Constructed Preferences and Value-focused Thinking: Implications for AI research on Preference Elicitation

Giuseppe Carenini and David Poole

University of British Columbia,
Computer Science Department ,
2366 Main Mall , V6T 1Z4, Vancouver, Canada
carenini@ubc.cs.ca

Abstract

Decision theory has changed considerably in the last decade. In behavioral decision theory, a large number of studies have shown that human decision making is inherently adaptive and constructive. In prescriptive decision theory, we have witnessed a move from an alternative-focused approach to a value-focused approach. In this paper, we discuss the implications of these new ideas in behavioral and prescriptive decision theory for AI research on preference elicitation.

Introduction

In the last three decades the field of classical decision theory has witnessed two rather independent conceptual shifts, which have become mainstream in the 90s with the publication of two corresponding milestone books [Payne, Bettman et al. 1993] and [Keeney 1992].

The first conceptual shift has occurred in the field of behavioral decision making, where a large number of studies have shown that *human choice is inherently adaptive and constructive*. Individuals, in deciding how to decide, are adaptive to both the decision task and the decision environment. They have several decision strategies at their disposal and when faced with a decision they select a strategy depending on a variety of factors related to the task, the context and individual differences. Also, additional studies investigating the contingent nature of decision making indicate that individuals often do not possess well-defined preferences on many objects and situations, but construct them in a highly context-dependent fashion during the decision process.

The second conceptual shift in decision theory has occurred in the field of prescriptive decision making and it is called *value-focused thinking*. The traditional approach to decision making that value-focused thinking criticizes is called alternative-focused thinking. In this approach the decision-maker, given a decision problem should follow three basic steps. The first step is to identify a set of plausible alternatives, the second to specify the values relevant to evaluate the alternatives, and the last to apply these values to choose the best alternative for her.

Value-focused thinking turns the decision process upside down. Once a decision problem is recognized, full specification of fundamental, relevant values is the next step. After that, the identified values are used to creatively identify possible alternatives and to carefully assess their desirability.

Although decision theory and the treatment of preferences are gaining more and more attention in AI [Doyle and Thomason 1999], it seems that AI research has somehow overlooked these two major conceptual shifts that have occurred in decision theory .

We argue that AI research on treating preferences should seriously consider the implications of the two conceptual shifts: the adaptive and constructive nature of decision making, and of value-focused thinking.

In this paper, we present few ideas on the issues involved in examining the implications of these new theories with respect to preference elicitation.

For each of the two conceptual shifts we follow the same presentation scheme. We first describe its main principles and findings in detail. Next, we start a hopefully stimulating discussion on the implications of the conceptual shift on AI research on preference elicitation.

On the Adaptive and Constructive Nature of Decision Making

Principles and Findings form Behavioral Decision Theory

Plenty of evidence from behavioral studies indicates that the achievement of four main metagoals drives human decision making [Bettman, Luce et al. 1998]. Although individuals clearly aim at *maximizing the accuracy* of their decisions, they are often willing to tradeoff accuracy to *reduce cognitive effort*. Also, because of their social and emotional nature, when making a decision people try to *minimize/maximize negative/positive emotions* and *maximize the ease of justifying* a decision.

The adaptive nature of human decision making arises from the fact that people, when faced with a decision, make three critical assessments contingent on the decision task (e.g., number of alternatives) and the decision environment (e.g., how information is presented to the DM). The first assessment involves establishing the relative importance of the four metagoals in that situation. The second assessment is to determine to what extent each of the several decision strategies that the decision-maker (DM form now on) may have at her disposal achieves each of the metagoals. Finally, the DM assesses which strategy (or which combination of strategies) best compromises among the metagoals.

For illustration, consider the prototypical decision of preferential choice, in which the DM has to select a preferred alternative (e.g., a car) out of a set of available options by evaluating the alternatives with respect to their attributes (e.g., the car's emission levels of air pollutants). Let's assume that the DM knows only two strategies to make this type of decision: (a) the weighted adding strategy (WADD), in which first each alternative is evaluated by multiplying the subjective value of each of its attributes times the attribute's importance weight, and then the alternative with the highest evaluation is selected (notice that this strategy is normative for preferential choice, i.e., the most accurate in absolute terms); (b) the elimination by aspect strategy (EBA), in which alternatives are eliminated if they do not meet a minimum cutoff threshold for the most important attribute. And the process is iteratively repeated on less important attribute until only one alternative survives

The key point of adaptive decision making is that when our DM is faced with a preferential choice she will decide which of the two strategies to use contingent on her detailed assessment of the situation. In particular, assuming that emotions and justifiability are not relevant in this context, the DM will assess a contingent tradeoff between decision accuracy and effort. Then the DM will assess to what extend WADD will be more/less accurate and will require more/less effort than EBA in this specific situation. Finally, she will select the strategy that best compromises between decision accuracy and effort.

So far we have described a top-down view of strategy selection in which the DM, after assessing the decision task and environment, selects and applies the best strategy contingent on the situation. However, several studies have shown that especially in unfamiliar and complex decision tasks DMs construct their decision strategy bottom-up by reassessing the metagoals and switching from one strategy to another as they learn more about the task structure and the environment during the course of decision making (see [Payne, Bettman et al. 1993] Chp. 5 for details).

The constructive and adaptive nature of decision making has important implications for preference elicitation. It is commonly recognized that people do not possess well-defined preferences on many objects and situations. Therefore, many expressions of preference are built when people are asked an evaluation question. But, as described above, the situational component can be a major determinant when people are asked to make a choice or express a judgment, and consequently the elicited preferences may be contingent on the elicitation setting (e.g., how the evaluation question is asked) [Fishhoff, Welch et al. 1999].

As suggested in [Payne, Bettman et al. 1999], metaphorically speaking, preference elicitation is best viewed as architecture (building a set of values) rather than archeology (uncovering existing values). In this perspective, special care should be applied in eliciting well-constructed preferences. The "building code" for preferences presented in [Payne, Bettman et al. 1999] is a first step in specifying how well-constructed preferences should be elicited. The authors start from a detailed analysis of the faults inherent in the preference construction process and for each of them they propose possible remedies. For instance, a critical fault in the preference construction process is the DM's avoidance of tradeoffs among attributes of the alternatives [Luce 1998]. Several remedies are suggested for this fault including the traditional swing weights technique that forces the DM to consider the attribute ranges, and also techniques to provide feedback to the DM on the weights implied by her judgments.

By following the building code, the outcome will be a well-constructed preference model, which is based "on thorough processing of information (reason and reflection) that is transparent and in proportion to the importance of the question at hand".

To summarize, research on human judgment and choice indicates that human decision making is adaptive and constructive. A DM, when faced with a decision, adapts to the situation by either selecting or constructing a decision strategy that she believes will be the most effective in achieving her decision metagoals with respect to the specific decision task and environment. The same constructive adaptation occurs during preference elicitation when, as it is often the case, the DM does not have pre-existing preferences on the entities involved. To cope with the detrimental effects of adaptive preference construction researches are working on a "building code" for well-constructed preferences.

The main points to keep in mind for later discussion are:

- Decision accuracy is not the only metagoal of decision making. DMs can rationally select strategies which are sub-optimal with respect to accuracy (i.e., non-normative) because to some extent they also care about the other metagoals (i.e, minimize cognitive effort, minimize negative emotions and maximize ease of justifiability).
- Since people often do not possess well-defined preferences on many objects and situations, a substantial part of their values and preferences may be constructed during preference elicitation
- A building code for preferences is an elicitation methodology that attempts to enhance the quality of the resulting preference model.

Implications of the Adaptive and Constructive Nature of Decision Making for AI research on Preference Elicitation

Let's first clarify a key distinction regarding the goal of preference elicitation [Bettman, Luce et al. 1998]. Preferences can be elicited for two very different reasons: (i) to predict what "similar" people will do in "similar" circumstances (e.g., what option they will choose), (ii) to design a preference model for a DM to help her to make an informed and balanced decision that is consistent with her values and objectives.

In the first case, when the goal is *prediction*, the elicitor does not want to eliminate the biases and/or construal effects that we have discussed in the previous section. The idea is that the preferences for prediction should be elicited in a decision environment that is as close as possible to the one in which the decision we are trying to predict will be made. In consumer research this approach is called context-matching. In implementing context-matching all relevant factors that are expected to influence the consumer (whose behavior we want to predict) are matched in the elicitation environment.

In the second case, when the goal is *design*, the elicitor should apply the "building code", briefly described in the previous section, and by doing this try as much as possible to avoid all the faults due to the adaptive and constructive nature of the elicitation process.

This basic distinction between the prediction and the design goals of preference elicitation will be relevant in our analysis of the implications of the constructive and adaptive nature of decision making on how AI techniques can be applied to facilitate the preference elicitation process.

Simplifying Elicitation: Clustering, Matching and Refining

A common complain with standard preference elicitation techniques from decision theory is that, by requiring a possibly large number of cognitively demanding questions, they are time-consuming, tedious and error-prone. The promise of some AI techniques is that they would allow one to simplify the elicitation process by reducing the number of questions and by simplifying their complexity (e.g., making them qualitative). Typically, the proposed elicitation methodology goes as follows (see [Chajewska, Getoor et al. 1998], [Ha and Haddawy 1998] for examples).

- Preference models for a sufficient number of users must be acquired using "complete and reliable" elicitation techniques.
- These models can be grouped into qualitatively different clusters – A practice consistent with considerable literature in market segmentation which indicates that people tend to form clusters according to their preferences.
- Given the clusters, the determination of a preference model for a new user is decomposed into two arguably simpler sub-processes: (i) find the cluster to which the new user more likely belongs, (ii) refine the preference model associated with that cluster for the new users

The rationale is that finding and refining a matching cluster would require significantly less elicitation steps than building a preference model from scratch.

At first sight this methodology seems quite appropriate. However, we claim that it can be problematic when the goal of the elicitation process is to design a preference model to help the DM make an accurate decision (whereas it could be quite effective for prediction).

In light of the constructive nature of preference elicitation, significantly reducing the number and complexity of the elicitation questions will cause the decision-maker to construct only a partial model of her preferences. This happens because, as preferences are generally constructed at the time the elicitation is performed, if elicitation is simplified, less sophisticated preference construction occurs. Notice that the problem is not with the model that we eventually would assign to the decision-maker by refining the model associated with her cluster. If the clustered models had been acquired applying the preference "building code", we may well expect that the refined model would actually precisely represent the preferences of the decision-maker (i.e., we would have obtained the same model by means of a thorough elicitation process). The problem is that since the elicitor did not go through full elicitation, the decision-maker did not have a chance to construct (in his mind) the "complete" model. A possible key detrimental consequence of this lack of model construction on the part of the decision-maker is that she may not understand and accept any advice based on the "complete" model, because she did not have a chance to construct the preferences on which this advise is based. For similar ideas in the apparently very different context of multi-agents decision making for public policy see [Schwarz 1999].

Also, notice that this problem does not go away if techniques more sophisticated than "clustering, matching and refining" are used (e.g., [Chajewska, Koller et al. 2000]). Again, if the elicitation process is simplified, the decision-maker preference construction process is reduced and consequently any advice based on a refined model will unlikely be understood by the decision-maker.

How to address the problem we have highlighted in this section is an open research issue for AI research. Ideally, we would still like to simplify the elicitation process for a DM by using preference models elicited for other DMs, but we would like to do this without missing elicitation questions that are critical in the preference construction process. What this might require is the ability to assess for each elicitation question a measure of its contribution to the construction process. This is clearly an issue for future research.

Another possibility might be to compensate the lack of preference construction due to reduced elicitation with an effective explanation component that by generating

detailed explanations of the proposed advice may stimulate additional preference construction.
As a final note, consider that the problem we have discussed would be particularly severe for unfamiliar and complex decision (e.g., medicine, finance) for which preferences are more likely to be constructed during the elicitation process.

Learning DM's preferences from DM's behavior
A second promising application of AI to preference elicitation is to acquire a DM's preference model by applying machine learning techniques to data about the DM's behavior. Depending on the underlying application the data can be about either the DM's previous decisions [Chajewska, Koller et al. 2001], or about the DM's holistic comparisons among pairs of alternatives [Geisler and Ha 2001], or about the DM behavior in an interface for exploring and criticizing a set of available alternatives [Shearin and Lieberman 2001]. A similar approach to acquire preferences from explicit (or declared) behavior has a long history in Economics, where stated and revealed preferences are used to build statistical models to predict consumers' behavior and consequently forecast product demand [Brownstone 2000].
Let's now consider the implications of the adaptive nature of decision making on learning preferences from behavior. First, given that the behavior of the DM is contingent on the decision task and decision environment, learning a preference model from behavior is sensible for prediction of the DM behavior in similar circumstances (as it is done in Economics), but not for preference design. The DM behavior in a particular decision task and in a particular environment will be the result of a decision strategy adaptively selected by the DM for those circumstances, and it may well be completely unrelated with the DM preferences that we would elicit by following the "building code". Notice that the only exception to this claim occurs if we assume that the behavior of the DM is guided by a normative strategy [Chajewska, Koller et al. 2001], which however is unrealistic in most situations.
So, is there any hope to learn from the DM's behavior when the goal of elicitation is design? The answer is a tentative yes. If we want to learn a preference model for design by observing the DM behavior we must first use the DM behavior to identify/track what decision strategy the DM is executing at each stage of decision making (and AI techniques may help to tackle this problem). Only in light of a specific decision strategy data about behavior can provide information on the DM's preferences. For instance, consider the example on preferential choice that we discussed in the previous section. If by observing the DM behavior we could infer that she is applying an EBA (elimination by attribute) strategy, we could derive from that same behavior information about attribute importance. This information may be useful in "designing" a linear/compensatory preference model for the DM (see [Payne, Bettman et al. 1993] Chp. 7 for similar ideas).

Value-Focused Thinking

A Glimpse at the Theory

In the last decade, the landscape of prescriptive decision theory has changed dramatically with the introduction of Value-Focused Thinking (VFT) [Keeney 1992]. This new method has influenced both the practice and the teaching of prescriptive decision theory. For instance, the second edition of the leading textbook in decision theory [Clemen 1996], has been substantially revised in light of VFT.
DMs naturally deal with complex decisions by focusing on a set of obvious (easily accessible) alternatives. Once the set of alternatives is circumscribed, the DM thinks hard to identify what are the values that would allow her to assess the desirability of each alternative and to select the most valuable one. Traditional prescriptive decision theory [Keeney and Raiffa 1976] to a large extent has assumed and accepted this alternative-focused thinking approach to decision making.
In contrast, VFT claims that the specification and clarification of values should precede as much as possible the identification of a complete set of alternatives. In this way, the DM's values, the principles used to evaluate alternatives, will not be "framed" by a pre-selected set of easily accessible alternatives, and as a result will have a more fundamental nature. In other words, VFT suggests that the objectives the DM wants to achieve in a particular decision context should creatively determine the set of alternatives that she should consider and not the other way around.
For illustration, consider the following example. Let's assume that one day you receive an email offering you an attractive new job. According to alternative-focused thinking you should spend your time trying to determine whether this new job is better than your current one (i.e., the status quo). In contrast, by following VFT, if you decide to consider a job move, you should start by identifying what you value in a job, and only once your fundamental values are clarified for this decision context, you should creatively consider what jobs you may try to get.
Although the essence of VFT can be simply stated, many aspects of decision making can benefit from this conceptual shift (see [Keeney 1992] for details):

- Preference elicitation - Because of the emphasis posed on the DM's values, a key step of VFT consists of the identification and structuring of the objectives the DM wants to achieve in a given decision context. Also, VFT provides several techniques to uncover hidden objectives (more on this later).
- Creation of new alternatives - Focusing on values first stimulates the DM to search for more desirable alternatives than the ones readily available. And whenever it is possible to creatively devise new

alternatives that better achieve her fundamental objectives.
- Communication and understanding of final decision - Fundamental objectives tend to be removed from technical language.
- Interconnecting decisions - Fundamental objectives are more likely to be general and therefore applicable to a variety of decision contexts.
- Evaluation of alternatives - A more clearly defined value model allows a more precise evaluation of the alternatives.
- Identifying decision opportunities - VFT is a proactive methodology. Once fundamental values are identified, the DM may routinely apprise to what extent they are achieved. And every time it appears that the DM could do better on any fundamental value, the DM has an opportunity to decide how to improve her situation.

Given the focus of this paper, we will only discuss the effects of VFT on preference elicitation. However, we believe that a detailed analysis of the effects of VFT on all the stages of decision making may provide useful insights to develop novel AI approaches to the treatment of preferences and in general to automate or support decision making.

Preference elicitation in VFT involves articulating and clarifying the DM's basic values for a particular decision context. The steps are similar to more conventional approaches [Keeney and Raiffa 1976], but the details are rather different.

According to VFT, the DM should qualitatively distinguish between *fundamental* and *means* objectives. Fundamental objectives should reflect what the DM really wants to accomplish with a decision, while means objectives simply help to achieve other objectives[1]. For instance, in deciding on a policy for the safety of automobile travel (see [Keeney 1992] pag. 70), minimize loss of life could be a fundamental objective, while minimizing driving under influence could be a means objective.

In a model of the DM's values, fundamental objectives are structured in a hierarchy going from general objectives to more specific ones (e.g., from minimize loss of life to minimize loss of children lifes). Means objectives, on the other hand, are organized in a network of causal relationships (e.g., minimizing driving under influence will maximize the quality of driving). In the integrated model, the hierarchy and the network are connected by linking means objectives with the fundamental objectives they help to achieve.

To assign each objective to the appropriate type is important because in VFT only the leaves in the hierarchy of fundamental objectives provide the basis on which alternative are evaluated (means-objectives play other important roles in successive steps of VFT, but because of the limited focus of this paper we will not discuss them here).

Once a preliminary hierarchy of fundamental objectives is built, VFT prescribes rather conventional steps to quantify the value model. First, we have the identification for each objective of a measurable attribute (of plausible alternatives) that can be used to assess the achievement of the objective (e.g., an attribute for the objective minimize loss of life could be total years of expected life lost). After that, the form of the utility function must be determined by verifying conditions of independence among objectives (i.e., preference, utility and additive independence). Then value tradeoffs among objectives are quantified. And finally, a utility function for each attribute is elicited.

Although VFT and traditional decision theory [Keeney and Raiffa 1976] do not differ in the basic steps applied to quantify the value model, the key aspect that distinguish VFT is its emphasis on how the iterative process of refinement and quantification of the fundamental objectives can reveal further fundamental objectives, which had remained hidden in preliminary stages of value elicitation.

Any stage of the iterative process of refinement and quantification of the value model can generate insights for uncovering hidden objectives. In this paper, we will discuss only insights from the stage in which independence conditions among objectives are verified, because this stage appeared in this preliminary analysis to be the most interesting in terms of implications for AI research on qualitative preferences.

A basic tenet of VFT is that any violation of an independence conditions among fundamental objectives is almost always an indication that either a fundamental objective is missing or that means objectives are being used in place of fundamental ones. In other words, the appropriate set of fundamental objectives will typically be additive independent. [Keeney 1992] Chp. 6 presents several examples from a variety of domains in which the violation of one of the independence conditions leads to the discovery of a hidden objective.

One of these examples is particularly relevant to understand a second important issue related to the elicitation of preferences among objectives measured at different time periods. When we are in such a situation, we will often find that attributes measuring objectives in time period t are preferentially dependent of those measuring objectives in time $t-1$. And this typically would indicate that the DM has hidden fundamental objectives concerned with the change of the objectives from time $t-1$ to time t. For instance, following the example presented in [Keeney 1992], preference for unemployment and inflation level in one year may be preferentially dependent on the measures for the same two objectives in the previous year. This finding may imply that the DM has the hidden fundamental objective of minimizing the number of jobs lost (i.e., the change in unemployment form one year to the other).

To summarize, the main points to keep in mind are:

[1] A similar distinction between means and ends objective was already present in [Keeney and Raiffa 1976], but its implications for decision making were only cursorily discussed there.

- In VFT, the quantification of the hierarchy of fundamental objectives is a powerful tool to aid the DM in qualitatively identifying and clarifying hidden objectives for a specific decision context.
- Evidence from practical experience in preference elicitation suggests that any violation of an independence condition should lead to a better understanding of the fundamental objectives of the decision problem. In general, this violation very likely implies that either a fundamental objective has been overlooked or means objectives are being used in place of fundamental ones.
- Special care is required in eliciting preference for objectives measured at different time periods, as this elicitation may involve uncovering fundamental objectives concerned with how the achievement of an objective changes over time.

Implications of Value-focused Thinking for AI research on preference elicitation

The discussion of the implications of VFT for research in AI on preference elicitation is even more preliminary than the one we outlined for the implications of the adaptive and constructive nature of decision making. As mentioned in the previous section, we limit our analysis to the implications of VFT on the stage of quantification and refinement of the value model, when independence conditions among objectives are verified.

Fundamental and means objectives: verification and representation of independence conditions

VFT indicates that a basic distinction in developing a value model for a DM is the one between objectives that should reflect what the DM really wants to accomplish with a decision (i.e., fundamental objectives) and objectives that simply help to achieve other objectives (i.e., means objectives). According to VFT, this qualitative distinction has several critical consequences on how preferences are elicited and on all other key stages of decision making.

As for preference elicitation, a basic tenet of VFT is that any violation of an independence conditions among fundamental objectives is almost always an indication that either a fundamental objective is missing or that means objectives are being used in place of fundamental ones. It is our suggestion that AI research should also distinguish between fundamental and means objectives in term of how they are acquired as well as how they are represented and reasoned about.

Several issues and questions related to AI research may be investigated. For instance, if fundamental objectives are very likely to be additive independent (once all hidden fundamental objectives have been elicited), it seems that for fundamental objectives, we should not need sophisticated graphical models to represent dependencies and independencies (like the ones proposed in [Bacchus and Grove 1995] and [Shoham 1997]). However, these models might be quite suitable to elicit and represent means objectives. In this regard, the causal interpretation of utility networks discussed in [Shoham 1997] seems to indicate that utility networks might be more suitable to express means objectives (once they have been identified), rather than fundamental objectives. Finally, what about modeling the connection between fundamental and means objectives, which appears to take care of the dependencies among objectives?

Preferences among objectives measured over time

In VFT, discovering a violation of an independence condition among fundamental objectives indicates that they should be revised. When this happens while eliciting preferences among objectives measured at different time periods, the discovery may lead to identify fundamental objectives concerned with the change of the objectives from one time period to another.

Uncovering hidden fundamental objectives concerned with the dynamic aspects of a domain may be quite relevant in eliciting preferences for agents performing decision theoretic planning. In fact, in decision-theoretic planning, to select an effective policy for an agent, we frequently need to associate utilities to consequences of a behavior over an extended period of time [Bacchus, Boutilier et al. 1996]. In Markov Decision Process's terminology, we need to specify rewards that are a function of the system trajectory or history and not of the state alone.

Since decision-theoretic planning is a major area of research in AI in which preferences play a critical role, we argue that the analysis of the structure of fundamental objectives when time induced dependencies are involved deserves considerable attention. In particular, it should be clarified whether the findings form VFT can be integrated with other approaches to describe rewards over extended periods of time [Bacchus, Boutilier et al. 1996].

Conclusions

Behavioral and prescriptive decision theory have changed considerably in the last decade. In this paper, we argue that AI research on treating preferences should seriously consider the implications of these changes.

Although our investigation is still in a very preliminary stage, we have drawn some interesting conclusions.

The adaptive and constructive nature of human decision making emerged in behavioral decision theory has at least two implications for AI research on preference elicitation. First, AI techniques that use preferences elicited from other DMs to simplify the preference elicitation effort for a DM should be applied with care. While they can be very effective when the goal is to predict the DM's behavior, their application can be problematic when the goal of preference elicitation is to design a preference model to support the DM's choice. Since in many decision contexts preferences are constructed dynamically, a simpler elicitation process may lead an AI system to suggest alternatives that the DM will unlikely understand and accept, because the DM did not have a chance to construct the relevant preference.

A second implication of the adaptive and constructive nature of decision making concerns the use of machine learning techniques to learn the DM's preferences from data about the DM's behavior. Again, this may well be extremely effective when the goal of preference elicitation is prediction, but may be problematic when the goal is design. Since the DM adaptively select decision strategies depending on the features of the decision task and environment, any sensible inference about the DM's preferences based on her behavior should be preceded by the identification of the decision strategy the DM is applying at that stage of decision making.

In prescriptive decision theory, value-focused thinking has emerged in the 90s as a very influential theory. In this paper, we have considered implications of value-focused thinking for AI research on preference elicitation. On this issue our investigation is even more tentative. We have simply highlighted a set of hopefully interesting and provoking questions related to the distinction between fundamental and means objectives, the identification and representation of independence conditions among objectives, and the elicitation of preferences among objectives measured over time.

References

[Bacchus, Boutilier et al. 1996] F. Bacchus, C. Boutilier and A. Grove. Rewarding Behaviors. Proc. of the *American Association of Artificial Intelligence Conference*, 1996.

[Bacchus and Grove 1995] F. Bacchus and A. Grove. Graphical models for preference and utility. Proc. of the *Uncertainty in Artificial Intelligence Conference*: 3--10, 1995.

[Bettman, Luce et al. 1998] J. R. Bettman, M. F. Luce and J. W. Payne. Constructive Consumer Choice Process. *Journal of Consumer research* 25: 187-217, 1998.

[Brownstone 2000] D. Brownstone. Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-fuel Vehicles. *Transportation Research* 34: 315-338, 2000.

[Chajewska, Getoor et al. 1998] U. Chajewska, L. Getoor, J. Norman and Y. Shahar. Utility Elicitation as a Classification Problem. Proc. of the *Uncertainty in Artificial Intelligence Conference*: 79-88, 1998.

[Chajewska, Koller et al. 2001] U. Chajewska, D. Koller and D. Ormoneit. Learning an Agent's Utility Function by Observing Behavior. Proc. of the *International Conference on Machine Learning*, 2001.

[Chajewska, Koller et al. 2000] U. Chajewska, D. Koller and R. Parr. Making rational decisions using adaptive utility elicitation. Proc. of the *American Association of Artificial Intelligence Conference*, 2000.

[Clemen 1996] R. T. Clemen. *Making Hard Decisions: an introduction to decision analysis*. Belmont, California, Duxbury Press, 1996.

[Doyle and Thomason 1999] J. Doyle and R. H. Thomason. Background to Qualitative Decision Theory. *AI Magazine* Summer 99: 55-68, 1999.

[Fishhoff, Welch et al. 1999] B. Fishhoff, N. Welch and S. Frederick. Construal Processes in Preference Assessment. *Journal of Risk and Uncertainty* 19(1): 139-164, 1999.

[Geisler and Ha 2001] B. Geisler and V. Ha. Modelling User Preferences via Theory Refinement. Proc. of the *Conference on Intelligent User Interfaces*, Santa Fe, New Mexico: 87-90, 2001.

[Ha and Haddawy 1998] V. Ha and P. Haddawy. Toward Case-Based Preference Elicitation: Similarity Measures on Preference Structures. Proc. of the *Proceedings UAI98*: 193-201, 1998.

[Keeney and Raiffa 1976] R. Keeney and H. Raiffa. *Decision with multiple objectives: Preferences and value tradeoffs.*, Cambridge University Press, 1976.

[Keeney 1992] R. L. Keeney. *Value-Focused Thinking: A Path to Creative Decision Making*, Harvard University Press, 1992.

[Luce 1998] M. F. Luce. Choosing to Avoid: Coping with Negatively Emotion-Laden Consumer Decisions. *Journal of Consumer Research*(24): 409-433, 1998.

[Payne, Bettman et al. 1993] J. W. Payne, J. R. Bettman and E. J. Johnson. *The Adaptive Decision Maker*, Cambridge University Press, 1993.

[Payne, Bettman et al. 1999] J. W. Payne, J. R. Bettman and D. A. Schkade. Measuring Constructed Preference: Towards a Building Code. *Journal of Risk and Uncertainty* 19(1): 243-270, 1999.

[Schwarz 1999] N. Schwarz. Defensible Preferences and the Public: Commentary on "Measuring Constructed Preferences Towards a building Code" by Payne, Bettman and Schkade. *Journal Risk and Uncertainty* 19(1): 271-272, 1999.

[Shearin and Lieberman 2001] S. Shearin and H. Lieberman. Intelligent Profiling by Example. Proc. of the *Conference on Intelligent User Interfaces*, Santa Fe, New Mexico: 145-151, 2001.

[Shoham 1997] Y. Shoham. Conditional Utility, Utility Independence, and Utility Networks. Proc. of the *Uncertainty in Artificial Intelligence Conference*: 429-436, 1997.