Multi-angle view on preference elicitation for negotiation support systems

Alina Pommeranz, Joost Broekens, Wietske Visser,

Willem-Paul Brinkman, Pascal Wiggers, Catholijn M. Jonker

Section Man-Machine Interaction

Delft University of Technology
Mekelweg 4, 2628CD, Delft, The Netherlands
a.pommeranz@tudelft.nl

ABSTRACT

Motivation – Elicitation of preferences is crucial in negotiation support. This is a non-trivial task which could be supported by computers.

Research approach – Experiment in which 32 participants have to order holidays using different preference elicitation techniques including a navigational task and affective scoring. The results were used as input for a lexicographic ordering algorithm.

Findings/design – Traditional property rating approach seems most preferred by the participants and resulted in one of the best orderings of the outcomes space to match their preferences, at least when using the lexicographic algorithm.

Originality/value – The elicitation process is approached from an algorithmic perspective as well as from a user-centred perspective for both navigation and affective attitude.

Take away message – A multi-angle approach gives a richer understanding of the process of preference elicitation.

Keywords

Preference elicitation, Recommender Systems, Lexicographic ordering, Affective scoring

INTRODUCTION AND RELATED WORK

The success of a negotiation depends on the specific preferences of the negotiating partners. In general, a negotiation is only successful if both negotiators are satisfied with the final outcome. To reach an optimal agreement, both party's preferences have to be taken into account. Hence, an important aspect of designing and implementing intelligent systems that give users decision support during a negotiation process is eliciting the users' preferences in order to build a user preference model within the system (Boutilier, 2002; Pu et al. 2003; Chen and Pu, 2004).

As pointed out by Rashid et al. (2002) and Boutilier (2002) preference elicitation is a non-trivial task. From a technical point of view, the outcome space of a negotiation is typically very large; it is not feasible to let a user specify a complete preference ordering of all outcomes directly. Therefore, a feasible and user-friendly method has to be found to elicit user preferences in such a way that an appropriate outcome ordering can be derived that resembles the user's actual preferences.

Research in this direction has mainly been done in an Artificial Intelligence context (Guo, Müller, and Weinhardt, 2003; Dastani et al., 2001) and in usability testing of specific Recommender Systems (Chen and Pu 2004; Rashid et al., 2002; Shearin and Liebermann, 2001).

Artificial Intelligence approaches typically focus on creating a user preference model, i.e. how the preferences for each user can be represented in an efficient way within the system. An interface that facilitates that model is then added. A problem that we see in this approach is that the interface design might not correspond to the way the user is capable of, and feels comfortable with, revealing his preferences. As Pu and colleagues (2003) explain, "[…], without an adequate interaction model and guidance, it is difficult for users to establish a complete and accurate model of their preferences". It is therefore important to look at preference elicitation not only from a system perspective but also from a user perspective.

Unfortunately, most user-based evaluation studies that were conducted on preference elicitation methods used for Recommender Systems focus specifically on a particular system's interface (Chen and Pu, 2004; Rashid et al., 2002; Pu et al., 2003). In addition, many of these systems are used in the same domains, such as travel assistance or shopping recommendations. There has been little research on different user-centred interaction styles for preference elicitation that could be used for any domain.

Besides the satisfaction and the cognitive limitations of the user, another aspect of preference elicitation from a user perspective is the role of emotions. Preferences can be considered affective constructs, as they are about liking versus disliking objects and properties and liking is a fundamental affective quality (see e.g. Mehrabian, 1980; Russell, 2003). In addition, emotions influence the negotiation process (Barry & Oliver, 1996; Mastenbroek, 2002; Fisher & Shapiro, 2005). Therefore, one should take the emotions that might play a role into consideration as well. However, if this influence is to be dealt with in a negotiation support system, we first and foremost need to be able to measure emotion in a reliable and valid way.

An exhaustive review of the emotion measurement literature is out of scope here, but recent studies (Isomursu et al, 2007) show that there are currently no simple-to-use, validated digital measurement tools that can easily be integrated in any application. Many approaches exist towards explicit emotion feedback in computerized systems. However, these approaches typically have a fundamental trade-off between precision and measurement speed/ease of use (Isomursu et al., 2007). Further, many methods ask a user to input categorical emotions with or without intensity (Desmet, 2002; Sanchez et al., 2006; Isomursu et al., 2007). Finally, the measurement method must be easy to integrate in any interface and the data measurements produced by the method must be valid and consistent as well as usable in the application in which the measurement tool is used.

In summary, a preference elicitation method should derive a preference ranking from the (incomplete) information that users can provide. In this process the goals, cognitive capacities and emotions of the user should be taken into account. No single research area can address all these issues and the associated risks and pitfalls. Therefore, it is important to have a multi-angle view on preference elicitation especially for negotiation support systems. To obtain effective systems and avoid sub-optimization of either the system side or the user side, one should acknowledge that they are interrelated and follow a holistic design approach.

In terms of research methods this implies that one should combine theoretical modelling, quantitative statistical measures and qualitative data (De Dreu and Carnevale, 2005; Hopmann, 2002; Moore and Murnighan, 1999). As Buelens et. al. (2008) put it, one should *triangulate* research methods to obtain a rich understanding of the problem domain. In this paper we approach preference elicitation from various angles.

RESEARCH QUESTIONS AND HYPOTHESIS

Preferences over objects, situations or outcomes of negotiations are often dependent on preferences over their properties. This dependency can be modelled in different ways. One common approach is that of multi-attribute decision theory, in which the utility of outcomes is computed from weights associated with the properties (Keeney and Raiffa, 1993). However, it is difficult to obtain such numerical values. Therefore, various qualitative approaches to multi-attribute preference definitions have been proposed, such as the lexicographic ordering. This ordering compares two objects according to the property that is rated most important. Other properties will only be considered if the value of the most important property is the same for both objects. It has been argued that this is a natural and intuitive way to derive preferences over objects or outcomes from an importance ranking of properties (see for example Liu, 2008). This suggests that the resulting order over objects reflects the user's real preferences. To our knowledge, this claim has never been confirmed by user studies. This paper presents a first attempt to fill this gap.

In order to extract preferences in the first place, different interaction styles can be used for preference elicitation. Ordering or rating properties or objects, or profiling by example (Shearin and Liebermann, 2001) are just a few examples. These styles influence how well the preferences of the user can be represented inside the system and also impact the user satisfaction. If the user is not satisfied with the interface or finds it hard to use it valuable information about her true preferences and also hidden preferences could be lost. Therefore, it is important to find the best suited method to elicit preferences. As Pu and colleagues (Pu et al., 2003) point out: "stating preferences is a process rather than a one time enumeration of preferences that do not change over time". They suggest to give the user immediate feedback of results, visual feedback and to allow the user to give any preference in any order. Based on these suggestions we investigate whether a navigational interface where people can browse through the outcome space by changing any one attribute at a time and get immediate, visually supported feedback of their choices is more preferred by the user than a simple interface based on ordering alternatives of properties. In particular we would like to compare *effort*, *intuitiveness*, *ease of use* and how much people *like* the two interaction styles. Furthermore, we would like to test whether we can extract the same information about the user's preferences from such a navigational interface as we would get from the explicit preferences by ordering the properties.

Regarding affective input of preferences, our main research question in this study is to find out if affective feedback is useful for expressing preferences. Four subquestions were investigated: (a) do users like to give affective feedback, (b) how much perceived effort is involved, (c) what is the perceived quality of the resulting ordered holiday lists, and (d) is affective feedback useful to predict the "ideal" holiday ordering of a user as generated by that user.

As argued in the introduction preference ranking methods and interaction styles for preference extraction should not be developed independently. To see how the interaction between those components influences the end result we have tested how well different ways of extracting user preferences over properties work as input for the lexicographic ordering method. We would like to see how similar the different preference orderings, discussed in the next section, generated with this method are to preference orderings as specified by users.

METHOD

In order to test different methods of extracting user preferences we ran an experiment that consisted of 8 ordering/rating tasks (tasks will be numbered throughout the paper), 2 comparisons of results and a final questionnaire. An overview of the ordering/rating tasks is presented in Table 1 (each task will be discussed in more detail below). After each task we asked participants to rate (on a 7 point scale) how much effort the task cost as well as how much they liked the task.

Task	Description
A1	Order 9 property values (given at the same time)
B1	Order 27 holidays
A2	Navigation through holidays
B2	Order 3x3 property values (given three at a time)
A3	Likert rating of holidays
B3	Affective rating of holidays
C3	Likert rating of properties
D3	Affective rating of properties

Table 1. Overview of 8 preference elicitation tasks.

We chose holidays as our domain, since most persons can relate to holidays and have preferences about different aspects of holidays. Each holiday has the properties type, location and accommodation, with respective alternative values relaxation, active and city trip, Mediterranean, Scandinavia and Alps, and hotel, camping and apartment (Table 2).

Location	Accommodation	Type
Mediterranean	Apartment	Relaxation
Alps	Hotel	City trip
Scandinavia	Camping	Active

Table 2. Properties of holidays and the alternative values for each property used in the experiments.

Material

The study material consisted of two sets of 9 cards showing one alternative value for a property of a holiday each, one set with and the other without pictures. Further, there were two sets with 27 cards showing complete holidays; one set with 4 pictures to give an orientation about what the holiday could look like, and one set without pictures. Furthermore, we used a computer interface that included 4 different tasks. In these tasks participants were asked to rate one at a time either holidays or alternatives for properties of holidays. Rating was done using either a 9-point Likert scale from *like* to *dislike* or with the *AffectButton* (Figure 1). This interface component functions, looks and behaves like a button but enables a user to input dynamic (i.e. graded) emotions. The button itself renders a face that changes directly according to the mouse position in the button as well as the scroll wheel. The mouse coordinates within the button and the scroll wheel define the values on the affective dimensions Pleasure, Dominance and Arousal (PAD) (Mehrabian, 1980) respectively. These values can be between -1 and 1. The user can therefore select an affective triplet from the PAD space by using the mouse within the button. An emotional expression that represents the PAD triplet is selected by clicking the button. We used this button to ask users about their affective preference for an item (i.e. a holiday in our experiment). We have chosen the PAD dimensions as these have proven to be fundamental and independent variables of, amongst other things, affective attitude. As we want to measure a person's affective attitude towards preferences, this is a promising model.

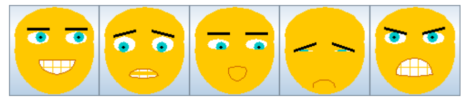

Figure 1. Example expressions: from left to right Happy (PAD=1,1,1), Afraid (-1,1,-1), Surprised (1,1,-1), Sad (PAD=-1,-1,-1), Angry (-1,1,1)

Participants

We tested 32 participants, 10 female and 22 male, which were mainly students and researchers within the field of information technology aged between 21 and 31. Each participant had to do all tasks the experiment consisted of. The order of the tasks was counterbalanced per participant.

Design

Navigation through the outcome space

To test the effect of navigating through the outcome space, i.e. complete holidays, two tasks were presented to the user. The first task is a navigation task (A2). In this task, the subject was presented with a random card with a complete holiday in the beginning and first had to find her most preferred holiday by changing the value of one property at a time to any of the 2 alternative values of that property. However, the subject could have a look at all 6 holidays related to the present one before deciding which one to navigate to. The task was presented as a paper prototype of a mobile interface. Considering the small screen on such interfaces the subject could see only the 2 alternative values for each property at a time. Once the subject found her most preferred holiday the procedure was repeated for the least preferred holiday starting with the most preferred one. The cards showed three property values of a holiday and four pictures, which were used to give the participant an idea about the kind of holiday.

In the second task (B2), the subject had to come up with a complete ordering of the alternative values of each of the 3 properties presented on cards with one value and a related picture each. Furthermore, the subject was asked to order the 3 properties type, location and accommodation according to *importance* when searching for a holiday.

A final questionnaire was presented to the user containing a number of questions about the *intuitiveness* and *ease of use* of these two tasks as well as more detailed questions about how much the subjects liked the navigation (A2) and property ordering (B2) tasks and the use of pictures.

Affective Feedback

In this part of the whole experiment we used a 2x2 experimental setup, with affect versus normal (Likert 9-points scale) rating as one independent variable and property values versus whole holidays rating as the other. As such, we had four different conditions: Likert rating of holidays (A3), affective rating of holidays (B3), Likert rating of property values (C3) and affective rating of properties (D3). For the holiday rating tasks (affect versus Likert scale rating) nine holidays were presented one by one and in random order, just as the 9 property values in the two lists of property values.

For each condition a simple algorithm generated an ordered list containing 9 holidays based on the user input. In the Likert&Holiday case the list was ordered directly based on the user's holiday preference feedback. In the Affect&Holiday case feedback variables pleasure, arousal and dominance were summed and then used to order the list. In the Likert&Property case the weight of the property value entered by the user as feedback was used to calculate a sum for each holiday that was to be ordered This sum was used to order the list of holidays. In the Affect&Property case the pleasure, arousal, and dominance feedback was summed and then used to order the property values; from this property ordering a ordering of the nine holidays was derived. These algorithms resulted in four differently sorted lists, each containing the same holidays. After the complete experiment, users were asked to score the extent to which the ordering of each of these four lists matched their own preferences (E3). They were also asked to order the four lists based on the same criterion.

Preference Ordering

Lexicographic ordering is one of the best known approaches to derive a preference ordering over objects from a given ordering over properties and property values (alternatives). As input for the lexicographic ordering method we need the order of importance of the three properties, and for each property, an ordering over that property's values. A holiday is preferred over another holiday if the former's value of the most important property is better than the latter's value of the same property. If both values are the same, the alternatives of the next most important property are considered, and so on. Consider for example Table 2, and suppose that the properties are ordered from left to right and the property values from top to bottom. Then a relaxation holiday in an apartment in the Alps is less preferred than an active hotel holiday in the Mediterranean, because the latter scores better on the most important property (location), even though it scores worse on both other properties. Likewise, a city trip in a hotel in Scandinavia is more preferred than an active camping holiday in Scandinavia, since they score the same on the first property and the former scores better on the second.

The input needed for this preference ordering method was gathered in task B2 as described above.

The input described above has a two-dimensional structure; values of each property are grouped together. It is possible to 'flatten' this structure in order to express different property value orderings. This is done by 'promoting' property values to the property level. The new property values are left implicit (they are booleans where true is better than false for every property). This approach also gives the opportunity to give two properties equal importance. In this case, the number of true properties of a given importance level are compared. Consider for example Table 3, in which the properties are again ordered from left to right. An active holiday in an apartment in Scandinavia would be preferred over a citytrip in a hotel in Scandinavia, because the former has two of the most important properties and the latter only one.

active, Scandinavia	citytrip, hotel	apartment	alps, camping	mediterranean, relaxation

Table 3. Example ordering of properties in a flattened structure.

The input needed for this was gathered in task A1, an ordering task of 9 cards showing one alternative of a holiday property each. Equally preferred alternatives could be put on the same level. All cards should be laid out on the table from most preferred to least preferred.

The same kind of input was also derived from the rating of property values on a Likert scale (C3) and by means of the AffectButton (D3), such that the property value that was rated best is considered most important.

The last task (B1) consists of ordering 27 cards showing a complete holiday, each consisting of a combination of the three properties. Equally preferred holidays could be put on the same level. All cards should be laid out on the table from most preferred to least preferred. This user-specified preference ordering is used as a standard against which the orderings generated with the lexicographic method from the different inputs will be compared.

Besides this objective comparison, we asked participants to judge which of two ordered lists of 27 holidays better reflected their preferences: the list they specified themselves in task B1 or the list generated with the lexicographic ordering method from the input from task A1.

Procedure

The study was conducted during 2 weeks. Each experiment took about 45 minutes and consisted of 8 tasks considering preference input, 2 comparisons of resulting lists and a final questionnaire. Before the tasks were explained and executed a general introduction was given about the goal of the experiment and the holiday domain. Furthermore, subjects were told that each task stands for itself, which means there is no need to remember anything between the tasks.

The presentation of tasks to users was counter balanced to avoid order of presentation effects.

RESULTS AND DISCUSSION

Preference Ordering

We have used the different methods of rating and ordering properties as input for the lexicographic ordering algorithm to investigate how well this algorithm can perform given a variety of inputs. These methods thus include affective rating (D3 in two ways, as explained in the result section on *Affective Feedback*) 9-points rating (C3), ordering 9 property values (A1), ordering the properties and then 3x3 values (B2). The algorithm generated ordered lists for each user, and these lists were compared with the lists that the users specified themselves in the 27-card ordering task (B1).

This is essentially a comparison between two rank-ordered lists containing the same items. The similarity between these lists is computed in two ways. Kendall's τ can be seen as a distance measure; it is based on the minimal number of switches between two adjacent items in one list that is needed to attain the second list. Spearman's ρ is another well-known rank correlation method. Both measures are normalized and range from -1 to 1, where 1 indicates that the lists are identical, 0 no relation at all, and -1 indicates reverts ordering.

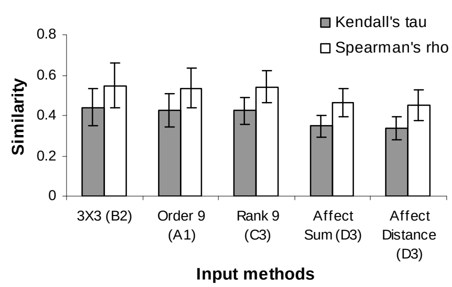

Figure 2: Rank correlations between generated holiday preference lists and user-specified holiday preference lists (including 95% confidence interval).

Figure 2 shows the correlation coefficients averaged over participants between the standard list (specified by the participant in task B1) and the lists generated with the lexicographic ordering method with different types of user input. All correlations are significant above 0 ($p < 0.001$), which indicates that the generated lists are much more similar to the standard list than random lists. It is important to note that the best any preference ordering method can do is not as high as 1, since the participants were not always consistent between tasks. Hence the human-specified 'standard' ordering is not fixed. Current results give rise to an estimate of around 0.9 as highest attainable score (see the next section, square root of 0.81).

As the τ and Spearman results are strongly correlated ($r = 0.99$) analysis on the difference between the methods focused only on τ results. An ANOVA with repeated measures with τ results as dependent variable and the methods as within-subject variables revealed a significant main effect ($F(2.77, 85.75) = 3.23$; p. = 0.027, df adjusted for sphericity violation with Greenhouse-Geisser method) for the methods. Examining Figure 2 shows that both the Affect Sum ($F(1,31) = 5.60$, p. = 0.024) and Affect Distance ($F(1,31) = 7.64$, p. = 0.010) method resulted in lower level of similarity on average, which was confirmed by a Deviation Contrast with the 3x3 method as reference category.

Navigation through the outcome space

Participants were asked to rate the tasks A2 (navigation through the outcome space) and B2 (ordering the alternatives for each property) on its intuitiveness, ease of use, effort of use, and how much they liked it. As these two tasks can provide similar information, it would be interesting to see if participants perceived them differently. Therefore a MANOVA with repeated measures was conducted which took the various ratings as dependent measures, and the task as independent within-subject variable. The multivariate analysis found a significant main effect ($F(4,28) = 3.14$; p. = 0.030) for task, which was only found again in univariate analysis on effort ($F(1,31) = 9.02$; p. = 0.005) and intuitiveness rating ($F(1,31) = 4.64$, p. = 0.039). Examining the means shows that participants rated the navigation through the outcome space (A2) task ($M = 3.0$, $SD = 1.65$) as more effortful than the B2 task ($M = 2.0$, $SD = 1.16$) and as less intuitive ($M = 4.9$, $SD = 1.48$) than task B2 ($M = 5.6$, $SD = 1.32$). Assuming that both input methods give the same information, this data suggests that the more traditional B2 method is preferred. The preference of using pictures in A2 did not seem to correlate ($r = 0.09$; p. > 0.05) with how much participants liked task A2. Some participants mentioned afterwards that pictures helped them with imagining the holidays. These participants rated the helpfulness of pictures significantly ($t(27.5) = -5.0$; p. < 0.001) higher ($M = 5.8$, $SD = 1.01$) than those who did not mention this issue ($M = 3.5$, $SD = 1.57$). Likewise participants that mentioned pictures as a distraction from their own imagination rated the usefulness of pictures significantly ($t(30) = 3.69$; p. = 0.001) lower ($M = 3.0$, $SD = 1.32$) than participants that did not make this comment ($M = 5.17$, $SD = 1.56$).

Studying task A2 in more detail however revealed that a considerable group of the participants (34%) did not consider the properties independent, which is an important assumption when using data from ordering the alternatives for each property. For example, in task A2 one participant selected relaxation, Alps, and apartment as most preferred holiday and city trip, Alps, and camping as the least preferred holiday. For this participant the Alps were both in his most and least preferred holiday. For 11 of the 32 participants the most and least preferred holiday had at least one similar value. One participant even had two similar values as his most preferred holiday was active, Alps and camping, and his least preferred holiday was city trip, Alps, and camping. This means two things. First, a property independent approach is not suitable for all people to describe their preferences. Second, the navigation through outcomes task might be an effective approach to determine whether for a specific individual preferences over properties are dependent.

Based on the values of participants' least and most preferred holidays it was possible to order the values of each property. Spearman correlations between the ordering of the value derived from the A2 task and the B2 task ranged from 0.62 to 0.91 with a 0.73 mean. Therefore it seems that A2 can obtain similar data as B2 when it comes to ordering value of a property. The B2 task also provides information about the ordering between the properties. The hypothesis was that this information could also be obtained by looking at what property participants changed first and what last in the A2 task. However, Spearman correlations between the data sets from A2 and B2 on this issue did not reach a significant level, thus this hypothesis should be rejected. Comparing the least and most preferred holiday with holidays at the top and bottom of the user-specified list of all holidays (B1) shows a match of 81% (26 out of 32) for the most preferred holiday, and a match of 44% (14 out of 36) for the least preferred holiday. This can mean several things. First, the participants were not very consistent when it came to their least preferred holiday. Second, there might be a bias in task A2 causing that participants end up with different least preferred holiday; however there was no indication to support this idea. Third, 81% seems as upper limit for an algorithm's prediction accuracy to match a person's preference list (algorithms and input methods such as those presented in Figure 2). Fourth, people are far more consistent in identifying their most preferred holiday than their least preferred holiday.

Affective Feedback

Statistical analysis of the data using a MANOVA with repeated measures showed that there is a main effect of affect versus Likert scale rating ($F(2,30) = 24.00$; $p. < 0.001$)and property versus whole holiday rating ($F(2,30) = 6.73$; $p. = 0.004$) with no significant interaction effect. These main effects were found again in the univariate analysis on effort for affect versus Likert scale rating ($F(1,31) = 46.32$; $p. < 0.001$) as well as for property versus holiday rating ($F(1,31) = 13.90$; $p. = 0.001$). This means that both affective-, as well as holiday-based feedback are associated with a higher perceived effort in preference elicitation (Table 4).

With regards to the perceived quality of the resulting nine-item lists as generated by the simple algorithms we found a significant main effect for affect versus Likert scale rating ($F(1,31) = 6.12$; $p. = 0.019$) and no main effect for holiday versus property rating or interaction effect. With regards to the ordering of the resulting generated lists based on their perceived quality we found similar results. No significant effects were found apart from an effect of affect versus Likert scale rating for property values (Wilcoxon Sign Rank Test, $z = 2.28$; $p. = 0.023$) Together these findings indicate that the algorithmically-generated sorted lists based on affective feedback matched the user's preferences less well than the lists that were generated based on normal feedback. This could be due to two reasons; (a) users did not understand the AffectButton as input device, and (b) the algorithm to generate the lists was too simplistic. The first explanation is unlikely, as pleasure and dominance strongly correlated with the Likert-scale feedback ($r=0.7$ and $r = 0.6$ respectively, $p < 0.001$). Also, previous research suggests that the AffectButton is a valid and reliable affective feedback device (Broekens, submitted). To test the second explanation we used the Lexicographic ordering method to generate lists based on affectively scored property values. Property values were ordered according to their affective Euclidian distance to "happy" (Pleasure = 1, Arousal = 1, Dominance = 1) as well as according to their affective sum (as described earlier). These property orderings were then used in the lexicographic ordering algorithm to generate 2 different lists and these lists were compared to the baseline preferences as given by the user in the 27-holiday card ordering task. On average over all 32 participants, the resulting orderings were worse than the orderings based on Likert scale rated property values as compared to the orderings given by the participants in the 27-card ordering task. This suggests that affective input might not be very useful as input for the lexicographic ordering as well, or at least that mapping affective dimensions to algorithms that are intended for one dimensional preference values is not trivial.

Finally, we conducted a regression analysis given the holiday Likert rating and pleasure, arousal, dominance ratings (backward stepwise) over all items to predict the item ranking using the user's baseline preferences given by the 27-holiday card ordering task. The same analysis was repeated for the property values (now predicting the property baseline ranking given by the 9-property card ordering task). The regression analysis with holiday-ranking as dependent variable resulted in a significant model ($F(2,285) = 110$; $p. < 0.001$) with a correlation between actual ranking and predicted ranking of $r=0.66$. The model included as significant parameters the item Likert rating ($Beta = -0.55$; $t = -9$; $p. < 0.001$) and the item pleasure rating ($Beta = -0.15$; $t = -2.5$; $p. = 0.012$). The regression analysis with property-ranking as dependent variable also resulted in a significant model ($F(2,285) = 110.6$; $p. < 0.001$) with a correlation between actual ranking and predicted ranking of r = 0.66. The model included as significant parameters the property Likert rating (Beta = -0.60; t = -11.1; p. < 0.001) and the property dominance rating (Beta = -0.11; t = -2.00; p. < 0.045). This means that, even in a simple linear model, affective feedback does add something unique in order to predict user preferences and can therefore be used to better understand human preferences.

Condition	Liking Mean and Std	Effort Mean and Std	Rated quality of generated lists	Order of generated list (bigger is better)
Likert & Holiday	3.938	2.750	6.188	2.656
	1.318	1.191	1.786	1.260
Affect & Holiday	4.188	3.906	5.500	2.438
	1.575	1.594	2.064	1.105
Likert & Property	4.188	1.938	6.031	2.781
	1.731	1.014	2.177	0.870
Affect & Property	4.313	3.250	5.156	2.125
	1.655	1.481	2.329	1.157

Table 4. Summary of average liking and effort scores for the tasks A3-D3.

Multi-angle view

After each of the eight ordering tasks participants were asked to rate the task on how much they liked it and how much effort it took them. A MANOVA with repeated measures was conducted to examine an effect for the ordering/rating style (independent within-subject variable) on the perceived effort and liking (dependent variables). The multivariate analysis found a significant main effect for ordering/rating style ($F(14,18) = 10.71$; $p. < 0.001$), which was found again in the univariate analysis of the effort rating ($F(7, 217) = 27.91$; $p. < 0.001$), and the liking rating ($F(7, 217) = 3.17$; $p. = 0.003$). As could be expected, Figure 3 shows that the ordering/rating task B1 (ordering all 27 cards) clearly stands out as least preferred and required the most effort to complete. This simply confirms the motivation behind preference elicitation research as people are not much in favour of evaluating all individual items in the outcome space. Figure 3 also shows the other side of the spectrum. The more traditional individual property ordering (B2) or rating (C3) tasks were rated low on effort and relatively high on liking. This suggests that people appreciate the relative cognitive simplicity of this task; dealing only with a small part of the outcome space complexity. From the tasks that involved evaluating the complete holidays (B1, A2, A3, and B3) it seems that the navigation through the outcome space (A2) is the most preferred one. However, the question remains why participants liked it. Was it because they liked the interaction style of browsing through the outcome space or was it because this task often involved the evaluation of relative fewer holidays compared to the other three holiday ordering/rating tasks?

Taking the results of accuracy in ordering the outcome space (Figure 2) and Figure 3 together, the first observation is that the more traditional property rating approach seems a relative good solution, as participants like it, it does not involve much effort and the ordering output was among the best matches to the participants own ordering output.

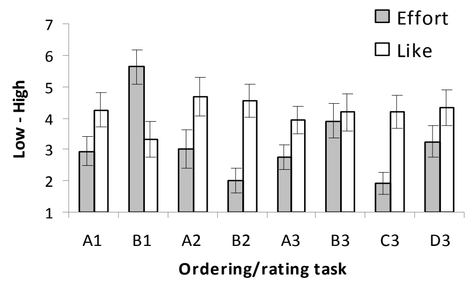

Figure 3: The mean liking and effort rating of ordering/rating tasks, including a 95% confidence interval

CONCLUSION AND FUTURE WORK

This study into preference elicitation provides a number of observations:

- Traditional property rating approach seems most preferred by the participants and resulted in one of the best orderings of the outcomes space to match their preferences, at least when using the lexicographic algorithm.
- Properties of holidays were not independent for a considerable group of the participant.
- Considering the affective attitude toward a holiday or holiday property in addition to overall attitude can improve understanding of preference elicitation.
- As illustrated by the previous point and by the navigation task that identified the dependencies in the preferences a multi-angle approach gives a richer understanding of the process of preference elicitation.

Like any study, this study also had a number of limitations. For example, for practical reason the set-up of the experiment only considers a limited number of properties and alternative values within each property to allow participants to order the whole outcome space. The real potential of some of the preference elicitation methods might therefore not have come to light in this study. Future research might look at these elicitation methods when considering a much larger outcome space with far more properties and values. The multi-angle approach also shows that even the most preferred elicitation method has serious limitations for example when it comes to dependencies of properties. A more effective approach therefore would be to combine methods to overcome pitfalls and use the strength of specific methods. For example navigation through the outcome space can be used initially to see if dependence exists, if so another method can focus on these dependencies. In addition, combination of methods that addresses both affective and overall attitude would help to increase understanding about preferences.

Another interesting observation is that applying a multi-angle approach enriches understanding of people's preference orderings. Future research might like to extend this approach by combining formal research, laboratory studies, case studies etc. including both qualitative and quantitative methodologies as was already suggested by Beulens et al (2008). This will help research to get understanding of underlying factors and how these operate in the field.

ACKNOWLEDGMENTS

We would like to thank all our participants of the experiment. This research is supported by the Dutch Technology Foundation STW, applied science division of NWO and the Technology Program of the Ministry of Economic Affairs. It is part of the Pocket Negotiator project with grant number VIVI-project 08075.

REFERENCES

Barry, B., and Oliver, R.L. (1996). Affect in Dyadic Negotiation: A Model and Propositions. *Organizational Behavior and Human Decision Processes 67* (2): 127-143.

Boutilier, C. (2002). A POMDP Formulation of Preference Elicitation Problems. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 239–246, Edmonton.

Broekens, D.J. AffectButton: Towards a Standard for Dynamic Affective User Feedback. *Submitted* (2008).

Buelens, M., Van De Woestyne, M., Mestdagh, S., Bouckenooghe, D. (2008). Methodological Issues in Negotiation Research: A State-of-the-Art-Review, *Groep Decis Negot*, 17:321--345

Chen, L., and Pu, P. (2004). Survey of Preference Elicitation Methods. Technical Report IC/200467, Swiss Federal Institute of Technology in Lausanne (EPFL).

Dastani, M., Jacobs, N., Jonker, C. M. and Treur, J. (2001). Modeling User Preferences and Mediating Agents in Electronic Commerce. In: *F. Dignum, C. Sierra (eds.), Dynamics and Management of Reasoning Processes. Series in Defeasible Reasoning and Uncertainty Management Systems*, vol.6. Kluwer Academic Publisher.

De Dreu, C., Carnevale, P. (2005). Disparate methods and common findings in the study of negotiation. *Int Negot* 10:193--203

Desmet, P. (2002). Designing Emotions. Doctoral dissertation, Delft University of Technology, The Netherlands

Druckman, D., (2005). Doing research: methods of inquiry for conflict analysis. Sage Publications, Thousand Oaks

Guo, Y., Müller,J.P. and Weinhardt, C. (2003). Learning User Preferences for Multi-attribute Negotiation: An Evolutionary Approach. *Lecture Notes in Computer Science*, Springer, Berlin.

Hopmann, P. (2002) Negotiating data: reflections on the qualitative and quantitative analysis of negotiation processes. *Int Negot* 7(1):67—85

Isomursu, M., Tahti, M., Vainamo, S., and Kuutti, K. (2007). Experimental evaluation of five methods for collecting emotions in field settings with mobile applications. *Int. Journal of Human-Computer Studies 65*(4): 404-418.

Keeney, R.L. and Raiffa, H. (1993). Decisions with Multiple Objectives: Preferences and Value Trade-Offs. Cambridge University Press.

Liu, F. (2008). Changing for the Better: Preference Dynamics and Agent Diversity. PhD Thesis, University of Amsterdam.

Mehrabian, A.. (1980). Basic Dimensions for a General Psychological Theory. OG&H Publisher.

Moore, D., Murnighan JK (1999). Alternative models of the future of negotiation research. *Negot J.,* 15(4):347—353

Pu, P., Faltings, B., and Torrens, M. (2003). User-involved Preference Elicitation. *IJCAI Workshop on Configuration*.

Rashid, A. M., Albert, I., Cosley, D., Lam, S.K., McNee, S., Konstan, J. A., and Riedl, J. (2002). Getting to know you: Learning new user preferences in recommender systems. In *Proc. of the 2002 Conference on Intelligent User Interfaces*.

Russell, J.A.. (2003). Core affect and the psychological construction of emotion. *Psychological Review 110(*1): 145-172.

Sánchez, J.A., Hernández, N.P, Penagos, J.C., and Ostróvskaya, Y. Conveying Mood and Emotion in Instant Messaging by Using a Two-Dimensional Model for Affective States. In: *Anais do IHC 2006*, ACM Press (2008), 66-72.

Shearin, S. and Liebermann, H. (2001) Intelligent Profiling by example. In Proceedings of the Conference of Intelligent User Interfaces (IUI'01), ACM Press.

Swearingen, K. and Sinha, R. (2001). Beyond Algorithms: An HCI Perspective on Recommender Systems. ACM SIGIR 2001 Workshop on Recommender Systems