Revealed or hidden?: insights into ways of measuring mental representations online: A comparative study of APT and CNET applied to an online agent

Revealed or hidden? Insights into ways of measuring mental representations online

A comparative study of APT and CNET applied to an online agent

OLIVER HORENI, THEO A. ARENTZE, BENEDICT G.C. DELLAERT[1)]
AND HARRY J.P. TIMMERMANS
Eindhoven University of Technology, [1]Erasmus University
PO Box 513
5600 MB Eindhoven
The Netherlands
O.Horeni@tue.nl

Key words: Mental Representations, Causal Networks, Interview Techniques, CNET, APT

Abstract: This paper presents and compares newly developed interview techniques (APT and CNET) which were implemented and tested by an online agent in order to measure mental representations underlying activity-choices. The comparison is supported and completed by the results from a first online survey with both new methods which was raised among respondents in the Netherlands. Their resulting mental representations for a simple activity-travel task are analysed and compared and conclusions for further investigations are drawn.

1. INTRODUCTION

When faced with a choice situation, individuals are assumed to activate some mental representation (MR) of the context-specific decision problem by selecting a subset of attributes of the faced choice alternatives, situational impacts, and individual needs being relevant to the decision at hand. Hence, a mental representation is assumed to consist of attributes, benefits, situational variables and the causal links between them. Once having established a MR the decision maker is able to interpret the choice situation and evaluate the consequences of different courses of action, matching each benefit in the MR. These matching evaluations lead to specific choices, according to decision rules and trade-offs of different benefits.

Eliciting MRs from individuals facing activity-travel choices would not only allow modellers to implement such individual variability into transport demand models but also provide insights for planners into the underlying attributes of choice alternatives and situational aspects that are decisive in activity-travel planning in any particular context.

Arentze et al. (2008) and Dellaert et al. (2008) suggested a semi-structured interview protocol for eliciting mental representations called CNET (Causal Network Eliciation Technique). CNET was successfully tested using face-to-face interviews. The main disadvantage of face-to-face interviews, however, is that these are very costly to administer and potentially sensitive to interviewer bias. This may prohibit application in large-scale surveys. The authors (Horeni et al, 2008) have therefore explored the possibilities of developing web-based online techniques. Both CNET and an existing alternative technique, association pattern technique (APT), were implemented in a web application. While APT is a highly structured method, where respondents indicate their considerations by ticking off revealed attributes and benefits, CNET works dynamically with open questions where respondents recall attributes and benefits and type them into input fields. The advantage of CNET over APT is that respondents are not influenced by the variables, chosen by the researcher.

This paper compares the outcomes of a first survey with these two techniques which collected mental representations of 70 respondents in Eindhoven, the Netherlands. The basic findings of this study indicate that the average interview duration for online APT and CNET is less than one third of conventional face-to-face interviews. Furthermore, MRs elicited by CNET are significantly smaller than MRs elicited by APT; a fact which may be attributed to the influence of an explicit listing of variables in APT. The strong differences in the elicited attributes support this assumption. While all listed attributes in APT have been indicated as being part of the mental representation by 8% to 66% of the respondents, some of them have not been elicited in CNET at all. The differences for benefits are less salient.

The next section will summarize the theoretical background of this study, including a brief presentation of the online APT and CNET applications. The third and fourth section will present the study and the results conducted in Eindhoven, The Netherlands, before the paper closes with the conclusions.

2. BACKGROUND

The vast majority of activity-based (AB) models assume that individuals share the same utilities and preference functions, and that these functions are context-independent. Admittedly, taste variation has recently been captured in terms of allowing for a distribution of estimated parameters or latent classes, but the specification of the utility functions in terms of the selected attributes is shared by all individuals. Moreover, context-dependent utility functions are rarely used. Some notable exceptions are proposed by Tversky and Simonson (1993), and Oppewal and Timmermans (1991). Apart from attribute selection, individuals and contexts may differ in terms of the benefits an individual expects from choice alternatives. Existing discrete choice analysis does not consider this layer and, hence, is limited as a means to better understand choice behaviour.

Thus, the degree of heterogeneity allowed for in mainstream models is relatively limited. Individuals may use a different set of attributes for the same choice problem, and the set of attributes used by the same individual for the same choice problem may also vary as a function of for example, constraints, involvement, available time, interest, etc. Hence, a potentially valuable line of research is to better understand the context-dependent mental representations of decision problems and to judge whether models, allowing for context-dependent mental representations perform better than current models which assume that this variation is sufficiently captured by the error terms of the model.

According to Mental Model theory (Johnson-Laird, 1983) the concept of mental representations describes how humans mentally map variables of their environment to be able to oversee the consequences of their behaviour. MRs can further be subdivided into their components: attributes, benefits, situational variables and the causal links between them. Whereas attributes relate to physically observable states of the system, benefits describe outcomes in terms of the dimensions of more fundamental needs. Situational variables describe states of the system which are beyond reach for the decision maker or they result from a far-reaching decision in the past. The links represent the causal relationships as If-Then rules between attributes, benefits and situational variables. Because individuals hold their MRs in working memory, and the capacity of that memory is limited, they will experience limitations on the amount of information that can be represented. Consequently, MRs will generally involve a significant simplification of reality. Figure 1 shows an exemplary mental representation for an activity-travel task represented as causal network.

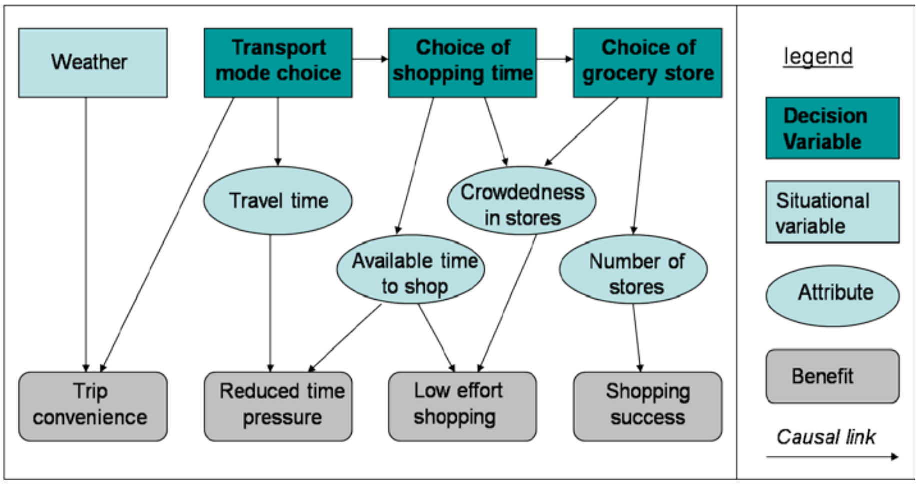


Figure 1. Mental representation for an activity-travel task.

Although MRs are established each time the decision maker faces a new choice situation, it does not necessarily mean that all underlying attributes and benefits are conscious to the decision maker. This fact complicates the measurement of MRs and makes completely unstructured techniques such as the think-aloud protocol inappropriate. There are, however, several (semi-) structured techniques which measure MRs with more or less success.

2.1 Cognitive mapping

In earlier work (Arentze et al. 2008; Dellaert et al. 2008), the authors have formulated a conceptual framework to collect data on such MRs. A semi-structured interview protocol (CNET) has been developed and tested. These face-to-face sessions were very time-consuming as they require the interaction of the interviewer and the respondent. In the worst case both interview parties need to travel to the interview location and the interviewee needs to get compensated for his travel costs. Furthermore, the interviewers

need to be coached and the data need to be digitalised after the interview. Moreover, a possible impact of the interviewer and analyst increases the risk for biases. All these potential shortcomings prevent a large-scale application of this interview protocol.
Some other approaches have been developed in the context of means-end-chain theory (Reynolds & Gutman, 1988). Basically, two groups of techniques in this domain are relevant: laddering and the association pattern technique. The former technique shows parallels to the semi-structured interview protocol suggested by the authors in that the concepts underlying a certain choice are elicited in a stepwise manner by structured questions. Russell et al. (2004a,b) compared different variants of laddering for a complex food choice problem. Their emphasis was on mothers' opinions of the role of breakfast on their children's physical and psychological well-being.

This technique was applied as soft and hard laddering. The former variant comes closest to the authors' semi-structured protocol. A major difference however is that respondents select between one and three attributes only from a list in the beginning of the face-to-face session. The underlying consequences and values are then elicited without auxiliaries by recall.

Hard laddering (see also Botschen and Thelen, 1998) in turn was performed as a computerised version and as paper-and-pencil version. For both hard laddering variants respondents had to select three important attributes, and their underlying consequences and values from revealed lists of variables. The results showed that the hard laddering techniques yielded more ladders than soft laddering; a fact which is attributed to differences in participants' cognitive processing (recall vs. recognition). While Russell et al. recommend hard laddering if the focus of the research is on investigating strong links between certain pre-determined elements, soft laddering would be more appropriate for gaining a fuller picture of participants' cognitive structure. However, the drawbacks of a face-to-face interview remain which make soft laddering not suitable for large-scale surveys.

Ter Hofstede et al. (1998) suggested another measurement technique, called the association pattern technique (APT). Similar to the hard laddering variants respondents are faced with revealed attributes, consequences and values. The difference is only that the variables are not shown in list format and that the ladders are not elicited one-by-one. Rather, APT consists of two matrices (one for attributes and consequences and one for consequences and values) where respondents can indicate causal links by ticking off the corresponding cells. Hence, all ladders are elicited simultaneously which makes this technique quite difficult. The high complexity of the matrix format with which respondents might struggle can hardly be outweighed by the short interview duration. The advantage of APT is due to its simple analysis the convenience it brings for the researcher. Thanks to the predefined labelling of attributes and benefits no post-processing of the responses is necessary, thus, making MRs conveniently comparable. Yet, the downside of this convenience is, that respondents are limited in their response freedom and possibly influenced by the revealed presentation of attributes and benefits which might rather evoke recognition than recall.

Although the presented techniques proved to work mainly under laboratory conditions for small samples, they are not very convincing for applications in large-scale surveys aiming at eliciting MRs underlying activity travel choices. Hence, we see the need for an interview technique that works automatically without an interviewer but does not influence the interviewee by showing variables. Furthermore, the structured techniques such as APT do not allow skipping layers of the MR. We believe, however, that attributes may under some circumstances not occur in some MR subsets. APT, in turn, forces respondents to indicate variables of each category. A less structured interview technique would, thus, come closer to respondents' unbiased and individually tailored MRs. The contribution of this paper is hence to test whether CNET can fulfill these requirements when applied to an online tool. In order to be able to draw comparative conclusions, also a technique with revealed variables will be brought online and tested. Although we refer to the latter technique as APT, it does not work with matrices. Rather, the variables are presented in list format separately for each category. For both techniques attention will also be directed to the respondent – machine interaction. The next section illustrates these online applications briefly by means of the experimental activity-travel task used in the survey.

2.2 Online applications of APT and CNET

2.2.1 The Online Association Pattern Technique (APT)

Having read the instructions, respondents are faced with the three interdependent decision variables involved in the experimental task (time of shopping, shopping location, and transport mode) which appear in random order on screen. After sorting them in the order in which respondents would make their decisions the revelation of the mental representations takes place separately for each decision variable in the indicated order.

Accordingly, respondents are faced with a list of eligible attributes tailored to the decision at hand which is illustrated by images of three choice alternatives (see Figure 2). Respondents are instructed to tick off the attributes being part of their MR. In case any considered attribute is missing it can be added by ticking off the “not on list”-option. Having indicated all considered attributes respondents continue with the indication of the underlying benefit(s) for each considered attribute (see Figure 3). Also the list of shown benefits is tailored to the attribute at hand. Missing benefits can be added likewise.

Figure 2 – Attribute elicitation in APT.

This procedure is repeated for the second and the third ranked decision variable until the complete MR for the underlying activity-travel task has been elicited. Finally, respondents are asked to indicate their preferred choice option for each of the three decision variables. The interview concludes with some evaluative post-experimental questions.

Figure 3 – Benefit elicitation in APT and CNET.

2.2.2 The Online Causal Network Elicitation Technique

Interpretation of Responses in online CNET

Before presenting the online CNET application some modifications of the semi-structured interview protocol need to be introduced. In contrast to face-to-face CNET where the interviewer interprets the responses in terms of attributes and benefits, this task has to be taken over by a string recognition tool. The basic auxiliary thereby are the pre-defined attributes and benefits being likely to occur in mental representations for the choice task at hand. These attributes and benefits are stored in a database from which also APT retrieves its variables. For CNET the variable set is only much larger. Besides additional attributes it comprises also synonyms and hypernyms of the defined variables to cover different individual wordings. However, the database can never cover all possible response variables. Yet, exhaustive testing ensures a sufficient comprehensiveness. Whereas new attributes or benefits could be understood by the human interviewer in face-to-face CNET, this is impossible in online CNET. The reason lies in the classification of variables into attributes and benefits according to Myers

(1976). Whether a variable is an attribute or a benefit is determined by the researcher, but it is not a semantic feature of its label. Yet, due to the open and dynamic character of the interview, the categorization is necessary as the type of the elicited variable determines the subsequent question in the interview protocol.

Nonetheless, the database of attributes and benefits alone would not suffice to understand respondents’ input. A string recognition algorithm has been developed and applied in order to find matches between the input string and the stored variables from the database. The string similarities are calculated by means of the PHP-inbuilt Soundex function and the Levenshtein distance. The respondent will then select the variable he intended from the set of possible matches. Hence, the interview protocol had to be modified. The following section will illustrate better when and how the string recognition algorithm comes into play.

The course of online CNET

The CNET interview starts with instructions and sorting of decision variables as in APT. The revelation of mental representations happens again separately for each decision variable in the indicated order with the same questions as for APT. The difference to APT is, however, that respondents are not faced with a list of revealed attributes but eight input fields instead (see Figure 4). For each typed consideration respondents are then faced with a list of variables that the string recognition algorithm detected as potential matches (Figure 5). Consequently, respondents have to select the variable that comes closest to his original consideration. In case no match is found the respondent can either retype his consideration or continue the interview with the unidentified input which will then be treated as an attribute. Hence, this step serves not only as harmonization of different labels in order to increase the inter-individual comparability of MRs but also to classify the considered issue as attribute or benefit. In case the selected label stands for a benefit the interview continues with the interpretation of further typed considerations. If, however, the selected label stands for an attribute the interview proceeds with the elicitation of the underlying benefit(s) like in APT (Figure 3). Having completed the indication of benefits the two steps illustrated in Figures 5 and 3 are repeated until all typed considerations from Figure 4 are processed. In case there are no further considerations, the whole procedure is repeated for the remaining decision variables.

Figure 4 – Eliciting MRs in CNET in open question format.

Figure 5 – Interpretation of respondents' considerations in CNET.

Another concession to the original interview protocol had to be made to allow for indicating missing links or even missing attributes. In contrast to face-to-face CNET where respondents could interact verbally if something was forgotten, online CNET needs to cope with that differently. Therefore, an additional summary step has been added to the end of the interview (Figure 6). The summary is done separately for each benefit which was elicited during the whole interview process. Respondents can see which of the elicited attributes were causally linked to the benefit at hand and which not. Furthermore, they can indicate missing links and add attributes which they forgot to state earlier. After the summary step has been performed for all benefits the respondent is asked to state his choices and evaluate the experiment by means of six post-experimental questions just like in APT.

Figure 6 – Summary step of a CNET interview.

3. EXPERIMENT

This section reports a first survey of the online tool aiming at pilot testing the interview tool and collecting first data on mental representations to allow for a comparison of both applied techniques.

3.1 Participants and design

Respondents were invited to participate in the experiment by orange paper cards in A6 format which were systematically distributed in four neighbourhoods in Eindhoven, The Netherlands. These neighbourhoods were selected such as to avoid neighbourhoods previously selected by our research group and to ensure diversification of respondents. Within these neighbourhoods all households were approached except the ones which explicitly excluded impersonal postings to their letterboxes. Besides the invitation text and the link to the interview the invitation cards included the logo of the TU Eindhoven, the research subject, the name of the researcher and his email address. As incentive for participation a lottery was announced where 10 respondents would win shopping vouchers each worth €50. Furthermore, a date was mentioned by which the interview could be performed. Depending on the neighbourhood in which the addressed household was located this deadline amounted between one and three weeks.

From a total of 3945 households which were addressed 276 started the interview (≈7%). Yet, only 137 respondents (49.64%) finished the interview successfully which yields a net response rate of 3.47%. This paper reports however only on the outcomes of the 70 respondents who were randomly assigned to a basic experimental scenario in CNET and APT which is described in the next section. Respondents who were faced with modifications of the basic experimental scenario are not regarded in this paper.

Table 1 presents sample descriptors calculated from responses to questions concerning socio-demographic information. It shows that there are only little differences between the sub-samples. Remarkable however is the high number of participants with a university degree (73.7% vs. 78.1%). This outcome cannot only be attributed to the fact that the survey took place in neighbourhoods close to the university. Rather, it indicates a greater appeal of scientific online surveys to higher educated people and a stronger interest in participation among this group. The licence ownerships of 100% are caused by the fact that respondents without driving licence were assigned to another scenario for experimental reasons.

Table 1 – Sample Descriptors

Characteristics	**APT**	**CNET**
N	38	32
Gender (% men)	60.5	59.4
Age (years) (M/SD)	47.5/17.6	48.1/17.2

Status	*Single (%)*	34.2	18.8
	Childless Couple (%)	36.8	37.5
	Couple with child (%)	23.7	37.5
	Lone parent (%)	5.3	0
	Other (%)	0	6.3
Education	*Secondary school (%)*	15.8	6.3
	MBO (%)	7.9	15.6
	University (%)	73.7	78.1
Driving licence (%)		100	100
Vehicle ownership (%)			
	Bicycle	92.1	96.9
	Scooter	2.6	0
	Motorcycle	2.6	6.3
	Car	78.9	84.4
Possession of PT passes (%)			
	40% discount card	31.6	43.8
	Annual ticket	7.9	6.3
	Route bound discount	0	3.1

3.2 Experimental design

In the interview respondents were exposed to a complex activity-travel task consisting of three inter-depending decisions for transport mode (car vs. bus vs. bicycle), the location for daily grocery shopping (central market vs. corner store vs. supermarket) and time of shopping (during lunch break vs. after work vs. later in the evening) for a usual workday in a fictive environment. They were instructed about the environmental conditions and the alternatives for each choice that had to be taken (see Figures 7). A map of the fictive city and small images for the choice alternatives served as mental support. These maps and the provided information differed slightly between the experimental scenarios to which respondents were assigned randomly. However, this paper deals only with the basic scenario. The interview technique (APT vs. CNET) was assigned randomly, too.


Prologue - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
http://www.cnet.ddss.nl/survey/APT/prolog_1.php
Neue Anfrage Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten
Prologue
Beschrijving van de taak
thuis
werk
Stelt u zich voor dat u een nieuwe baan heeft gekregen in een andere stad en dat u daarom naar die stad moest verhuizen. Op de plattegrond hiernaast ziet u uw nieuwe huis en uw werkplek. U werkt acht uur met een flexibele begintijd en kunt maximaal een uur lunchpauze nemen. Voor het doen van dagelijkse boodschappen heeft u daarom drie mogelijkheden:
• tijdens de lunchpauze
• meteen na het werk
• later 's avonds
Bedenk wat uw overwegingen zijn bij de keuze van het tijdstip van boodschappen doen voor dagelijkse behoeften.
Terug
Doorgaan
Fertig


Prologue - Mozilla Firefox
Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe
http://www.cnet.ddss.nl/survey/APT/prolog_2.php
Neue Anfrage Meistbesuchte Seiten Erste Schritte Aktuelle Nachrichten
Prologue
Beschrijving van de taak
grote supermarkt
kleine supermarkt
markt
In uw stad zijn drie winkellocaties te vinden voor dagelijkse boodschappen:
• een kleine winkel dicht bij uw huis
• een grote supermarkt aan de stadsrand
• een weekmarkt in het centrum
De producten op uw boodschappenlijstje zijn op elke locatie de hele dag verkrijgbaar.
Bedenk wat uw overwegingen zijn bij de keuze van de locatie van boodschappen doen voor dagelijkse behoeften.
Terug
Doorgaan
Fertig

Figures 7 - Screenshots of the experimental description in APT and CNET.

4. RESULTS

Firstly, respondents´ ranking of decision variables is presented. Thereafter, the structure of MRs in terms of number of attributes, benefits and causal links is compared and their content in terms of the nature of frequently considered variables is analysed. The final subsection deals with general issues encountered with the online surveys.

4.1 Sorting of decision variables

The ranking of the decision variables did not yield clear differences (see Table 2). The average rank scores for all decision variables are around 2, suggesting that the ranking is quite balanced. Whereas APT respondents preferred to plan the time of shopping before transport mode and shopping location, CNET respondents showed the reversed order. Given the fact that APT and CNET do not differ in this interview step, i.e. the technique cannot have an influence on the order of decisions, the averaged values of both techniques are presented in the column ‘mean’. The order there is the same as for CNET respondents only, but even closer to 2. Yet, it has to be noted that actually only a few respondents ranked the transport mode choice second. Rather, respondents who ranked it as the first and the third decision were almost balanced.

Table 2 – Ranking of the decisions (average rank scores)

Variable	APT	CNET	Mean
Transport Mode	2.05 (SD 0.928)	1.91 (SD 0.777)	1.99 (SD 0.860)
Shopping Location	2.05 (SD 0.695)	1.88 (SD 0.751)	1.97 (SD 0.722)
Shopping Time	1.89 (SD 0.831)	2.22 (SD 0.906)	2.04 (SD 0.875)

4.2 Parameters of the MRs in APT and CNET

The elicited MRs were analysed in terms of the following variables: number of associations, total number of attributes, number of added attributes, number of benefits, number of added benefits, and number of benefits per attribute. Table 3 reports means for each dependent variable and each experimental group. Furthermore, the average interview duration is shown in Table 3.

Table 3 – Means of the dependent variables for each experimental group

Variable	APT	CNET	t	df	p
Interview duration	13min 33s	19min 05s	-2.605	68	.011
No. of associations[1)]	41.66	22.13	2.877	52.6	.006
Total no. of attributes[2)]	11.71	6.31	5.773	68	<.001
No. of added attributes[3)]	0.21	2.16	-5.541	36.58	<.001
No. of benefits[4)]	12.47	9.00	3.388	68	.001
No. of added benefits[5)]	0.82	0.31	1.734	56.0	.088
No. of benefits/attribute[6)]	1.09	1.56	-3.407	38.2	.002

1) Number of associations counts link chains of the form: Decision Variable – (Attribute -) Benefit
2) Total number of attributes counts attributes which were ticked off (APT), typed in (CNET) or added (APT)
3) Number of added attributes counts added attributes (APT) and not-interpretable inputs (CNET)
4) Number of benefits counts benefits which were recalled (CNET), ticked off or added (APT and CNET)
5) Number of added benefits counts all benefits which have been added to the list (APT and CNET)
6) Number of benefits per attribute is the ratio between number of benefits and number of attributes

An examination of Table 3 reveals that APT yields significantly different means than CNET for almost all dependent variables. Compared with the average interview duration for CNET (19min 05s), APT respondents, on average, spend 5min 32s less to complete their task. Apart from the fact that the longer interview duration of CNET is caused by additional and repetitive interview steps (see Figures 5 and 6) and the probably longer pauses for thought, it is striking how much faster respondents finished the online CNET interview compared to face-to-face interviews. Dellaert et al. (2008) report an average interview duration of 55 minutes, but their interview included an additional set of questions to reveal parameters of the causal network (i.e., conditional probabilities and utilities).

The number of associations is almost twice as high for APT than it is for CNET which might be caused by an induction effect of presenting variable lists to the respondent. It is conceivable that APT respondents indicated causal links between variables which they recognized as plausible reasons but which were not necessarily part of their MR. The t-test showed that APT differs significantly from CNET (p = .006) in this respect.

The total number of attributes is roughly twice as high for APT and, therefore, significantly different (p <.001) from CNET. This finding supports the hypothesis that CNET is a more sensitive methodology for measuring MRs as it prevents induction of variables that might be part of the broader causal knowledge of the respondent but are not brought to bear for making the decisions. The number of added attributes also shows a significant difference between APT and CNET (p < .001). On the one hand, the low value for APT (0.21) speaks to the completeness of the list of attributes in APT. On the other hand, this recognition-oriented methodology might hamper respondents in rendering their MR completely consciously, i.e. attributes which are not on the list are not recalled. The higher values for CNET in turn do not necessarily speak to the incompleteness of the database. Rather, it might be caused by the imperfect performance of the string recognition algorithm. Whenever wordings were used for which no match was found, the respondent could go on with the not interpreted input which was then treated as an added attribute. It does, however, not necessarily mean that this attribute is not stored in the database under a different label.

The difference in number of benefits is significant when comparing APT with CNET (p = .001). The higher values for the number of benefits among APT respondents might be a multiplication effect as also the total number of attributes was higher in APT. Hence, APT respondents were more frequently faced with the interview step aiming at eliciting benefits (Figure 4). The difference is not related to the technique per se since respondents of both techniques were able to recognize benefits. Thus, the number of added benefits does not differ significantly between the techniques.

When comparing the ratio of benefits and attributes, APT yields significantly ($p = .002$) lower numbers than CNET. While this ratio is almost 1:1 for APT, CNET yields around 1.5 times more benefits than attributes. The reason for this perhaps unexpected finding is, as mentioned above, the comparatively low number of recalled attributes to the high number of recalled (mean 0.72) and recognized benefits in CNET.

4.3 Frequently considered variables

The elicited attributes and benefits were analysed separately for each decision variable in order to examine which variables were frequently considered by the respondents for different decisions.

Transport mode (TM) decision

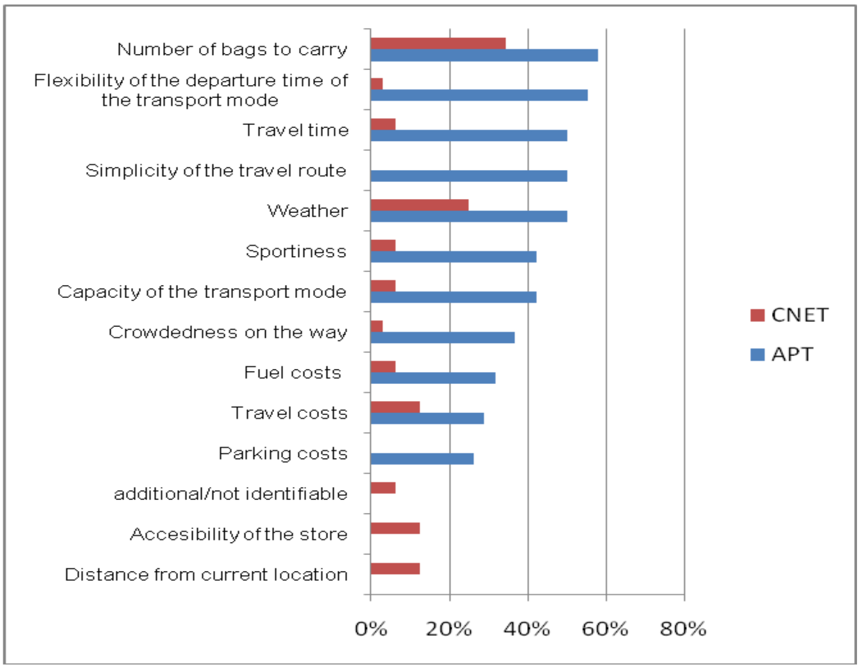


Figure 8. Elicited attributes for TM decision for CNET and APT.

Figure 8 shows all attributes that were elicited by either technique for considering transport mode in the experimental activity-travel task by at least 10% of the respondents. It is remarkable that all attributes being shown to the respondents in APT were also indicated as being considered by 26 to 58% of the respondents. In contrast, none of the APT respondents indicated to consider an attribute which was not on the list. While one could ascribe this result to the completeness of the list of attributes we rather interpret the outcome as a bias APT entails.

When having a look at the CNET results it is apparent that the attributes were elicited less frequently than in APT. Merely five attributes were elicited by more than 10% of the respondents (number of bags to carry, weather, travel costs, accessibility of stores and distance from current location). In line with APT, number of bags to carry and weather belong to the most considered attributes. Interestingly, none of the CNET respondents considered "simplicity of the travel route" (APT 50%) and "parking costs" (APT 26.3%). This finding speaks to the biasing impact the revealed format has in APT. Only two inputs could not be interpreted in CNET. These were "bycicle" and a longer input about the alternatives the respondent would choose under different conditions. Hence, this results rather from a misconception of the task than from a failure of the string recognition tool.

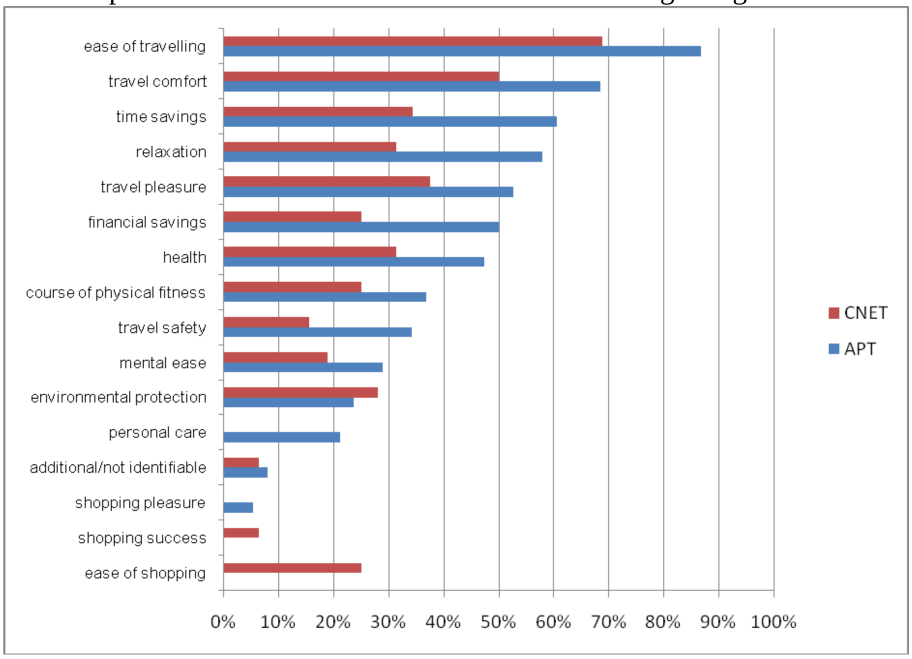


Figure 9. Elicited benefits for TM decision for CNET and APT.

The elicited benefits do not differ as much as the attributes for the transport mode decision (Figure 9). However, this would also be very surprising as most of the benefits in CNET are elicited like in APT in revealed format.

Only 15 benefits in total were recalled in CNET, the rest was ticked off from the list. CNET scores only lower because less attributes were elicited by this technique. Hence, a smaller number of underlying benefits is the logical consequence. Yet, both in CNET and APT interviews “ease of travelling” (APT 86.8%, CNET 68.8%) and “travel comfort” (APT 68.4%, CNET 50%) are the most frequent elicited benefits.

Shopping location (SL) decision

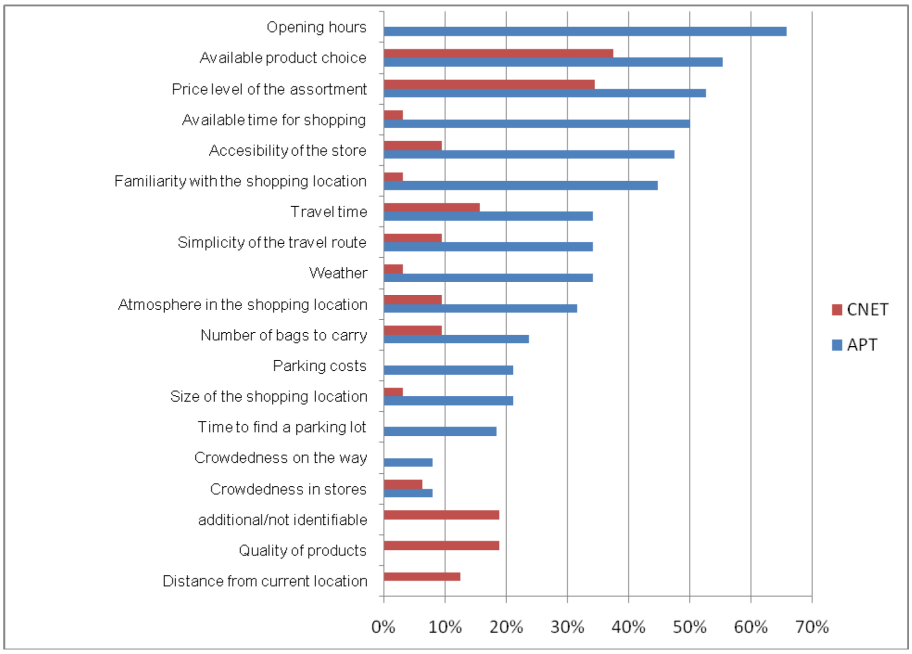


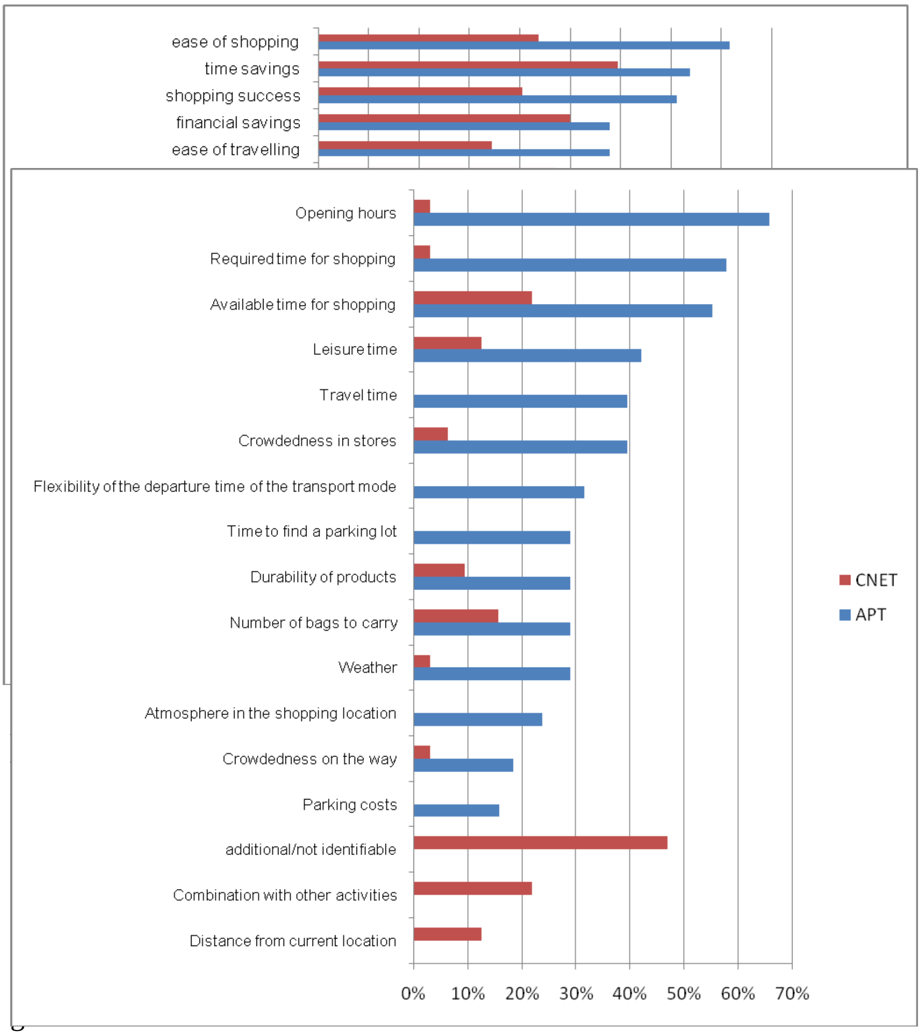
Figure 10. Elicited attributes for SL decision for CNET and APT.

Also for the shopping location decision all attributes that were provided in APT have been indicated as considerations by 8 to 66% of the respondents. Like for transport mode, none of the APT respondents indicated to consider an attribute which was not on the list.

In CNET the attributes were elicited less frequently than in APT. Attributes being indicated by less than 10% of the CNET respondents and not being provided in APT are not shown here. Hence, merely four attributes reach a somewhat frequent level of indication (the available product choice, price level, quality, and distance from current location). In line with APT, the available product choice and the price level belong to the strongest considered attributes (more than 30%). Strikingly, none of the CNET respondents considered “opening hours” (APT 66%). Probably, this obvious general condition has been taken for granted. Twenty CNET inputs could not be interpreted by the string recognition algorithm or the respondent rejected all suggestions, respectively. However, after a manual post-processing only six items remained which could not be assigned to an attribute label. Most of these items were statements about the preferred choices which might be caused by respondents' misconception of the task.

Figure 11. Elicited benefits for SL decision for CNET and APT.

Figure 11 lists the benefits which were elicited for the MRs for the shopping location decision by APT and CNET. As for TM, APT yields higher frequencies. However, the differences between APT and CNET seem to be



Timing of Shopping (TS) decision

Figure 12. Elicited attributes for TS decision for CNET and APT.

The attributes elicited for Timing of shopping (TS) differ extremely between APT and CNET (Figure 12). Again, all provided attributes were selected by at least 17% of the APT respondents. Also for TS no additional attributes were added in APT. In contrast, 21 CNET respondents typed in items which could not be interpreted by the string recognition algorithm. When checking the inputs afterwards it turned out that only six of them were really attributes. The other 15 inputs were either nonsense or respondents commented on their preferred choice option.

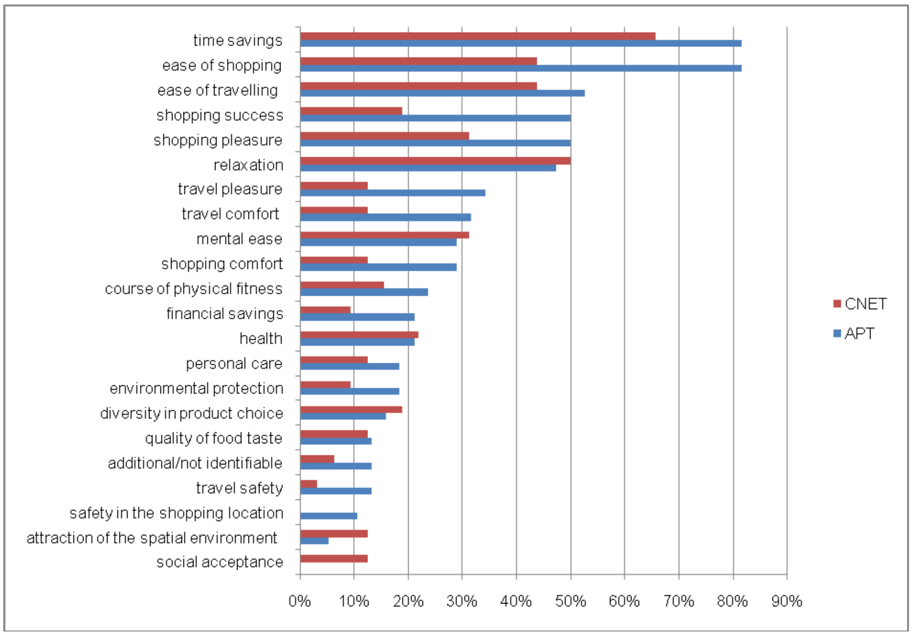


Figure 13. Elicited benefits for TS decision for CNET and APT.

Figure 13 shows the benefits which were elicited for the MRs for timing of shopping (TS) by at least 10% of the respondents. Although APT yielded in general higher benefit frequencies, five benefits were elicited more frequent in CNET. Four benefits were recalled in CNET, the rest was ticked off. Consistently, “time savings” is the most important benefit in both techniques (81.6% in APT vs. 65.6% in CNET). Further, frequently elicited benefits are “ease of shopping” (81.6%) and “ease of travelling” (52.6%) for APT and relaxation (50%), “ease of shopping” (43.8%) and “ease of travelling” (43.8%) for CNET. Three items were added to the list of benefits in CNET and nine in APT. However, only two of them could be interpreted as benefits afterwards.

4.4 Recruitment and interaction in online surveys

An issue which is not negligible concerns the low response rate of this study. Only 7% of the addressed households started the interview. Why 93% of the addressed households did not even start the online interview can only be speculated. Missing internet access can only excuse few societal groups from nonparticipating. Rather, an explanation could be that people do not feel encouraged by an impersonal invitation to an automatic online survey where personal contact to the researcher is lacking. Although the invitations were designed seriously with the official logo of Eindhoven University of Technology it might not imply trustworthiness among all addressees. Surely, also the need to start the computer, open the browser and type in the address of the survey website is a burden which does not occur in face-to-face interviews or paper-and-pencil questionnaires.
Also the high number of dropouts (50%) needs further discussion. When checking where exactly respondents left the interview it is striking that 71 of 139 dropouts (51%) happened before the decision variables had to be sorted. This may suggest that the instructions given in the introduction were not clear or too fatiguing or that the subject of research did not arouse interest among respondents. However, it may also mean that many respondents struggled with the sorting task as it required dragging and dropping the items on screen with the mouse. Although an instruction was provided, it has to be assumed that not all respondents read it carefully.

Another group of 27 respondents (19%) dropped out when facing the prompt to type in considerations (Figure 6) for the first ranked decision variable. Either this burden for respondents was too demanding or they did not expect the open format questions but a more common multiple choice questionnaire.

From the 178 respondents who typed in their considerations for the first decision, another 14 dropped out (10% of all dropouts) when they were asked to select a corresponding label among the suggestions from the string recognition algorithm. Apparently, this algorithm failed in finding proper labels which might have frustrated respondents.
The subsequent interview step (Figure 3) aiming at eliciting the underlying benefits caused the dropout of another 10 respondents (7% of all dropouts). Probably, thinking about this layer of the mental representation was too abstract for some respondents. All subsequent interview steps repeat the

earlier mentioned steps for the remaining two decision variables. Therefore, respondents are already somehow experienced with the task and the number of dropouts is much less.

Comparing APT (dropout rate 27%) to CNET (dropout rate 52%) the difference in dropouts is obvious. The higher mental effort, the somewhat longer interview duration, misinterpretations of the string recognition algorithm and the unexpected open format might be possible causes for the higher dropout rate for CNET.

Respondents who completed the interview had the chance to comment on it in a final step. These comments were grouped into three categories: scenario related comments, technique related comments and personal comments.

Typical scenario related comments regarded for instance elucidations of respondents´ activity-travel considerations or that the scenario did not match their real life situation. For these respondents, the comment box served mainly as a relief to finally express what could not be stated during the interview.

The second group of comments regarded statements about technical or procedural features of the corresponding interview technique. Some respondents criticized for instance the abstract questions or the long instructions.

Finally, a third group of comments comprises statements such as “interesting research”, “Good luck!” or “I would like to get informed about the outcomes of this survey.”

Number of comments and their relation to respondents is shown in Table 4.

Table 4. Respondent’s final comments.

		scenario	**technique**	**personal**
APT	# of comments	4	3	2
	relative	11%	8%	5%
CNET	# of comments	7	5	7
	relative	22%	16%	22%

According to Table 4 the tendency to comment is higher in CNET than in APT. As this could be expected for technique related comments, it is a bit striking for scenario/experimental related comments as the given situations do not differ. Furthermore, the stronger positive echo for CNET in terms of personal comments is even more surprising. However, statistical tests are not reliable as the observed frequencies are too little.

5. CONCLUSIONS

This paper compares two different online techniques, namely APT and CNET, in measuring MRs of 70 respondents for a fictive activity-travel task. This task consisted of interrelated choices for time of grocery shopping, shopping location and transport mode.

First of all, the paper proved that CNET can be brought online albeit with some concessions to its original protocol and some caveats like the lower response and the higher drop out rate. The complexity of online CNET is assumed not to be higher than for offline CNET. The threshold to drop out is in the anonymous online version only much lower.

The results of the study have clearly shown that MRs elicited by CNET are smaller than the MR elicited by APT. The number of associations, the total number of attributes and number of benefits are all significantly smaller in CNET than in APT. The explicit a priori listing of variables in the latter technique might thus trigger the mentioning of attributes which are not necessarily part of the MR. This explanation is supported by the fact that all attributes which are listed in APT were also indicated as being considered by at least some respondents. In order to check how respondents evaluated their opportunities to indicate (all aspects of) their considerations a post-experimental question addressed this issue ("Could you indicate all your considerations?"). On a scale from 1 (never) to 7 (always) APT scored highest (5.83). The difference to CNET (5.25) is, however, not significant. There is no correlation between this post experimental rating and the number of times the string recognition could not find a match (r = -0.047 with p = .800). Nevertheless, in order to guarantee a successful interpretation of respondents' inputs, a comprehensive pre-experimental collection of likely and unlikely attributes, benefits and their synonyms is unavoidable. This effort is, however, paid off for large-scale surveys by the relief that electronic data collection brings for post-experimental data processing.

Further research is necessary to check why the elicited attributes differ that much between the techniques. Apparently, some attributes are too obvious to be recalled (e.g. opening hours). Hence, they are highly indicated as considerations in APT but by far not in CNET. In contrast, attributes which score high in CNET, belong also to the most frequent considered attributes in APT (although to a less nominal extent). Exceptions are merely "the distance from the current location" for all three decisions, "quality of products" for SL decision and "combination with other activities" which could not be indicated in APT.

All three examples yield, however, interesting insights. The latter one was not provided in APT as the situational task did not include other activities on that certain day. Apparently, CNET evokes considerations which go beyond the fictive experimental task into the personal experiences.

Quality of products was not provided in APT and none of the APT respondents missed it. However, for CNET respondents it was the third most frequent mentioned attribute for the SL decision. This finding underpins our assumption that CNET supports respondents in rendering their MR consciously whereas APT hampers respondents in that.

Finally, the frequent indication of “distance” in CNET complies with the high indication of “travel time” in APT which is a hint that individuals naturally tend to think in spatial and not in temporal terms.

In conclusion then the question whether online versions of these techniques are to be preferred to face-to-face versions, whether CNET outperforms APT and even whether the conceptualization underlying these methods require simplification is open for further debate and empirical results, but first results reported here are encouraging.

6. REFERENCES

Arentze, T.A., B.G.C. Dellaert, and H.J.P. Timmermans (2008). Modeling and Measuring Individuals’ Mental Representations of Complex Spatio-Temporal Decision Problems. Environment and Behavior, Vol. 40, No. 6, 843-869.

Botschen, G. and Thelen, E. (1998). Hard versus Soft Laddering: Implications for Appropriate Use. In: New Developments and Approaches in Consumer Behavior Research (I. Balderjahn, C. Mennicken, E. Vernette, ed.), pp. 321-339. Stuttgart.

Dellaert, B.G.C., Arentze, T.A., Timmermans, H.J.P. (2008). Shopping context and consumers’ mental representation of complex shopping trip decision problems. Journal of Retailing, 84, 219 -232.

Horeni, O., Arentze, T.A., Timmermans, H.J.P., Dellaert, B.G.C. (2008). Interview techniques for measuring individuals’ mental representations in space-time choices. In: Proceedings International Conference on Design & Decision Support Systems. Leende, The Netherlands. CD-ROM: Steering Committee for the International Conference on DDSS.

Johnson-Laird, P. N., 1983, Mental models, Harvard University Press, Cambridge, MA.

Myers, J. H., 1976. Benefit Structure Analysis: A New Tool for Product Planning. Journal of Marketing, 40, 23-32.

Oppewal, H., H. Timmermans (1991), Context Effects and Decompositional Choice Modelling, Papers in Regional Science 70(2), 113 131.

Reynolds, T.J., Gutman, J. (1988). Laddering Theory, Method, Analysis and Interpretation. Journal of Advertising Research, 28(1), 11-31.

Russell, C.G., Flight, I., Leppard, P., van Lawick van Pabst, J.A., Syrette, J.A., and Cox, D.N. (2004a). A comparison of paper-and-pencil and computerised methods of “hard” laddering. Food Quality and Preference, 15, 279-291.

Russel, C.G., Busson, A., Flight, I., Bryan, J., van Lawick van Pabst, J.A., and Cox, D.N. (2004b). A comparison of three laddering techniques applied to an example of a complex food choice. Food Quality and Preference, 15, 569-583.

ter Hofstede, F., Audenaert, A., Steenkamp, J.-B.E.M., Wedel, M. (1998). An investigation into association pattern technique as a quantitative approach to measuring means-end chains. International Journal of Research in Marketing, 15, 37-50.

Tversky, A., Simonson, I. (1993). Context-Dependent Preferences. Management Science, 39(10), 1179-1189.