivalent to the WAD strategy). The advantage of fitting the parameters of a linear additive utility function is that other strategies, like LEX, can be modeled as sub-models of WAD if the utility functions fulfill certain conditions (Martignon & Hoffrage, 2002; Kohli & Jedidi, 2007).

Comparative model fitting is a more recent approach in which several models compete for the best explanatory power for a given data set (Dhami & Ayton, 2001; Hoffrage, Hertwig, & Gigerenzer, 2000; Gilbride, & Allenby, 2004; Garcia-Retamero, Hoffrage, & Dieckmann, 2007). Parameters of the competing models are defined beforehand, and maximum likelihood methods are used to compare the final choices with the predictions made by the models (Glöckner, 2009). Thus, in contrast to structural approaches, different decision strategies can be assumed and compared when using comparative model fitting. Bröder and Schiffer (2003), for instance, used a maximum likelihood approach to assess whether a simplified variant of LEX, WAD, or an equal weight linear model provides the best likelihood estimation for the observed decision behavior. Recently, other authors have also developed approaches to identify non-compensatory strategies, including EBA (Hauser, Toubia, Theodoros, Silinskiai, & Berfurt, 2010; Kohli & Jedidi, 2007; Gilbride & Allenby, 2006; Rieskamp & Hoffrage, 1999).

Both procedural and outcome-based methods have limitations. Procedural methods are able to follow the attention paths of respondents with high precision and uncover whether the decision is compensatory versus non-compensatory or whether the choice process is attribute-wise versus alternative-wise. However, these insights often reveal only the general category of choice processes, rather than the particular strategy used (Bröder & Schiffer, 2003; Ford et al., 1989; Ball, 1997; Rieskamp & Hoffrage, 1999). This leads to problems when competing strategies do not differ in terms of the predicted decision processes (Glöckner, 2009). Given that many different decision strategies can fit into the same general category of choice process, and given that each individual strategy can lead to a different final choice, the usefulness of procedural approaches diminishes as the number of analyzed strategies increases (Payne, Braunstein, & Carroll, 1978). In the case where the predictive accuracy of one strategy is to be tested, however, process tracing is useful for showing that a particular strategy does not explain the observed patterns (Johnson et al., 2008). Another limitation of using verbal protocols and tracing mouse clicks is the potential influence of the decision process (Lohse & Johnson, 1996; Glöckner & Betsch, 2008; Reisen et al., 2008). In addition, less intrusive methods such as eye tracking can usually only be applied in laboratory experiments.

Structural approaches are constrained by a singular, a priori chosen model and rarely evaluate multiple decision strategies at the same time; an exception are consider-then-choose that assume the customer first screens all the alternatives, then applies a compensatory process to finalize his or her choice (Gilbride & Allenby, 2006; Liu & Neeraj, 2011). Comparative model fitting can include various competing choice models but suffers greatly from the task selection problem (Bröder, 2000; Bröder & Schiffer, 2003; Rieskamp & Hoffrage, 1999, 2008; Myung & Pitt, 2009). Glöckner (2009) addressed this problem by including two behavioral measures into the maximum likelihood estimation, namely decision time and the participant's reported confidence judgments. Classification improved when the differences in decision time and confidence for the predicted strategies were sufficiently large. Therefore, Glöckner (2009) suggested to generate choice tasks where large differences in terms of the mapping of strategies to alternatives, the decision times, and confidence can be expected.

Jekel et al. (2011) took up this idea and presented an approach for generating diagnostic choice tasks: the EDTS. For two alternatives and four attributes, the authors generated all possible choice tasks with binary attribute levels (+and −) resulting in 40 different choice tasks. Then, they created hypothetical decision makers with different preferences. For each decision maker, each choice task, and each strategy, they simulated the choice decision, decision time, and confidence judgments. They also allowed for different error rates for choices and noise levels for decision times and confidence judgments. Choice tasks with high diagnosticity are the ones where the application of one strategy leads to a different choice, a different decision time, and a different user confidence than the application of other strategies. Hence,

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

diagnosticity is measured by the average Euclidean distance between all possible pairs of strategies, in the normalized three-dimensional space of the predicted chosen alternative, decision time, and confidence. The authors found that their method increases the proportion of correct classifications in comparison with the random sampling of choice tasks.

AN OPTIMIZATION APPROACH CONNECTING DECISION BEHAVIOR AND DECISION STRATEGIES

We address the diagnostic task selection problem by constructing choice tasks in such a way that selected alternatives are consistent with only one out of a predefined set of decision strategies. In this case, no other behavioral measurements are needed for discriminating between different strategies. Furthermore, we allow for heterogenous functions, that is, respondent-specific preference functions, and take into account measurement errors related to these functions.

Objectives of the approach

To design respondent-specific choice tasks, a two-step approach is mandatory: In the first step, respondents' preferences need to be collected. In the second step, the respondent-specific choice tasks need to be presented to the respondents and their decision strategies must be recorded. In order to create individual and preference-specific choice tasks where each alternative is consistent with one predefined decision strategy, optimization techniques need to be applied.

In order to link the respondent's choice with a decision strategy, data concerning the respondent's preferences need to be collected. For example, if the decision strategies under consideration are WAD, MAJ, LEX, and EBA, the following measures need to be collected: the utility $u(l)$ that the decision maker assigns to each attribute level l (necessary for WAD, MAJ, and LEX), the importance $w(a_i)$ of each attribute a_i (WAD, LEX, and EBA), and the aspiration function of each attribute $s(a_i)$ (EBA). If other decision strategies are under consideration, other aspects of the respondents' preferences would have to be measured.

In our experiment, the choice tasks presented to a respondent needed to be designed in such a way that observing the respondent's choice would allow us to determine the decision strategy used. Therefore, we wanted to achieve a bijective mapping between a set of predefined decision strategies and alternatives. We denote a mapping to be *optimal* if each strategy out of a set of predefined strategies corresponds to exactly one alternative. Figure 1 shows two different respondent-specific choice tasks, which differ in terms of their attribute levels. The respondent-specific choice task A is an example of a non-optimal mapping. It is a non-injective (and non-surjective) mapping from the set of decision strategies to the alternatives. In our example, the respondent will select phone A if he or she follows WAD or MAJ. We will denote such cases, where two or more strategies map to the same alternative, as ambiguous mapping. Following EBA leaves the respondent with phone B or C. We will denote such cases, where a decision strategy leads not only to one but to several alternatives as multiple mapping. Keeping in mind that the choice of different attribute levels leads to different mappings, all choice tasks that do not lead to ambiguous or multiple mappings are denoted as *choice tasks with optimal mapping*. An example of an optimal mapping is provided in choice task B in Figure 1. Choice task B was designed in such a way that the respondent would select phone A if he or she followed a WAD strategy. Similarly, if the respondent followed a MAJ, EBA, or LEX strategy, he or she would select phone B, phone C, or phone D, respectively. If a respondent selected phone E, he or she was not following any of the four strategies. We denoted this case as OTHER.

Respondent-specific choice tasks with optimal mapping depend on the preferences of the respondents. However, when measuring their preferences, measurement errors are inevitable. To reduce the effects of errors, we wanted to ensure that choice tasks were *robust*, which means that small changes in the attribute values of a choice task would not strongly increase the number of ambiguous and multiple mappings.

Finding robust choice tasks with optimal mapping is difficult because the number of different choice tasks is high even in scenarios with a small number of alternatives and attributes. For a choice task with n alternatives and m attributes, where the

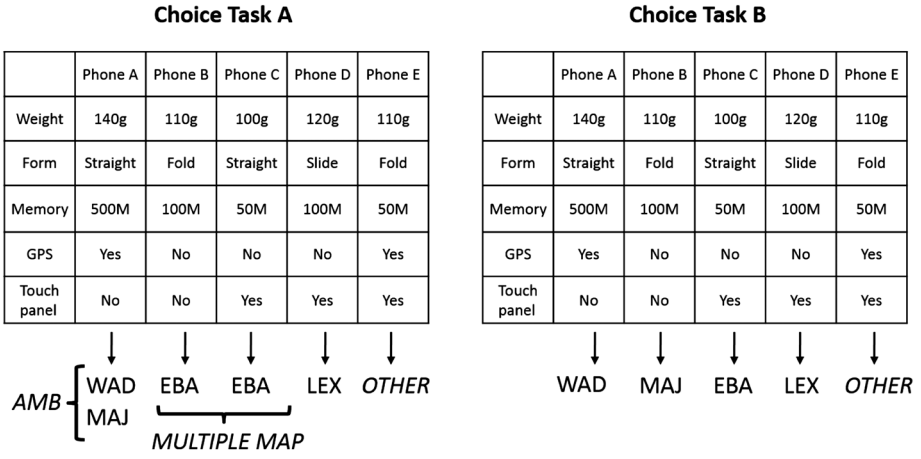


Figure 1. Two example choice tasks. EBA, Elimination by Aspect Strategy; LEX, Lexicographic Heuristic; MAJ, Simple Majority Decision Rule; WAD, Weighted Additive Difference; AMB, ambigious mapping

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

m attributes are chosen from a set of m_{all} attributes and each attribute has on average $|L_i|$ possible attribute levels, we have $|L_i|^{nm} \cdot \binom{m_{\text{all}}}{m}$ different choice tasks. As an example, for a choice task with $n=5$ alternatives, $m=5$ attributes that are chosen out of a set of $m_{\text{all}}=15$ possible attributes, and $|L_i|=3$ attribute levels per attribute, we have $2.5*10^{15}$ different choice tasks. This high number of possible choice tasks suggests that we must use optimization methods that find robust choice tasks with optimal mapping. Just applying exhaustive search methods is no choice as the problem is computationally intractable.

An optimization approach for finding robust choice tasks with optimal mapping

Genetic algorithms (Holland, 1975; Goldberg, 1989; Reeves & Rowe, 2003) are heuristic optimization methods that are inspired by the principles of evolution and apply recombination, mutation, and selection operators to a population of solution candidates. In contrast to exact optimization methods such as linear programming or branch-and-bound approaches, heuristic optimization methods do not guarantee an optimal solution but often return high-quality solutions in a reasonable amount of time (Bäck, Fogel, & Michalewicz, 1997; Glover & Kochenberger, 2003). Heuristic optimization methods are often applied with success to $\mathcal{NP}$—hard problems or problems where the size of the search spaces increases exponentially with the problem size.

Algorithm 1 Functionality of genetic algorithm

Create initial population P
Evaluate initial population
Update archive
while termination criterion is not met **do**
  Select high-quality solutions from P and obtain P'
  Apply recombination to P' and obtain P''
  Apply mutation to P'' and obtain P'''
  Evaluate P'''
  $P = P'''$
  Update archive
end while
Choose best solution from archive

Figure 2 and Algorithm 1 describe the basic functionality of the GA. After randomly creating and evaluating an initial set (population P) of candidate solutions (step 1), the algorithm evaluates the candidate solutions according to their ability to solve the optimization problem (in our case the diagnosticity task selection problem, step 2). The GA keeps an archive separately from the population. This archive stores the best solutions observed during the GA run whenever a new population is evaluated (step 3). Then, the GA iteratively creates new populations P''' by recombining (with probability p_c, step 5) the selected high-quality solutions (step 4) and applying mutations (with probability p_m, step 6). The iterative process stops when a termination criterion is met. After termination, the GA returns the archive.

For each candidate solution, the GA stores all information relevant to the solution of the optimization problem. In our case, the candidate solution was a vector containing all the information necessary to the construction of a choice task. Thus, a candidate solution had to define which m attributes and corresponding nm attribute levels were to be shown to a particular respondent. Figure 3 gives an example of how a choice task is constructed from a candidate solution (for a more formal technical description, please refer to APPENDIX A). In our example with $n=5$ alternatives and $m=3$ attributes, a candidate solution representing a choice task has $m(n+1)=18$ elements. The first (x_1), seventh (x_7), and 13th (x_{13}) elements of the candidate solution indicate the shown attributes. The other elements determine the15 attribute levels used. In particular, the first element x_1 of a solution specifies the attribute that is shown in the first row of the choice task. In our example, we assume an attribute vector $A^w=(form, weight, price)$, which is ordered with respect to the importance of the attributes. A value of $x_1=2$ indicates that the attribute at the second position of A^w is selected. The following $n=5$ elements $x_2,\ldots,x_6$ indicate the shown attribute levels for the $n=5$ different alternatives. We assume that the vector $L^w_{weight}=(100\text{g}, 110\text{g}, 120\text{g}, 130\text{g}, 140\text{g})$ contains the possible attribute levels of the attribute "weight," where the attribute levels are ordered with respect to the respondent-dependent utility. Then, a value of $x_2=5$ indicates that we use the attribute level at the fifth position of L^w_{weight} for the weight of phone A, for example, 140 g.

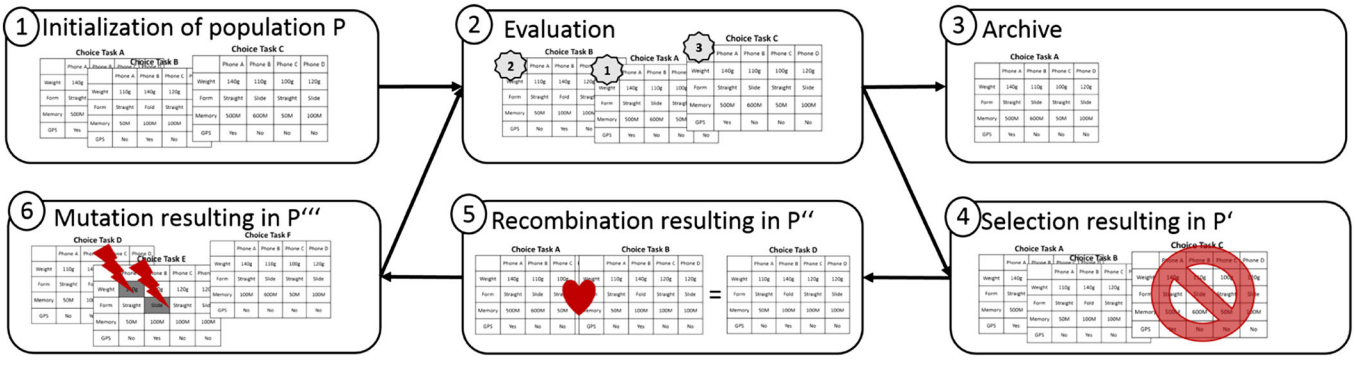


Figure 2. Basic functionality of the genetic algorithm

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

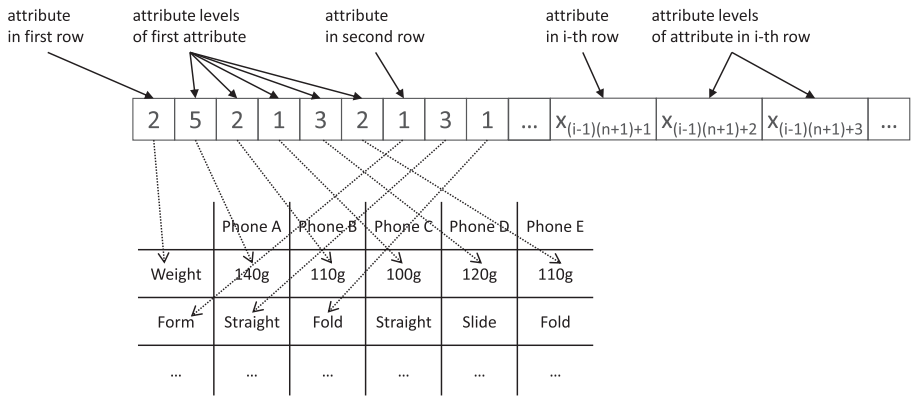


Figure 3. Construction of a choice task from a candidate solution

This construction process ensures that the search space is fully covered and that all possible choice tasks can be represented. The construction process follows the principle of problem space search (Storer, Wu, & Vaccari, 1992), where the values in the candidate solution vector select elements from a pre-ordered list. By selecting elements (attributes and attribute levels) from an ordered list, small variations of the vector result in new solutions with similar properties (attributes and attribute levels have similar importance and utility, respectively). Such a construction process often leads to a high performance of the heuristic optimization methods (Storer et al., 1992; Raidl & Julstrom, 2000; Rothlauf, 2009).

Population size is a parameter that the user chooses. Higher population sizes usually lead to better GA performance but increase running time. For our application of the algorithm, we decided that each population of the GA should consist of 300 individuals (choice tasks). In preliminary experiments, a population size of 300 turned out to be a good compromise between solution quality and running time. The candidate solutions of the first, initial population were created randomly, as described in APPENDIX A.

After the initial population of candidate solutions has been created, the GA iteratively creates and evaluates the population of candidate solutions in order to generate robust choice tasks with optimal mapping. For this purpose, we defined an objective function $f(c)$, which measures the task diagnosticity, that is, the quality of the mapping (step 2). Given a set of decision strategies S, we wanted to create choice tasks with a minimum number of multiple and ambiguous mappings. Thus, when evaluating a solution, we considered only non-multiple strategies $s_{\text{nonmult}} \subset S$ and calculated f_1 as $\sum_{s_{\text{nonmult}} \in S} 1/n_{\text{nonmult}}$, where n_{nonmult} denotes the number of strategies that map to the same alternative as strategy s_{nonmult}. For the example choice task A and $S = \{\text{WAD,MAJ,LEX,EBA}\}$ (Figure 1), there are no multiple mappings and we obtain $f_1 = 1 + 1 + 1 + 1 = 4$. Example B has a multiple mapping for EBA; WAD, and MAJ map to the same choice (are ambiguous). Thus, we obtain $f_1 = 1 + 1/2 + 1/2 = 2$, because EBA does not contribute to the fitness, while WAD and MAJ contribute 1/2 and LEX contributes 1.

The objective function can be extended to consider additional desired properties of a choice task. A straightforward approach is to assume an additive objective function $f(c) = \sum_i f_i$, where the different components f_i measure not only the quality of the mapping but also other desired properties such as how well treatment levels are met (for an example, see the *Evaluation* section, where the complexity of the choice tasks was manipulated). Other behavioral measurements such as the expected time of the decision or the predicted confidence can also be included as additional components of the fitness function. With respect to the expected decision time, we could, for example, define an additional fitness component f_t as the variance of the decision times that a user would need when applying decision strategies. The algorithm would then try to maximize this variance, which means that it would try to generate choice tasks with very different decision times dependent on which strategy is used.[2]

The GA applies standard selection and variation operators for steps 4 to 6 (Bäck et al., 1997, Chap. C3). Selection operators select high-quality solutions and ensure a continuous improvement of solutions in the population over the generations. We used a tournament selection of size 2. Thus, we iteratively selected two random candidate solutions from P and inserted the solution with higher objective value f into P' until P' was filled. The variation operators, recombination and mutation, ensure that either new candidate solutions are created by recombining candidate solutions (recombination) or by slightly modifying existing solutions (mutation) (see the *Recombination and Mutation* section in the appendix for details). The GA terminates after 600 generations (step 7) because improvements of the fitness become marginal and, in most cases, further improvements are not made, or the population has converged to an optimum. Furthermore, the GA uses an archive of size 300.

At the end of a run, the GA returns the archive that contains the 300 best solutions found. In order to address the robustness of a choice task and to reduce the effects of estimation errors stemming from inaccurate measurements of the preference function, we considered all 300 solutions in the archive and examined the effects of small attribute level modifications. We took each solution c and randomly modified one of the attribute levels shown. We did this

[2]Decision times can be predicted by measuring the decision effort with elementary information processes as defined by Johnson and Payne (1985) and described in detail by Pfeiffer and Scholz (2013).

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

50 times for each of the 300 choice tasks and computed the *robust objective value* $f'(c)$ as

$$f'(c) = \frac{1}{50}\sum_{i=1}^{50} f(c_i) \quad (3)$$

where the c_i are the 50 solutions that were created by randomly modifying one attribute level of c. The optimization approach finally returns the solution (choice task) that maximizes f'.

Related work

Jekel et al. (2011) presented the EDTS approach that also solves the task selection problem. However, it differs from ours in several aspects. First, Jekel et al. (2011) assumed homogeneous preference functions where the attribute level utilities are the same for all decision makers. Consequently, the diagnostic choice task problem is solved only once for a standard user. In contrast, our suggested approach assumes heterogenous preference functions for each respondent, making it necessary to solve the diagnostic task problem individually for each respondent. However, both their approach and our approach could easily be changed to mimic each other's assumptions of homogeneous versus heterogeneous preference functions. In particular, the approach of Jekel et al. (2011) could follow a two-step approach, where in the first step, respondent-specific preferences are measured. Likewise, our approach could easily be changed to mimic their approach by assuming preferences to be identical across respondents.

Another difference between Jekel et al.'s approach (2011) and ours is that Jekel et al. did not only observe choices, like we did; they also observed additional behavioral measurements. Additional behavioral measurement that can, for example, also be recorded with procedural methods, have the advantage to be able to help identify the applied strategy in case of multiple mappings of strategies to alternatives. To avoid multiple mappings if only choices were observed, larger choice tasks with at least as many alternatives as decision strategies would be needed. Large choice tasks with a higher number of alternatives, however, would be computationally demanding for their approach.

Third, EDTS is a very valuable approach for small search spaces. In their example, Jekel et al. (2011) searched for choice tasks with highest diagnosticity among a set of $2^8 = 256$ choice tasks.[3] By contrast, our approach can deal with much larger search spaces (e.g., $2.21 * 10^{34}$; see the *Evaluation: Generating Robust Choice Tasks with Optimal Mapping* section). Jekel et al. (2011) proposed an algorithm that guarantees to find the choice tasks with optimal diagnosticity (according to their Euclidean distance criteria), while, because of the sheer size of our search space, our GA optimizes for diagnosticity but cannot guarantee finding the optimal solution.

In sum, our approach nicely complements EDTS because it is suited for more complex choice tasks (more alternatives, more attribute, and more attribute levels) where optimal methods such as EDTS are unfeasible because of high computational costs.

EVALUATION: GENERATING ROBUST CHOICE TASKS WITH OPTIMAL MAPPING

We evaluated different aspects of our approach: We evaluated how well the approach could generate diagnostic choice tasks with optimal mapping, and we evaluated an extension of our approach, where we also tried to find optimal treatment levels. In particular, we generated an optimal experimental design that manipulated the complexity of choice tasks and evaluated two aspects of this extended version. First, we evaluated whether the GA is capable of not only ensuring an optimal mapping but also allows finding optimal treatment levels. Second, as proof-of-concept, we studied whether the GA is able to generate an experimental design that yields results that can contribute to literature about the influence of complexity on choice behavior.

As input for the evaluations, we use preference functions that we measured in an online experiment in Japan in 2010 with a pool of 301 participants selected by a Japanese advertising agency.[4] All participants were between 30 and 39 years old, 60.5% of them were female, and all owned a cell phone. For each participant, we measured the utility $u(l)$ and importance $w(a_i)$ of each attribute a_i using an adaptive conjoint analysis for 15 attributes that define cell phones. For the aspiration functions $s(a_i)$, we used a questionnaire asking the respondents which attribute levels they found unacceptable. Table 1 lists the attributes and attribute levels that were used in the study. They were specified by the Japanese advertising agency and viewed as most relevant to their customers.

First, we wanted to evaluate how well the GA can generate choice tasks with optimal mapping. Using respondent's preferences, we generated robust choice tasks with an optimal mapping. We assumed that the users followed one of the four strategies WAD, MAJ, LEX, and EBA, or some other strategy (OTHER). To assess the performance of the GA approach, we compared it to a benchmark where choice tasks were randomly sampled (denoted as *random sampling*). In order to increase the validity of the evaluation, we created choice tasks of different sizes with all possible combinations of five versus eight alternatives and attributes, respectively.

For the random sampling benchmark, for each of the choice tasks that were generated for each respondent, we generated 180 000 choice tasks by selecting random attributes and corresponding attribute levels. Analogously to the GA, we stored the 300 best solutions (with respect to f_1) in an archive and selected the most robust solution (with respect to f') from this set. The computational effort of this approach was high and about equal to the GA approach, which also performs 180 000 evaluations during a run (600 generations and 300 solutions) and 15 000 evaluations at the end of

[3]After the elimination of symmetric choice tasks, this number decreases to 40.

[4]Later on, this same group of participants was used in the proof-of-concept experiments.

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

Table 1. Attributes and attribute levels used in the experiment

Attribute	Attribute levels				
Sales rank	#1	#3	#5	#10	#20
Camera resolution	10 Mpx	8 Mpx	5 Mpx	3 Mpx	2 Mpx
Display	3.4″	3.2″	3″	2.8″	2.7″
Weight	100 g	110 g	120 g	130 g	140 g
Brand	NEC	Panasonic	Sharp	Fujitsu	
Form	Folding	Slide	Straight		
Thickness	10 mm	15 mm	20 mm		
iMode	iMode phone	Non-iMode	iMode smart-phone		
Memory	500 MB	100 MB	50 MB		
Continued use	300 minutes	240 minutes	180 minutes		
Wallet function	Yes	No			
GPS	Yes	No			
Touch panel	Yes	No			
Waterproof	Yes	No			
One seg TV	Yes	No			

the run (50 modifications for each of the 300 solutions in the archive). Table 2 compares mean and standard deviation (in brackets) of f', f_1, and the percentage of choice tasks with optimal mapping. The GA approach returns more optimal and more robust choice tasks than random sampling. Using the GA, 97.1% of the choice tasks are optimal (in comparison with 55.8% for random sampling). The worst choice task returned by the GA had an objective value of $f_1 = 1.50$, in contrast to the worst solution found by random sampling with $f_1 = .75$, where $f_1 \in [0, 4]$.

To conclude, our approach efficiently solves the task selection problem, even in scenarios with huge search spaces (in the largest choice tasks, with eight alternatives and eight attributes, the size of the search space was $|L_i|^{nm} \cdot \binom{m_{\text{all}}}{m} = 3^{64} \cdot \binom{15}{8} = 2.21 * 10^{34}$).

EVALUATION: OPTIMAL MAPPING AND MANIPULATION OF CONTEXT-BASED COMPLEXITY

To evaluate whether the GA also finds choice tasks with optimal treatment levels, we focus on manipulating the complexity of choice tasks. The complexity of a choice task describes how difficult it is for the decision maker to make a decision (Payne et al., 1993; Ford et al., 1989). It is influenced by several characteristics of the choice task. Common and well-studied characteristics are the number of alternatives and attributes shown to the decision maker (Scheibenhenne et al., 2010; Chernev, Böckenholt, & Goodman, 2010).

The difficulty of a choice task also depends on its *context-based complexity*. Measures for context-based complexity are respondent specific: They assess whether particular attribute levels and their relationship to one another make a choice task difficult for a particular decision maker (Swait & Adamowicz, 2001; Payne et al., 1993). Two examples of context-based complexity are the similarity of alternatives and trade-offs (conflict) between different alternatives. In the following case study, we will use the GA approach to study how decision behavior depends on context-based complexity. The approach generates choice tasks with either high or low context-based complexity. Because of the high computational effort, existing approaches would not be applicable for such a study.

Table 2. Quality of choice tasks found by random sampling versus genetic algorithm (GA) (mean and standard deviation)

	Random sampling	GA approach
Robust objective value f'	2.59 (.63)	2.86 (.48)
Optimality of mapping f_1	3.50 (.60)	3.96 (.21)
Percentage of optimal mappings	55.8%	97.1%

Hypotheses

We assume two example choice tasks: one choice task in which the attribute level utilities of different alternatives are similar to each other, and another choice task where the attribute level utilities vary strongly. For most decision makers, the choice task becomes more difficult (has higher complexity) if the attribute level utilities of the different alternatives are similar to each other (Biggs et al., 1985; Böckenholt, Albert, Aschenbrenner, & Schmalhofer, 1991).

An explanation is that the more similar the attribute level utilities are, the harder it is for the decision maker to eliminate one of the alternatives based on only a few comparisons (Schmalhofer, Albert, Aschenbrenner, & Gertzen, 1986; Böckenholt et al., 1991). This type of context-based complexity—the similarity of alternatives—can be measured using the *attribute range AR* of a choice task c, which calculates the average standard deviation of attribute level utilities (Bettman et al., 1993; Biggs et al., 1985; Böckenholt et al., 1991). Following Bettman et al. (1993) and Payne et al. (1993), the attribute range of a choice task c is defined as

$$AR(c) = \frac{1}{m} \sum_{a_i \in A} \sqrt{\frac{1}{n-1} \sum_{p_j \in P} \left(u\left(\alpha_i\left(p_j\right)\right) - \overline{u}_i \right)^2} \qquad (4)$$

where $\overline{u}_i$ is the average utility of all attribute levels of attribute i,

$$\overline{u}_i = \frac{1}{n} \sum_{p_j \in P} u\left(\alpha_i\left(p_j\right)\right) \qquad (5)$$

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

A low *AR* means that the attribute utilities of different alternatives differ only slightly. In this case, the choice task has high *similarity* (Payne et al., 1993). Analogously, a high *AR* means the choice task has low similarity.

Procedural approaches find that high similarity leads to a consistent search process, which means that respondents spend an evenly distributed amount of time on attributes and alternatives (Payne, Bettman, & Johnson, 1988; Bettman et al., 1993) or that they consider a constant number of attribute levels per attribute (Biggs et al., 1985). Choice tasks with high similarity are more difficult because decision makers spend more time on the decision and consider a greater amount of attribute levels (Böckenholt et al., 1991; Payne et al., 1988; Stone & Schkade, 1991). Strategies where decision makers acquire information more consistently and search for more information are usually compensatory (Riedl et al., 2008).

There are only a few outcome-based approaches that study the influence of similarity on choice behavior. The results of these studies are limited because they use a compensatory model that mimics non-compensatory strategies (Swait & Adamowicz, 2001). In these studies, researchers have observed that an increasing complexity is accompanied by an increased variance of the error term and/or utility values of a standard utility model. They interpret these findings as indications that decision makes are increasingly using simplifying strategies, because the decision makers are focusing on only a few attributes (Swait & Adamowicz, 2001) and choosing the expected alternative less consistently (Dellaert, Brazell, & Louviere, 1999).[5] Bettman et al. (1993) studied the relative accuracy of the chosen alternative, which is the deviation from the choice with maximal expected value. This normative choice is equivalent to the WAD strategy. The results of this study indicated that less accurate choices are made in choice tasks with high similarity.

In sum, while process-tracing studies suggest a more compensatory decision process with increasing complexity, outcome-based approaches argue that decision makers use more simplifying strategies. None of the existing studies examined which specific decision strategies best explain people's choices in tasks of low versus high complexity. We follow the argument of Bettman et al. (1993) and postulate that "individuals may intend to attain high accuracy but may not be able to actually implement the required mental calculations" [p. 944]. Instead, respondents apply simplifying non-compensatory strategies like EBA or LEX. Therefore, we formulate the following:

HYPOTHESIS 1 (H1) A decrease of the attribute range (higher similarity) of a choice task leads to an increase in the use of EBA and LEX and a decrease in the use of WAD and MAJ.

Some studies claim that attribute ranges alone are not a valid measure for determining context-based complexity and argue that the measure of conflict also plays a major role (Iglesias-Parro, De la Fuente, & Ortega, 2002; Bettman et al., 1993). A conflict occurs if an alternative has a higher utility than another alternative on one attribute, but a lower utility on another attribute. Conflicts require a decision maker to trade off the attribute levels on different attributes, thus making the choice task more difficult (Payne et al., 1993). Following Fasolo, Hertwig, Huber, and Ludwig (2009) and Luce et al. (1997), we measured the conflict of a choice task c as the average pairwise inter-attribute correlation AC of the m utility vectors $av_i = (u(\alpha_i(p_1)), \ldots, u(\alpha_i(p_n))) \in AV$ as

$$AC(c) = \frac{m(m-1)}{2} \sum_{av_i,\, av_j \in AV, i \neq j} r(av_i, av_j) \qquad (6)$$

where r is the Pearson correlation coefficient.

If the inter-attribute correlation AC is high, not many trade-offs between different alternatives exist and the comparison of two alternatives often returns a clearly better alternative. In this case, the number of conflicts is low. If AC becomes lower, more conflicts occur (especially if the correlation becomes negative).

In consumer decision making, conflicts often occur because consumers generally want low prices but a high product quality (Fasolo, McClelland, & Lange, 2005). Therefore, decision makers often use simplifying, non-compensatory strategies "to escape from the unpleasant state of conflict" (Shephard, 1964, p. 277). However, Bettman et al. (1993) and Luce et al. (1999) observed that respondents acquire more information and need more time in choice tasks with many conflicts. They concluded that a higher number of conflicts lead to an increased use of compensatory strategies because more information is considered more consistently. Nevertheless, Bettman et al. (1993) found that relative accuracy decreases in choice tasks with many conflicts. The outcome-based approach of DeShazo and Fermo (2002) examined the influence of conflict on the consistency of choice. They found that a higher value of conflict leads to a decrease in choice consistency.

We argue that people are not capable of successfully applying a compensatory strategy in choice tasks with a high number of conflicts and formulate the following:

HYPOTHESIS 2 (H2) A decrease of the average pairwise inter-attribute correlation (higher conflict) of a choice task leads to an increase in the use of EBA and LEX and a decrease in the use of WAD and MAJ.

Generation and evaluation of experimental design

We used our optimization approach to generate robust respondent-specific choice tasks for WAD, MAJ, LEX, and EBA with an optimal mapping and low versus high complexity. We manipulated the *AR* and the *AC* of the choice tasks and formulated the objective function $f(c) = f_1/2 + f_2 + f_3$, where f_1 measures the quality of the mapping and f_2 and f_3 measure how well we can manipulate a choice task's

[5]Dellaert et al. (1999) defined choice consistency as "the variance of the random error component in the consumer utility function: the smaller this variance, the higher choice consistency" [p. 1]. Thus, the higher the consistency is, the better the utility-maximizing strategy (WAD) explains the choice.

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

attribute range *AR* and inter-attribute correlation *AC*, respectively (details on f_2 and f_3 are in APPENDIX B). Overall, the contributions of $f_1/2$ and of f_2+f_3 to the objective function are equal to one another with $f_1 \in [0,4]$ and $f_2, f_3 \in [0,1]$. Thus, finding choice tasks with optimal mapping is just as important as manipulating the *AR* and the *AC*.

We used a 2×2 factorial design ((low vs. high AR) × (low vs. high AC)). Similarly to the first evaluation, we generated choice tasks of different sizes n and m: 5 versus 8 alternatives, and 5 versus 8 attributes. We manipulated the factors within subjects in order to control for respondent-specific effects. Thus, for each of the 301 respondents, we created 16 choice tasks resulting in overall 4816 respondent-specific choice tasks. The average GA running time was 1.21 minutes for one choice task, leading to an overall computation time of 27.09 hours on four parallel processors.

To assess the performance of the GA approach, we compared the objective values of the solutions returned by the GA approach with those obtained through the random sampling benchmark (the same as in the first evaluation). These results can be found in Table 3. In this table, we show the mean and standard deviation (in brackets) of the robust objective value f', the objective value f, and its three components f_1, f_2, and f_3. We also list the percentage of choice tasks with an optimal mapping.

The average objective value f is higher for the GA approach (3.26) than for random sampling (2.87, $p < 0.01$). Also, the partial objective values f_1, f_2, and f_3 are significantly higher for choice tasks found by the GA ($p < 0.01$). Thus, the GA approach yields better solutions if we assume that the measured preferences are correct. The similar f' values for the GA approach and random sampling indicate that small modifications of attribute levels lead to solutions of similar quality for both random sampling and the GA.

Table 3. Quality of choice tasks with manipulated context-based complexity found by the genetic algorithm (GA) versus random sampling (mean and standard deviation)

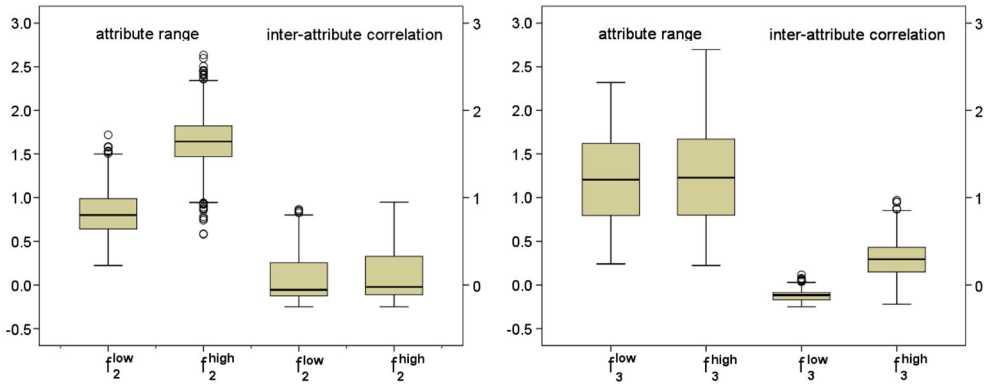
	Random sampling	GA approach
Robust objective value f'	2.36 (.45)	2.37 (.44)
Objective value f	2.87 (.45)	3.26 (.36)
Optimality of mapping f_1	1.75 (.30)	1.94 (.18)
Attribute range f_2	.55 (.17)	.62 (.19)
Inter-attribute correlation f_3	.56 (.33)	.70 (.26)
Percentage of optimal mappings	55.8%	90.4%

Thus, the high values of f obtained by the GA approach (where we assume an accurate measurement of preferences) do not lead to lower values of f' (where we assume some inaccuracy in the preference measurements); instead, the f' values of the GA approach are similar to those obtained through random sampling. Furthermore, most choice tasks returned by the GA approach have an optimal mapping (90.4%), despite the more complex objective function considering $f(c) = f_1/2 + f_2 + f_3$.

The roughly 10% non-optimal mapping can occur because of two reasons. First, for a measured preference function of a user, it is possible that no choice task with optimal mapping exists. If this is indeed the case, then it cannot be found by an optimization method. Second, the GA has to trade off between choice tasks that perform well with respect to both desired context-based complexity measures (part f_2 and f_3 in the fitness function) and tasks that perform well with respect to task diagnosticity (part f_1). Thus, the GA might decide to increase the performance of the context-complexity measures at the cost of optimal mapping.

We decided to allow the GA to create solutions with non-optimal mapping because even in choice tasks with non-optimal mapping, we were able to discriminate between strategies, as long as the observed choices did not involve multiple mappings or an ambiguous mapping. In choice task B in Figure 1, for example, only the choice of phone A, B, or C caused a discrimination problem. The choice of phones D and E did not. For applications where only choice tasks with optimal mappings are to be allowed, the design of the GA must be slightly modified. For example, all the non-optimal solutions can be dropped (by including only solutions with optimal mapping in the archive), or a higher weight can be assigned to the first part f_1 of the fitness function, leading to a lower percentage of non-optimal mappings.

We analyzed how well the GA approach manipulates the *AR* and the *AC*. For the choice tasks returned by the GA approach, Figure 4 plots the distribution of *AR* and *AC* when different variants of f_2 and f_3 are used.



(a) Using different variants of f_2 (b) Using different variants of f_3

Figure 4. Boxplots of attribute range and average pairwise inter-attribute correlation for solutions found by genetic algorithm

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

Figure 4(a) plots the results of the case where we sought to obtain choice tasks with either a high *AR* (objective function f_2^{high}) or with a low *AR* (objective function f_2^{low}). Similarly, Figure 4(b) plots the results of choice tasks with a high *AC* (objective function f_3^{high}) and a low *AC* (objective function f_3^{low}).

Both the *AR* and the *AC* can be well manipulated; moreover, the GA approach returns choice tasks with the desired properties. When searching for choice tasks with a high attribute range (f_2^{high}), we found choice tasks with a mean *AR* of 1.64. When searching for choice tasks with a low attribute range (f_2^{low}), we obtained an average *AR* of only .82. Similarly, when searching for choice tasks with a high *AC* (f_3^{high}), we obtained choice tasks with an average *AC* of .29, which is much higher than the average *AC* (−.12) we obtained when searching for choice tasks with a low *AC* (f_3^{low}).

Independently of whether the GA approach searches for choice tasks with high (f_2^{high}) or low (f_2^{low}) *AR*, the distribution of *AC* remains similar (Figure 4(a)). Likewise, the distribution of *AR* is similar regardless of the choice of f_3 (Figure 4(b)). Overall, the correlation between *AR* and *AC* is low (correlation coefficient = .018 and non-significant).

Figure 5 plots analogous results for the choice tasks obtained by random sampling. Both independent variables *AR* and *AC* cannot be well manipulated because it is difficult to find choice tasks with either high or low *AR* and *AC*. For random sampling, the difference between the mean *AR* of choice tasks with a high (f_2^{high}) versus low (f_2^{low}) attribute range is only .18, which is much smaller than for the GA approach (.82). Similarly, the difference between the mean *AC* for choice tasks with a high (f_3^{high}) versus low (f_3^{low}) inter-attribute correlation is low (.11) and much smaller in comparison with the GA (.42).

To conclude, our approach not only solves the task selection problem efficiently, but it also creates choice tasks with well-controlled treatment levels.

Results

In the main experiment, the afore-described 4816 respondent-specific choice tasks generated by the GA approach were shown to 301 respondents (16 were shown to each respondent). The attributes and alternatives of each choice task as well as the 16 choice tasks for each respondent were shown in random order. On average, respondents took 23.4 seconds per choice. Furthermore, as described in the *Generation and Evaluation of Experimental Design* section, we allowed choice tasks with non-optimal mapping but eliminated all choice tasks from the analysis where the choice could be explained by one or more strategies, or where the choice could be explained by a strategy using multiple mappings. Thus, from the 4816 observations, 4567 remained in the analysis.

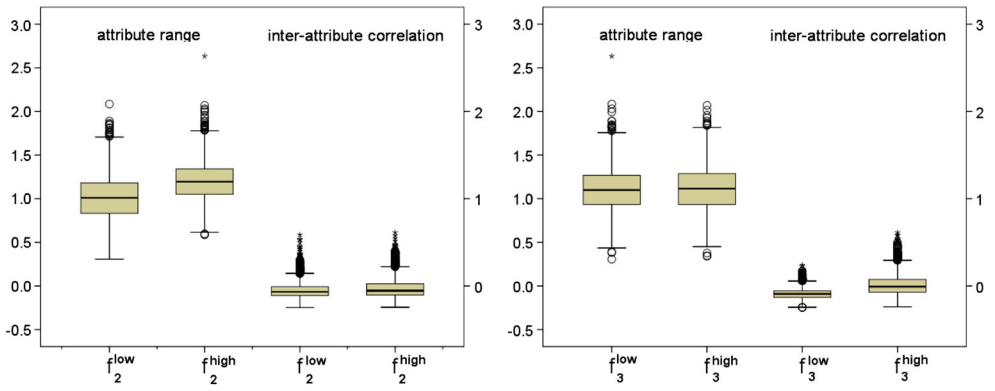
In order to test the hypotheses, we used a multinomial multilevel regression where the observed reactions to the context-based complexity were considered as nested within individuals. More specifically, we used the Generalized Linear Latent and Mixed Models (GLLAMM) analyses, where observed decision strategies are modeled as a linear combination of the complexity of the choice task and the random effect of respondents (Skrondal & Rabe-Hesketh, 2003) with the adaptive quadrature implement in GLLAMM in STATA (Rabe-Hesketh, Skrondal, & Pickles, 2005). The random effect reflects the unobserved preferences for strategy usage for each individual. The random effects were assumed to be constant for all 16 choice tasks and to follow a multivariate normal distribution. Finally, our model allowed for unobserved preferences for the distinct strategies to be correlated.

Tables 4 lists the relative frequencies of the observed strategies. Respondents applied WAD in most choice tasks (25.6%), followed by EBA, MAJ, and LEX. In 17.9% of all choice tasks, respondents selected an alternative that could not be explained by any of the four decision strategies (OTHER). We compared the observed relative frequencies with the baseline case, which assumes random choice. For example, for choice tasks with five alternatives, the baseline

Table 4. Relative frequencies of observed strategies (summary)

Strategy	Observed (%)	Baseline (%)
EBA	20.7	16.2
LEX	17.7	16.2
MAJ	18.2	16.2
WAD	25.6	16.2
OTHER	17.9	35.2

Note: EBA, Elimination by Aspect Strategy; LEX, Lexicographic Heuristic; MAJ, Simple Majority Decision Rule; WAD, Weighted Additive Difference.



(a) Using different variants of f_2 (b) Using different variants of f_3

Figure 5. Boxplots of attribute range and average pairwise inter-attribute correlation for solutions found by random sampling

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

model predicts the choices with a relative frequency of 20%. In our case, where choice tasks had either five or eight alternatives, the relative frequency of the baseline model was 16.2%. The usage of LEX and MAJ was only slightly higher than a prediction model that guesses randomly.

Figure 6 and Table 5 show the relative frequency of the observed strategies for similarity and conflict. We expected that an increase in the complexity of a choice task (higher value for similarity (lower *AR*) or conflict (lower *AC*)) would lead to an increased use of low-effort strategies (LEX and EBA) and a decreased use of high-effort strategies (WAD and MAJ). In a first step, we modeled no interaction between conflict and similarity. Both context-based complexity measures were inserted as binary variables with 0 for low similarity and low conflict (low complexity), and 1 for high similarity and high conflict (high complexity), respectively. As controls, the number of alternatives and the number of attributes (5 vs. 8) were also inserted into the model. Table 6 shows the relative frequency of the observed strategies for all four complexity measures.

The results of the model are shown in Table 7 with the exponential of the beta-coefficients to express the odds ratios, standard errors, and *p*-values. In our main analysis, we set the WAD strategy as reference because it is the most frequently used strategy. In addition, we show results for when LEX is set as reference category. The results show that the context-based complexity measures, similarity and conflict, have a stronger effect on strategy choice than the number of alternatives (*#alternatives*) or attributes (*#attributes*). A Wald test indicates that including the two context-based complexity measures and the number of alternatives creates a statistically significant improvement in the fit of the model (all $p < .01$), while this is not true for the number of attributes ($p = .21$). Furthermore, the significant influence of the number of alternatives on strategy choice is caused by a higher probability to choose OTHER compared with WAD when the number of alternatives is eight instead of five: The odds ratio for OTHER increases by 50.2% in the larger choice tasks. This is an intuitive result because in all choice tasks with optimal mapping and eight alternatives, OTHER explains four choices compared with one choice in choice tasks with five alternatives.

Hypothesis 1 states that a higher similarity makes the choice task more complex and thus that higher similarity leads to an increase in the use of EBA and LEX and a decrease in the use of WAD and MAJ. Our results indicate that this hypothesis can be supported. Compared with the reference category WAD, respondents applied EBA more often when the similarity was larger ($p < .001$ for both EBA and LEX). In fact, the odds ratio of applying EBA instead of WAD is increased by an estimated 61.71% when there was high similarity. For LEX, the odds ratio is increased by an estimated 51.4%. Furthermore, we found that similarity positively affects the choice of MAJ. Indeed, the odds ratio of applying MAJ instead of WAD was increased by an estimated 41% when similarity increased. Overall, the effects for similarity are quite strong. The higher the similarity of alternatives is, the less WAD is chosen as strategy.

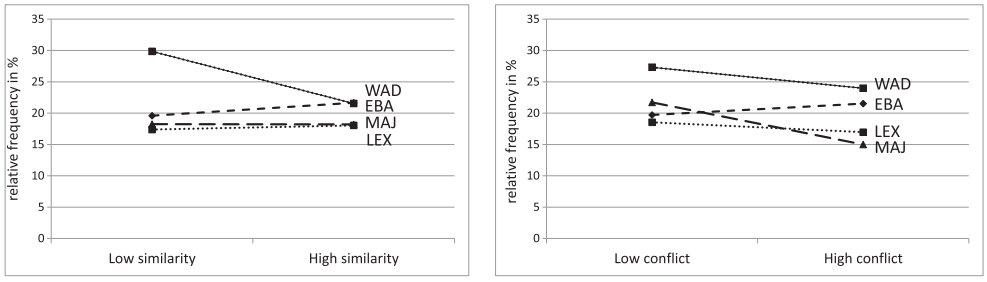
Hypothesis 2 states that an increase in conflict makes the choice task more complex and that this increase in conflict leads to an increase in the use of EBA and LEX and a decrease in the use of WAD and MAJ. The hypothesis can be supported for the usage of EBA but not for LEX. Compared with the reference category WAD, respondents applied EBA more often when the conflict was larger ($p = .005$), but this was not the case for LEX ($p = .454$). In fact, the odds ratio of applying EBA instead of WAD was increased by an estimated 29.9% for high conflict. Furthermore, we found that increasing conflict makes people apply WAD instead of MAJ. Indeed, the odds ratio of applying MAJ instead of WAD is decreased by an estimated 21.7% (100–78.3%). Thus, while similarity increased the likelihood of choosing MAJ instead of WAD, conflict decreased it.

To learn more about the effect of the complexity measures, we computed a second model, using LEX as reference category instead of WAD (Table 7). The results showed that conflict has an effect on the usage of EBA ($p = .073$) and similarity has no effect on the usage of EBA. Indeed, the odds ratio of applying EBA instead of LEX was increased by an estimated 20.6% for high conflict. Thus, when there

Table 5. Relative frequencies of the observed strategies for different combinations of conflict and similarity

Conflict	Similarity	EBA (%)	LEX (%)	MAJ (%)	WAD (%)
Low	Low	18.30	18.68	22.94	31.56
High	Low	20.71	16.25	14.14	28.37
Low	High	20.96	18.38	20.62	23.54
High	High	22.34	17.68	15.82	19.54

Note: EBA, Elimination by Aspect Strategy; LEX, Lexicographic Heuristic; MAJ, Simple Majority Decision Rule; WAD, Weighted Additive Difference.



(a) similarity (b) conflict

Figure 6. Influence of independent variables on the relative frequency of the observed strategies. EBA, Elimination by Aspect Strategy; LEX, Lexicographic Heuristic; MAJ, Simple Majority Decision Rule; WAD, Weighted Additive Difference

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

Table 6. Relative frequencies of the observed strategies

	Low sim. (%)	High sim. (%)	Low confl. (%)	High confl. (%)	5 alt (%)	8 alt (%)	5 attr (%)	8 attr (%)
EBA	19.59	21.65	19.71	21.52	22.33	19.01	20.97	20.32
LEX	17.38	18.03	18.53	16.96	19.52	15.95	17.23	18.20
MAJ	18.24	18.20	21.71	14.98	19.97	16.51	19.36	17.06
WAD	29.85	21.53	27.31	23.97	28.12	23.12	25.41	25.74
OTHER	14.95	20.59	12.74	22.57	10.08	25.41	17.01	18.69

Note: EBA, Elimination by Aspect Strategy; LEX, Lexicographic Heuristic; MAJ, Simple Majority Decision Rule; WAD, Weighted Additive Difference.

Table 7. Estimation results for the main model

	Exp(coefficient)	(Std. err.)	*p*-value
EBA when WAD is reference category			
Similarity	1.617	(.151)	.000
Conflict	1.299	(.121)	.005
#Alternatives	1.019	(.032)	.541
#Attributes	.976	(.030)	.426
Constant	.522	(.157)	.031
LEX when WAD is reference category			
Similarity	1.514	(.149)	.000
Conflict	1.077	(.106)	.454
#Alternatives	.997	(.033)	.929
#Attributes	1.009	(.033)	.786
Constant	.423	(.135)	.007
MAJ when WAD is reference category			
Similarity	1.410	(.136)	.000
Conflict	.783	(.076)	.011
#Alternatives	1.005	(.032)	.871
#Attributes	.947	(.030)	.088
Constant	.816	(.252)	.511
OTHER when WAD is reference category			
Similarity	2.119	(.207)	.000
Conflict	2.254	(.224)	.000
#Alternatives	1.502	(.051)	.000
#Attributes	1.020	(.033)	.532
Constant	.017	(.006)	.000
EBA when LEX is reference category			
Similarity	1.069	(.112)	.524
Conflict	1.206	(.126)	.073
#Alternatives	1.022	(.036)	.530
#Attributes	.967	(.034)	.335
Constant	1.238	(.425)	.534
MAJ when LEX is reference category			
Similarity	.932	(.100)	.513
Conflict	.727	(.079)	.003
#Alternatives	1.008	(.036)	.822
#Attributes	.939	(.034)	.076
Constant	1.932	(.679)	.061

Note: EBA, Elimination by Aspect Strategy; LEX, Lexicographic Heuristic; MAJ, Simple Majority Decision Rule; WAD, Weighted Additive Difference.

are more conflicts, people seem to face up to the conflicts and rather use EBA, which takes into account several attributes, than LEX, which focuses on only the most important attribute and thus cannot deal with conflicts. By comparing both compensatory strategies with LEX, we found that when the complexity in terms of similarity increases, there was no effect on MAJ, but there was a negative effect on the usage of WAD.[6] In terms of conflict, the results were reversed: When conflict increased, the odds ratio of applying MAJ instead of LEX was decreased by an estimated 27.3%, while conflict had no effect on WAD.

To sum up, an increase in conflict leads to a greater usage of EBA, while similarity only affects EBA when a compensatory strategy is assumed as the reference category. Furthermore, WAD usage decreases in cases of high similarity, whereas MAJ usage decreases in cases of high conflict.

To further understand the effects of the two context-based complexity measures, Table 5 shows the relative frequencies for the different treatment levels. In the easiest choice tasks (low similarity and low conflict), the compensatory WAD strategy explains a large portion of choices (31.56%, which is 1.95 times more than the baseline model could explain). In the most difficult choice tasks (high similarity and high conflict), the non-compensatory EBA strategy explains most choices (22.34%). Because 22.34% is only 1.34 times more

Table 8. Estimation results for the model with interaction effects

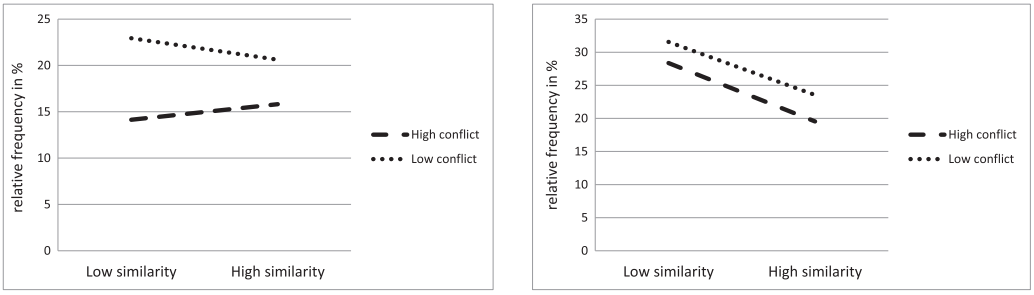
	Exp(coefficient)	(Std. err.)	*p*-value
EBA when WAD is reference category			
Similarity	1.622	(.218)	.000
Conflict	1.304	(.172)	.044
#Alternatives	1.019	(.032)	.540
#Attributes	.976	(.030)	.427
Similarity * conflict	.996	(.186)	.984
Constant	.520	(.159)	.032
LEX when WAD is reference category			
Similarity	1.385	(.192)	.019
Conflict	.983	(.136)	.899
#Alternatives	.996	(.033)	.915
#Attributes	1.010	(.033)	.772
Similarity * conflict	1.194	(.235)	.368
Constant	.443	(.143)	.012
MAJ when WAD is reference category			
Similarity	1.225	(.160)	.120
Conflict	.665	(.090)	.002
#Alternatives	1.004	(.032)	.902
#Attributes	.948	(.030)	.094
Similarity * conflict	1.379	(.266)	.096
Constant	.878	(.274)	.677
OTHER when WAD is reference category			
Similarity	2.878	(.455)	.000
Conflict	2.995	(.457)	.000
#Alternatives	1.503	(.051)	.000
#Attributes	1.019	(.033)	.554
Similarity * conflict	.635	(.129)	.025
Constant	.014	(.005)	.000

Note: EBA, Elimination by Aspect Strategy; LEX, Lexicographic Heuristic; MAJ, Simple Majority Decision Rule; WAD, Weighted Additive Difference.

[6]The significance levels for WAD are identical to the ones for LEX when WAD is the reference category.

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm



(a) MAJ (b) WAD

Figure 7. Interactions of similarity and conflict for Simple Majority Decision Rule (MAJ) and Weighted Additive Difference (WAD)

than the baseline of 16.2%, it seems that in these complex choice tasks, none of the four strategies explain a great number of choices. This indicates that, in the most complex choice tasks, respondents either applied a strategy that was not one of the four considered here or that the respondents mixed several strategies to arrive at their decision. Comparing the relative frequencies for the different strategies allows us to yet again conclude that the influence of the two context-based complexity measures is quite strong overall, and in particular for MAJ and WAD.

In order to analyze the interaction between similarity and conflict, we computed a second model that includes an interaction-term *similarity* conflict* in the model (Table 8). This term is 1 for choice task with both high similarity and high conflict. The results suggest that compared with the reference category WAD, in the most complex choice tasks, MAJ is applied more often than WAD ($p = .096$). The odds ratio of applying MAJ instead of WAD in the most complex choice tasks was increased by an estimated 37.9%.

To further clarify the findings, we plotted MAJ and WAD usage for the different levels of context-based complexity (Figure 7). The figure illustrates the following main effects (Table 7): The usage of MAJ is negatively influenced by conflict, and the usage of WAD is negatively influenced by similarity. Furthermore, for MAJ, we found an interaction (Table 8): Conflict has a higher impact on MAJ usage if similarity is low. Thus, when choice tasks get easier with respect to the similarity of alternatives, more respondents consider conflict when deciding for or against the use of MAJ.

CONCLUSIONS

First, this work contributes to the efficient design of choice task experiments and overcomes the problem of other approaches that are unable to discriminate between different decision strategies, in particular when the search space of potential designs is very large. Our approach generates choice tasks in such a way that following a particular, predefined decision strategy leads to a unique final choice. In addition, it creates robust choice tasks: Errors in measurements of attribute levels do not strongly modify the characteristics of the choice task. The design approach presented here does not rely on particular decision strategies to be tested.

Second, our approach is able to generate choice tasks for different treatment levels and thus allows researchers to manipulate a variety of choice task characteristics. In order to illustrate this, we generated choice tasks of varying complexity and observed the frequencies with which each of the four strategies EBA, LEX, MAJ, and WAD were used.

We found that context-based complexity has a strong influence on strategy choice. In choice tasks with low context-based complexity, WAD explains by far most of the observed choices (31.56% of choices are explained by WAD). In tasks with high context-based complexity, EBA explains most of the observed choices but only with 22.34%. It seems that in choice tasks with high conflict and similarity, none of the four strategies considered explain well the observed choices. In these scenarios, it seems that either other decision strategies were applied or that the decision makers used a mix of decision strategies to arrive at their decision.

By contrasting the different effects of similarity and conflict, we found that high similarity leads to a more than 50% increase in the probability of using both LEX and EBA, in comparison with WAD.

Comparing all four strategies with one another allows us to conclude that an increase in conflict leads to a greater usage of EBA, whereas an increase in similarity mainly increases the usage of LEX. Likewise, for the two compensatory strategies, WAD and MAJ, we can conclude that a decrease in conflict leads to an increase in MAJ usage, whereas a decrease in similarity mainly increases the use of WAD. An explanation of this result may be that MAJ neglects the importance of attributes and attribute level utilities and thus might not sufficiently consider conflicts. The low usage of WAD in scenarios with low similarity might be explained by the difficulty to detect the difference between the utility of alternatives if their attribute level utilities are very similar. In sum, the results are in accordance with Bettman et al. (1993), who observe that with increasing complexity, individuals make less accurate decisions and are unable to follow the more demanding compensatory strategies. Instead, they use less demanding non-compensatory strategies.

In comparison with the two context-based complexity measures, similarity and conflict, the size of the choice tasks has much less effect on which strategy is chosen. Indeed, both the number of attributes and the number of alternatives did not exhibit any significant effect on the choice of either

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

of the four strategies. The number of attributes only had a weak significant effect on MAJ with less usage of MAJ in the case of eight versus five attributes.

LIMITATIONS AND FUTURE WORK

One main limitation of the current approach is that for an unambiguous mapping, the number of alternatives needs to at least match the number of strategies. This problem can be solved by including other behavioral measurements, for example, by combining procedural methods with our approach. Combining outcome-based and procedural approaches seems to be a very promising way for studying choice behavior (Harte & Koele, 2001; Rieskamp & Hoffrage, 2008; Payne, Bettman, & Johnson, 1996; Senter & Wedell, 1999; Riedl et al., 2008). When the GA is applied in a multi-method approach, the fitness function could be modeled such that it favors ambiguous mappings of strategies whenever overlapping strategies can be differentiated by different decision processes (or other differences with respect to behavioral measurements). A second approach that would allow less alternatives than there are competing strategies would be to let the GA generate a whole portfolio of choice tasks for each experimental group. Then, one candidate solution would represent not just one choice task but rather a portfolio of choice tasks. The GA would minimize the overlap of the same strategies across different choice tasks with the disadvantage that each decision maker would need to answer several choice tasks per experimental group.

In our experiment, we tested four different strategies. WAD and MAJ are examples of compensatory strategies; EBA and LEX are examples of non-compensatory strategies. While WAD, MAJ, and EBA explained a high percentage of the choice decisions, LEX was only slightly more often used in comparison with random choice. In future work, we would like to consider more strategies and replace LEX with another strategy. Furthermore, we want to consider the possibility that decision makers do not follow pure strategies, but rather they combine standard strategies (Luce et al., 1997; Russo & Leclerc, 1994; Gilbride & Allenby, 2004; Gilbride & Allenby, 2006; Bettman & Park, 1980). Considering mixed strategies makes the creation of optimal mappings more difficult because many different strategy combinations exist. If we like to avoid to have as many alternatives as possible strategy combinations, we would recommend following the first approach mentioned earlier that allows for ambiguous mappings and incorporates procedural methods. Our approach could, for example, help to differentiate whether an EBA or LEX strategy was mixed with WAD: The GA could ensure that a user starting with EBA would always exclude other alternatives than a user starting with LEX. Process tracing could then help to determine when the strategy switch occurs, for example, by analyzing eye movements (Pfeiffer et al., 2014).

The objective of our approach was to create respondent-specific choice tasks and thus a heterogenous design of choice tasks, where different respondents were given different designs. Indeed, this approach was shown to achieve higher efficiency than when generating identical choice tasks for all respondents (Sándor & Wedel, 2005). This, however, requires the determination of some of the respondents' preferences. The prior elicitation of preferences was recently suggested by Liu and Neeraj (2011) to improve the construction of choice tasks. This approach, however, assumes that the preferences measured with adaptive conjoint analysis in the first step are stable and readily accessible. This is a rather strong assumption given the constructive nature of preferences (e.g., Payne, Bettman, & Johnson, 1992). However, our approach does not depend on a particular method for measuring the respondents' preferences in the first step. Any approach could easily be used to measure the preferences and unacceptable levels of respondents, which our algorithm needs in order to generate the respondent-specific choice tasks. It would also be possible to eliminate the respondent-specific measurement of preferences in step 1 and assume that all individuals have the same preference structure (in a similar manner to Jekel et al. (2011), where the attribute levels are coded as '+' or '−'). In fact, the influence of context-based complexity on choices has most often been measured using such symbolic attribute levels, which are assumed to be identical for all respondents (for notable exceptions, see Russo and Dosher, 1983; Böckenholt et al., 1991; Luce et al., 1997; and Pfeiffer et al., 2014). In fact, leaving out the first step requires no change to the algorithm. It just makes the approach less complicated, although probably also more inaccurate.

We partly address this estimation problem in the first step of preference measurement by anticipating inaccurate preferences: The algorithm creates choice tasks where small modifications of a choice task have only a low impact on its characteristics. Another way to address the issue of inaccurate measurements would be to directly assume some variations in the preferences. In this case, we would have to make assumptions about the type and extent of errors that occur in the preference measurements. Such an approach would increase the computational effort because each choice task generated would have to be evaluated for a variety of different preferences. Both approaches—assuming variations in the attribute level or directly in the preferences—are closely related to each other because the complexity of a choice task is calculated using preferences and attribute levels.

We hope that this work encourages other researchers to create optimal experimental designs with heuristic optimization methods, such as GAs, because this paper is a first step towards showing that heuristic optimization methods are very flexible and fast when it comes to solving complex experimental design problems.

APPENDIX A: SPECIFICS OF THE GA DESIGN

The GA evolves candidate solutions. Each candidate solution represents one respondent-specific choice task.

Construction of a candidate solution

Given a set of m_{all} possible attributes ($m_{\text{all}} \geq m$), we order the m_{all} possible attributes in a vector $A^w = (a_1, \ldots, a_{m_{\text{all}}})$ such that w

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

$(a_i) \geq w(a_{i+1})$. Similarly, for each attribute a_i, we order the l^i_{all} corresponding possible attribute levels in a vector $L^w_i = \left(l_1, \ldots, l_{l^i_{\text{all}}}\right)$ such that $u(l_i) \geq u(l_{i+1})$. A candidate solution is a vector $x = x_1, \ldots, x_{m(n+1)}$ with two different types of elements. First, the values of the elements $x_{(i-1)(n+1)+1} \in \{1, \ldots, m_{\text{all}}\}$ $(i = 1, \ldots, m)$ indicate which element of the attribute vector A^w is shown in the ith row of the choice task. To avoid multiple occurrences of the same attribute in a choice task, $x_{(i-1)(n+1)+1} \neq x_{(j-1)(n+1)+1}$ for $i \neq j$. Second, the values of the elements $x_{(i-1)(n+1)+(1+j)} \in \left\{1, \ldots, l_{l^i_{all}}\right\}$ $(i = 1, \ldots, m \text{ and } j = 1, \ldots, n)$ indicate which element of the attribute level vector L^w_i is used for the attribute a_i of product p_j.

Initialization of P

For the initial population, random solutions are generated as follows: For each solution, the vector elements that indicate the shown attributes $(x_{(i-1)(n+1)+1},\ i = 1, \ldots, m)$ are randomly drawn from the set $\{1, \ldots, m_{all}\}$ without replacement. The elements that indicate the shown attribute levels $(x_{(i-1)(n+1)+(1+j)},\ i = 1, \ldots, m$ and $j = 1, \ldots, n)$ are randomly drawn from $\left\{1, \ldots, l_{l^i_{\text{all}}}\right\}$.

Recombination and mutation

The recombination operator selects with probability $p_c = 1$ two random parent solutions from P' and copies them into P''. Then, with a probability of .5, the operator exchanges the values of all vector elements that belong to the ith attribute $(x_{(i-1)(n+1)+1}, \ldots, x_{i(n+1)})$ between the two new solutions. To ensure that no attributes are duplicated in one solution, no exchange occurs if an attribute exists in both solutions. Finally, we mutate each vector element with a probability of $1/(m(n+1))$ and copy all solutions from P'' to P'''. A mutation randomly increases or decreases the value of the vector element by 1. Thus, because of the ordering of A^w and L^w, a new solution is created with either a new attribute that has similar importance to the old attribute or a new attribute level with a similar utility as the old one. Mutation is not applied if it would duplicate the attributes in a solution.

APPENDIX B: FITNESS FUNCTIONS FOR *AR* AND *AC*

As we do not know a priori the minimum and maximum values of *AR* and *AC*, we generate, before each GA run, a set of 2000 random choice tasks. $AR_{\min}$ and $AR_{\max}$ are the minimum and maximum values of *AR* that occur in these random choice tasks. Analogously, $AC_{\min}$ and $AC_{\max}$ are the minimum and maximum values of *AC*. If we want to create a choice task c with a low attribute range, we define f_2 as $f_2^{\text{low}}(c) = \frac{AR_{\max} - AR(c)}{AR_{\max} - AR_{\min}}$; if we want to create a choice task with a high *AR*, $f_2^{\text{high}}(c) = AR(c)/(AR_{\max} - AR_{\min})$. Similarly, if we want choice tasks with a low *AC*, we define $f_3^{\text{low}} = \frac{AC_{\max} - AC(c)}{AC_{\max} - AC_{\min}}$, and if we want choice tasks with a high *AC*, we define $f_3^{\text{high}} = AC(c)/(AC_{\max} - AC_{\min})$. If choice tasks with lower minimum or higher maximum values are created during a GA run, the corresponding values of $AR_{\max}$, $AR_{\min}$, $AC_{\max}$, and $AC_{\min}$ are updated during the GA run. The normalization of *AR* and *AC* is necessary to ensure that the GA focuses equally on both *AR* and *AC*.

REFERENCES

Anderson, N. H. (1974). Algebraic models of perception. In E. C. Carterette, & M. P. Friedman (Eds.), Handbook of perception. New York: Academic Press.

Arrow, K. J. (1951). Social choice and individual values. New York: John Wiley & Sons.

Åstebro, T., & Elhedhli, S. (2006). The effectiveness of simple decision heuristics: Forecasting commercial success for early-stage ventures. *Management Science*, *52*(3), 395–409.

Atkinson, A., & Federov, V. (1992). Optimum experimental designs. Oxford University Press.

Bäck, T., Fogel, D. B., & Michalewicz, Z. (Eds.) (1997). Handbook of evolutionary computation. Bristol, UK: Institute of Physics Publishing and Oxford University Press.

Ball, C. (1997). A comparison of single-step and multiple-step transition analyses of multiattribute decision strategies. *Organizational Behavior and Human Decision Processes*, *69*(3), 195–204.

Ball, C. T., Langholtz, H. J., Auble, J., & Sopchak, B. (1998). Resource allocation strategies: A verbal protocol analysis. *Organizational Behavior and Human Decision Processes*, *76*, 70–88.

Bettman, J. R., Johnson, E. J., Luce, M. F., & Payne, J. W. (1993). Correlation, conflict, and choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 931–951.

Bettman, J. R., Johnson, E. J., & Payne, J. W. (1990, Feb). A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes*, *45*(1), 111–139.

Bettman, J. R., & Park, C. W. (1980). Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *Journal of Consumer Research*, *7*, 234–248.

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

Böckenholt, U., Albert, D., Aschenbrenner, M., & Schmalhofer, F. (1991). The effects of attractiveness, dominance, and attribute differences on information acquisition in multiattribute binary choice. *Organizational Behavior and Human Decision Processes*, *49*(2), 258–281.

Bröder, A. (2000). Assessing the empirical validity of the "take-the-best" heuristic as a model of human probability inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1332–1346.

Bröder, A., & Schiffer, S. (2003). Bayesian strategy assessment in multi-attribute decision research. *Journal of Behavioral Decision Making*, *16*(3), 193–213.

Cavagnaro, D. R., Gonzalez, R., Myung, J., & Pitt, M. A. (2013). Optimal decision stimuli for risky choice experiments: An adaptive approach. *Management Science*, *59*(2), 358–375.

Cavagnaro, D. R., Myung, J. I., Pitt, M. A., & Kujala, J. V. (2010). Adaptive design optimization: A mutual information based approach to model discrimination in cognitive science. *Neural Computation*, *22*(4), 887–905.

Cavagnaro, D. R., Pitt, M. A., Gonzalez, R., & Myung, J. (2013). Discriminating among probability weighting function using adaptive decision optimization. *Journal of Risk and Insurance and Uncertainty*, *47*, 225–289.

Chernev, A., Böckenholt, U., & Goodman, J. (2010). Choice overload: Is there anything to it. *Journal of Consumer Research*, *37*, 426–428.

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

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

Crow, L. E., Olshavsky, R. W., & Summers, J. O. (1980). Industrial buyers' choice strategies: A protocol analysis. *Journal of Marketing Research*, *17*(1), 34–44.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582.

Dellaert, B. D., Brazell, J. D., & Louviere, J. J. (1999). The effect of attribute variation on consumer choice consistency. *Marketing Letters*, *10*(2), 139–147.

DeShazo, J., & Fermo, G. (2002). Designing choice sets for stated preference methods: The effect of complexity on choice consistency. *Journal of Environmental Economics and Management*, *44*, 123–143.

Dhami, M., & Ayton, P. (2001). Bailing and jailing the fast and frugal way. *Journal of Behavioral Decision Making*, *14*, 141–168.

Ding, M., Park, Y.-H., & Bradlow, E. T. (2009). Barter markets for conjoint analysis. *Management Science*, *55*(6), 1003–1017.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, *32*, 53–88.

Fasolo, B., Carmeci, F. A., & Misuraca, R. (2009). The effect of choice complexity on perception of time spent choosing: When choice takes longer but feels shorter. *Psychology & Marketing*, *26*(3), 213–228.

Fasolo, B., Hertwig, R., Huber, M., & Ludwig, M. (2009). Size, entropy, and density: What is the difference that makes the difference between small and large real-world assortments? *Psychology & Marketing*, *26*(3), 254–279.

Fasolo, B., McClelland, G., & Lange, K. (2005). The effect of site design and interattribute correlations on interactive web-based decisions. In C. Haugtvedt, K. Machleit, & R. Yalch (Eds.), Online consumer psychology: Understanding and influencing behavior in the virtual world (p. 325–344). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Feit, E. M., Beltramo, M. A., & Feinberg, F. M. (2010). Reality check: Combining choice experiments with market data to estimate the importance of product attributes. *Management Science*, *56*(5), 785–800.

Fishburn, P. C. (1970). Utility theory of decision making. Huntington, N.Y: R. E. Krieger Pub. Co.

Fishburn, P. C. (1974). Lexicographic orders, utilities and decision rules: A survey. *Management Science*, *20*(11), 1442–1471.

Ford, J. K., Schmitt, N., Schechtman, 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*(1), 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*(9), 1197–1215.

Gigerenzer, G., & Selten, R. (2001). Bounded rationality: The adaptive toolbox. Cambridge, MA: MIT Press.

Gilbride, T. J., & Allenby, G. M. (2006). Estimating heterogeneous EBA and economic screening rule choice models. *Marketing Science*, *25*(5), 494–509.

Gilbride, T. J., & Allenby, G. M. (2004). A choice model with conjunctive, disjunctive, and compensatory screening rules. *Marketing Science*, *23*(3), 391–406.

Glöckner, A. (2009). Investigating intuitive and deliberate processes statistically: The multiple-measure maximum likelihood strategy classification method. *Judgment and Decision Making*, *4*(3), 186–199.

Glöckner, A., & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1055–1075.

Glover, F., & Kochenberger, G. A. (Eds.) (2003). Handbook of metaheuristics. Boston: Kluwer.

Goldberg, D. E. (1989). Genetic algorithms in search, optimization and machine learning. Bonn: Addison–Wesley Longman.

Goos, P., Vermeulen, B., & Vandebroeak, M. (2010). D-optimal conjoint choice designs with no-choice options for a nested logit model. *Journal of Statistical Planning and Inference*, *140*(2), 851–861.

Guo, L. (2010). Capturing consumption flexibility in assortment choice from scanner panel data. *Management Science*, *56*(10), 1815–1832.

Harte, J. M., & Koele, P. (2001). Modelling and describing human judgement processes: The multiattribute evaluation case. *Thinking and Reasoning*, *7*(7), 29–49.

Hauser, J. R., Toubia, O., Theodoros, E., Silinskiai, D., & Berfurt, R. (2010). Disjunctions of conjunctions, cognitive simplicity and consideration sets. *Journal of Marketing Research*, *47*(3), 485–496.

Hauser, J. R., & Wernerfelt, B. (1990). An evaluation cost model of consideration sets. *Journal of Consumer Research*, *16*, 393–408.

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 566–581.

Hogarth, R. M., & Karelaia, N. (2005). Simple models of multiattribute choice with many alternatives: When it does and does not pay to face trade-offs with binary attributes. *Journal of Management Science*, *51*(12), 1860–1872.

Holland, J. H. (1975). Adaptation in natural and artificial systems. Ann Arbor, MI: University of Michigan Press.

Iglesias-Parro, S., De la Fuente, E. I., & Ortega, A. (2002). The effect of context variables on cognitive effort in multiattribute binary choice. *Theory and Decision*, *52*(2), 101–125.

Inada, K. (1964). A note on the simple majority decision rule. *Econometrica*, *32*(4), 525–531.

Jasper, J., & Shapiro, J. (2002). MouseTrace: A better mousetrap for catching decision processes. *Behavior Research Methods, Instruments, & Computers*, *34*, 364–374.

Jekel, M., Fiedler, S., & Glöckner, A. (2011). Diagnostic task selection for strategy classification in judgment and decision making: Theory, validation, and implementation in R. *Judgment and Decision Making*, *6*(8), 782–299.

Johnson, E. J., & Payne, J. W. (1985). Effort and accuracy in choice. *Management Science*, *31*(4), 395–414.

Johnson, E. J., Schulte-Mecklenbeck, M., & Willemsen, M. C. (2008). Process models deserve process data: Comment on Brandsätter, Gigerenzer, and Hertwik (2006). *Psychological Review*, *115*(1), 263–273.

Kohli, R., & Jedidi, K. (2007). Representation and inference of lexicographic preference models and their variants. *Marketing Science*, *26*(3), 380–399.

Kuhfeld, W. F., Tobias, R. D., & Garratt, M. (1994). Efficient experimental design with marketing research applications. *Journal of Marketing Research*, *31*(4), 545–557.

Liu, Q., & Neeraj, A. (2011). Efficient choice design for a consider-then-choose model. *Marketing Science*, *30*(2), 321–338.

Lohse, G., & Johnson, E. (1996). A comparison of two process tracing methods on choice tasks. *Organizational Behavior and Human Decision Processes*, *68*(1), 28–43.

Luce, M. F., Bettman, J. R., & Payne, J. W. (1997). Choice processing in emotional difficult decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 384–405.

Luce, M. F., Payne, J. W., & Bettman, J. R. (1999). Emotional trade-off difficulty and choice. *Journal of Marketing Research*, *36*, 143–159.

Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory Decision*, *52*, 29–71.

Montgomery, H., & Svenson, O. (1976). On decision rules and information processing strategies for choices among multiattribute alternatives. *Scandinavian Journal of Psychology*, *17*, 283–291.

Myung, J. I., & Pitt, M. A. (2009). Optimal experimental design for model discrimination. *Psychological Review*, *116*(3), 499–518.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, *16*(2), 366–387.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology-Learning Memory and Cognition*, *14*(3), 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, *43*, 87–131.

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

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1996). When time is money: Decision behavior under opportunity-cost time pressure. *Organizational Behavior & Human Decision Processes*, *66*(2), 131–152.

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.

Pfeiffer, J., Meißner, M., Brandstätter, E., Riedl, R., Decker, R., & Rothlauf, F. (2014). On the influence of context-based complexity on information search patterns: An individual perspective. *Journal of Neuroscience, Psychology and Economics*, *7*(2), 103–124.

Pfeiffer, J., & Scholz, M. (2013). A low-effort recommendation system with high accuracy: A new approach with ranked pareto-fronts. *Business & Information Systems Engineering*, *5*(6), 397–408.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, *128*(2), 301–323.

Raidl, G. R., & Julstrom, B. A. (2000). A weighted coding in a genetic algorithm for the degree-constrained minimum spanning tree problem. In Proceedings of the 2000 ACM symposium on Applied computing. New York, NY, USA: ACM (pp. 440–445).

Reeves, C. R., & Rowe, J. E. (2003). Genetic algorithms: Principles and perspectives. Kluwer.

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*, *20*(3), 795–807.

Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. Todd, & the ABC Research Group (Eds.), Simple heuristics that make us smart. New York: Oxford University Press.

Rieskamp, J., & Hoffrage, U. (2008). Inferences under time pressure: How opportunity costs affect strategy selection. *Acta Psychologica*, *127*(2), 258–276.

Rothlauf, F. (2009). An encoding in metaheuristics for the minimum communication spanning tree problem. *INFORMS Journal on Computing*, *21*(4), 575–584.

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 (Vol. *5*(1), p. 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*(4), 676–696.

Russo, J. E., & Leclerc, F. (1994). An eye-fixation analysis of choice processes for consumer nondurables. *The Journal of Consumer Research*, *21*(2), 274–290.

Sándor, Z., & Wedel, M. (2005). Heterogeneous conjoint choice designs. *Journal of Marketing Research*, *42*(2), 210–218.

Scheibenhenne, B., Greifeneder, R., & Todd, P. M. (2010). Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research*, *37*, 409–425.

Schmalhofer, F., Albert, D., Aschenbrenner, K. M., & Gertzen, H. (1986). Process trace of binary choices: Evidence for selective and adaptive decision heuristics. *Quarterly Journal of Experimental Psychology*, *38*(A), 59–76.

Schulte-Mecklenbeck, M., Kuehberger, A., & Ranyard, R. (2011). A handbook of process tracing methods for decision research: A critical review and user's guide. New York: Psychology Press.

Schulte-Mecklenbeck, M., Murphy, R. O., & Hutzler, F. (2011). Flashlight: Recording information acquisition online. *Computers in Human Behavior*, *27*, 1771–1782.

Senter, S. M., & Wedell, D. H. (1999). Information presentation constraints and the adaptive decision maker hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 428–446.

Shephard, R. N. (1964). On subjectively optimum selection among multiattribute alternatives. In Human judgments and optimality (p. 257–281). New York: Wiley.

Shugan, S. M. (1980). The cost of thinking. *Journal on Consumer Research*, *7*, 99–111.

Skrondal, A., & Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychonometrica*, *68*(2), 267–287.

Stone, D. N., & Schkade, D. A. (1991). Numeric and linguistic information representation in multiattribute choice. *Organizational Behavior & Human Decision Processes*, *49*, 42–59.

Storer, R. H., Wu, S. D., & Vaccari, R. (1992). New search spaces for sequencing problems with application to job shop scheduling. *Management Science*, *38*(10), 1495–1509.

Swait, J., & Adamowicz, W. (2001). The influence of task complexity on consumer choice: A latent class model of decision strategy switching. *Journal of Consumer Research*, *28*(1), 135–148.

Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science*, *25*(3), 219–225.

Tsetlin, I., & Winkler, R. L. (2009). Multiattribute utility satisfying a preference for combining good with bad. *Management Science*, *55*(12), 1942–1952.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*, 31–48.

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

Wright, P. (1975). Consumer choice strategies: Simplifying vs. optimizing. *Journal of Marketing Research*, *12*(1), 60–67.

Yee, M., Dahan, E., Hauser, J. R., & Orlin, J. (2007). Greedoid-based noncompensatory inference. *Marketing Science*, *26*(4), 532–549.

Authors' biographies:

Jella Pfeiffer is a postdoctoral researcher at the Karlsruhe Institute of Technology. She received her PhD in information systems from the University of Mainz. Her research interests include decision support systems, consumer decision making, social networks, and Neuro-IS.

Dejan Duzevik is the chief technology officer of Concentric, a simulation software company in Cambridge, MA. He works on theories of human decision making and applies complexity science and agent-based modeling to social dynamics.

Franz Rothlauf is a professor in information systems at the University of Mainz. His research interests include metaheuristics and evolutionary computation, combinatorial optimization problems, and the economic value of Internet-enabled business models.

Eric Bonabeau is the chairman of Icosystem Corporation, a strategy consulting firm, and the dean of the College of Computational Sciences at the Minerva Schools at the Keck Graduate Institute. His research interests include the science of learning, complex decisions, and decision support tools.

Koichi Yamamoto is a general manager, Global Planning Office at Dentsu Inc, in Tokyo, Japan.

Authors' addresses:

Jella Pfeiffer, Institute of Information Systems and Marketing, Karlsruhe Institute of Technology, Karlsruhe, Germany.

Dejan Duzevik, Concentric, Cambridge, MA, USA.

Franz Rothlauf, Department of Information Systems and Business Administration, University of Mainz, Mainz, Germany.

Eric Bonabeau, Icosystem, Cambridge, MA, USA.

Koichi Yamamoto, Dentsu, Tokyo, Japan.

Copyright © 2014 John Wiley & Sons, Ltd.

DOI: 10.1002/bdm