Hierarchical Information Integration Experiments and Integrated Choice Experiments

Eric J.E. Molin
Associate professor of Transport Policy
Delft University of Technology
P.O. Box 5015
2600 GA Delft, The Netherlands
Phone: +31 15 2788510
Fax: +31 15 82719
Email: e.j.e.molin@tudelft.nl

Harry J.P. Timmermans
Professor of Urban Panning and Chair
Urban Planning Group
Eindhoven University of Technology
P.O. Box 513
5600 MB Eindhoven, The Netherlands
Phone: +31 40 247 4684
Fax: +31 40 243 8488
Email: H.J.P.Timermans@tue.nl

ABSTRACT

The use of fractional factorial designs in the study of complex decision making, involving many attributes, implies the problem of information overload and respondent burden, potentially jeopardizing the validity of such experiments. To avoid or reduce the impact of these potential problems, Hierarchical Information Integration has been suggested. The key notion is to classify the large number of potentially influential attributes into a smaller set of decision constructs, construct separate experimental designs for each of these constructs, and in addition a bridging design that allows the scaling of all part-worth utilities into a concatenated utility expression. The basic approach suggested for preference measurements has been elaborated for other measurement tasks and the original design strategy has been refined into an alternative approach. This paper summarizes these developments and discusses aspects of respondent burden and validity

1 INTRODUCTION

An important practical limitation of stated preference applications is that as the number of attributes and/or attribute levels increases, the size and complexity of the experimental task increases exponentially. Increased experiment size implies that respondents have to evaluate more hypothetical profiles with possible respondent burden as a result. Increased task complexity means that respondents have to evaluate more attributes per choice alternative with information overload as a possible result. Respondents may react to increased size and task complexity by adopting simplifying decision heuristics, such as paying attention to only a subset of attributes (e.g., Arentze et al. 2003, Caussade et al., 2005, Hensher, 2006). As these heuristics probably do not reflect their decision making processes in the real world, this may result in less valid models.

In the past various solutions have been proposed to overcome the problems related to respondent burden in decision problems involving many attributes. Among these solutions are applications of self explication tasks, or pair wise trade-off tasks, constructing separate experiments with a limited number of attributes with a single overlapping attribute, and adaptive conjoint analyses. For a further explanation and comparison of these methods we refer to Oppewal et al 1994 and Pullman et al 1999. A problematic aspect that these methods have in common is their rather ad hoc nature.

An alternative approach that is well rooted in theory is Hierarchical Information Integration (HII), first proposed by Louviere (1984). HII is a logical extension of information integration theory (Anderson, 1974, 1981, 1982) and assumes that decision makers use multistage (or hierarchical) decision strategies to process information in complex decision making tasks. In particular, HII assumes that when decision makers have to evaluate complex decision alternatives involving many influencing attributes, they will first classify the attributes into a set of higher order constructs, also called decision constructs. Examples of these decision constructs in a transportation context include 'service quality’, ‘safety’, and ‘comfort’. In fact, every attribute which is not described in physical terms can be regarded as a decision construct, because an additional experiment is needed to examine how physical terms load on the construct. It is further assumed that decision makers first form impressions for each of the decision constructs separately, and then integrate these impressions into overall preference values for the decision alternatives.

Consistent with these theoretical assumptions, the implementation of conventional HII models requires the construction of two different experimental designs. First, a subexperiment for each construct is required to measure the trade-off between the attributes defining that construct. Next, a bridging experiment is required to measure the trade-off between the decision construct evaluations to examine how the evaluations of the decision constructs are integrated into an overall evaluation of the decision alternative. As this bridging task implies a trade-off of summarizing construct evaluations, this experiment involves a rather abstract task for the respondents. To overcome this potential limitation, Oppewal et al. (1994) proposed the Integrated HII

Choice Experiments method. Both HII variants will be further discussed and illustrated in this paper.

As will be shown in this paper, by applying the HII approach the trade-off between many attributes is measured in several subexperiments, while it is still possible to arrive at a single utility function. As each of the subexperiments contains less attributes compared to a conventional full profile experiment, information overload is reduced. Furthermore, as the subexperiments can be allocated across different respondents, further reduction of respondent burden is possible, of course assuming that the subsamples are sufficiently homogeneous. Thus, HII provides a way of dealing with many attributes while overcoming task overload and respondent burden.

HII has been applied in marketing to model telecommunication services (Louviere, 1984), supermarket choice (Louviere and Gaeth, 1987, Oppewal et al 1994, Oppewal et al 2006), bank preferences (Oppewal and Vriens, 2000), in spatial sciences to model residential choices (Timmermans 1989, Louviere and Timmermans, 1990a, Borgers et al 1992, Timmermans et al 1992, Van de Vijvere et al 1998, Oppewal and Klabbers, 2003, and Molin, 1999), recreational choices (Louviere and Timmermans 1990b, 1992). More recently, HII has also been applied to transportation (Hensher, 1991, Chiang and Chang, 2003, Bos et al., 2004, Norojono and Young, 2003, Molin and van Gelder, 2008), although it is still not widely known in this area of research.

The aim of this paper is therefore to review the HII approach in order to bring this approach to the attention of a wider transport community as a possible way of reducing task complexity and information overload in SP experiments. First, the conventional HII approach is presented, followed by the integrated HII approach and a combined variant. This is followed by an overview of HII applications that illustrates how HII can reduce the respondents' tasks. Validity aspects of HII are then discussed and the paper finishes by drawing some conclusions and discussing some issues for further research.

2 HIERARCHICAL INFORMATION INTEGRATION

In this section the three variants of HII and the assumptions on which these are based are described. First, conventional HII is discussed, followed by the integrated HII and finally, a combined variant.

2.1. Conventional HII

Conventional Hierarchical information integration is based on the following set of assumptions (Louviere and Timmermans, 1990a&b; Van de Vijvere et al., 1998).

Assumption 1: Any particular choice or decision is influenced by a set of variables that may be quantitative or qualitative in nature. Let the total number of influential attributes be N.

Assumption 2: Faced with a complex decision making problem, individuals group or categorize a set of influential attributes into subsets, which we term ‘decision constructs’. Such categorization allows the individual to process several variables into a single decision dimension or construct. Let the set of N total attributes be divided into I subsets. Let these subsets of decision attributes be denoted by G_i ($i = 1, \ldots, I$). For each subset, there is a set of $\{ V_{i1}, \ldots, V_{in} \}$ attributes that represents the subset or decision construct. The number of variables $N(i)$ in each subset need not be the same for all decision constructs, but for notational convenience we assume that $N(i)$ is the same for all subsets. The total number of decision variables is therefore the sum of the separate variables in each subset.

Assumption 3: Associated with each attribute level is a part-worth utility, the value of which is inferred from an analysis of the evaluations of or choices among multivariable alternatives that individuals make in statistically controlled experiments.

Assumption 4: In each of the G_i subsets, individuals make judgments by integrating their part-worth utilities defined by the levels of the set of attributes contained in a particular subset G_i. The integration process follows some combination rule that may differ in each subset, but can be approximated by a statistical model with linear parameters and attributes. Thus, each subset G_i is defined by some combination rule g_i.

Assumption 5: Individuals respond to experimentally designed combinations of the levels of the attributes corresponding to the ith subset G_i. The experimenter records these responses on some psychological measurement scale that represents a transformation of the true, but unknown, utilities of the levels of the attributes contained in the ith subset. Let the ith set of attributes, G_i, map onto decision-construct dimensions A_i. Hence,

$$\begin{aligned} A_i &= g_i (G_i) \\ &= g_i (V_{i1}, \ldots, V_{in}) \end{aligned}$$

Where $\{A_i\}$ defines the levels of the attribute dimensions or decision constructs of the I subsets of attributes.

Assumption 6: Individual choices among, or judgments about, combinations of levels of decision construct A_i are observed on a psychological measurement scale, R, in an experimental setting. The responses can be decomposed into part-worth utilities for each level of each decision construct by means of some mapping, f. This mapping, f, can be expressed as a statistical model which is linear in its parameters and variables as follows:

$$\mathrm{R} = f\,(A_i), \qquad i = 1, \ldots, I.$$

Hence

$$\mathrm{R} = f\,[\, g_i\,(V_{i1}, \ldots, V_{in})], \qquad i = 1, \ldots, I.$$

The previously described equations are too general for modeling purposes; in practical applications of HII, therefore, f and g_i must be specified. Many possible specifications might represent the various combination rules g_i and the overall mapping f; however, a useful rather robust form is an additive or multilinear form. However, regardless of the "true" specification, the HII approach involves the following steps (recall that the basic idea is to structure decision tasks in such a way that one can study and analyze separate integration processes for each decision construct).

(1) Attributes are clustered into I sets based on logic, theory, or empirical evidence. While the clustering in most applications HII is based on logic or theory, Bos et al. (2003) shows how clustering can be based on an empirical analysis of the grouping of attributes conducted by a sample of individuals.

(2) Separate experimental designs are constructed for each of the I sets identified in step (1), to create alternative descriptions defined by the various combinations of levels, positions, or degrees of the variables that define the decision constructs represented by each set. Individuals evaluate combinations of the attribute levels, or positions in each construct set, on a category rating scale (or other suitable scale) that defines 'how much' of the construct is defined by each combination of attribute levels. In the HII literature, these experiments are often called subexperiments.

(3) The response data obtained in step (2) are analyzed separately for each set (and when the construction of the experiment allows it, for each individual in each set) to develop a statistical model that describes how the different attributes combine to define each decision construct. For example, one might use multiple linear regression models to develop the statistical descriptions, but any statistical model that preserves the nature of the category response scale can be used, for example, an ordered logit or probit model (Hensher, 1991).

(4) Each (higher-order) decision construct is then treated as a factor whose levels are the numerical categories of the ratings scales used to define the constructs in step (2). For example, if a ten-category rating scale is used in step (2), one might select the categories 2, 5 and 8 from this scale to the levels, and construct a fraction of the 3^I

design (I is the total number of decision constructs). Individuals are told that the ratings reflect those that they gave to each decision construct in step (2). Louviere (1984) originally framed this experiment, which in the HII literature is called a bridging experiment, as a rating task. Hence, the individual's task in the bridging experiment is to evaluate the combinations of decision-construct ratings by ordering them on a new and different ratings scale. Timmermans (1988, see also Louviere and Timmermans, 1990) generalized this approach by framing the bridging task as a choice task, making the individuals choose between construct rating profiles.

(5) The data obtained in step (4) can then be rated and treated as measures of the individual's overall utility, which are subject to the normal error assumptions of the multiple linear regression or related general linear models (for example, ANOVA). If discrete choice data have been obtained in step (4), one of several possible limited dependent variable models such as binary or multinomial probit or logit models can be used to analyze the data. In this way a statistical model of the integration of the decision-construct ratings can be derived.

(6) The separate statistical models estimated in steps (3) and (5) can be concatenated if one assumes that each decision process has a separate error distribution with expectation 0, which is not correlated with the error distribution of the other decision processes.

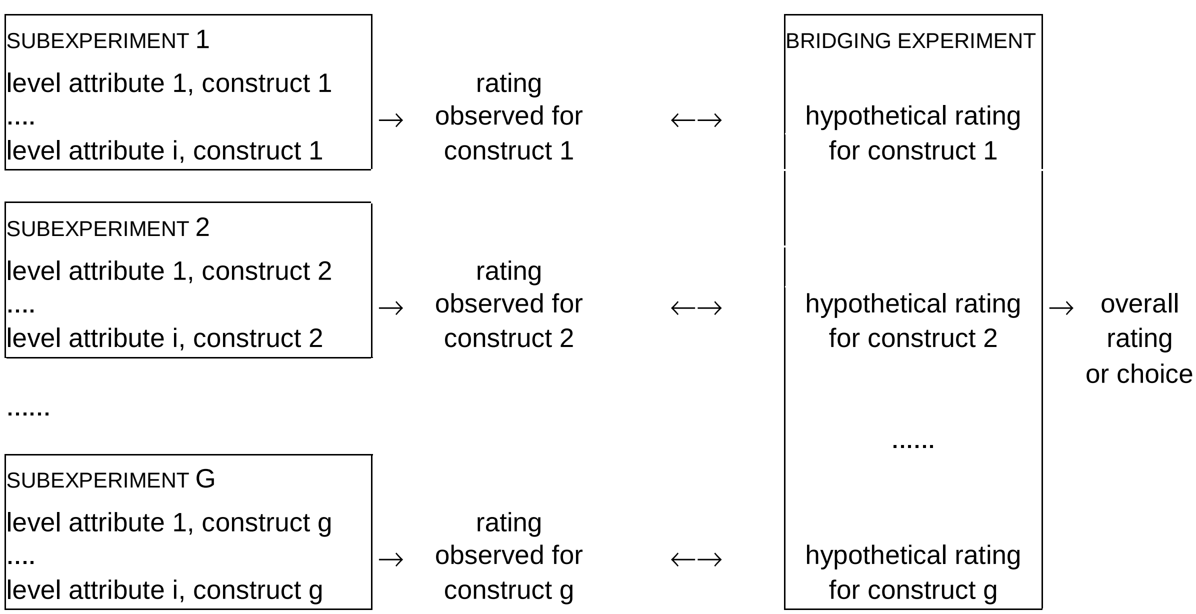

Figure 1 **Scheme of experiments underlying HII**

In sum, the application of conventional HII involves the construction of a subexperiment for each distinguished decision construct to evaluate each decision construct and a bridging experiment to model the integration of the decision construct evaluations, as summarized in Figure 1.

2.2. Integrated choice experiments

Oppewal et al. (1994, see also Van de Vijvere et al 1998) identified a series of potential limitations to the original HII approach. Firstly, they argued that the original HII approach produces several models rather than a single model for which overall goodness- or badness-of-fit measures and validity tests can be derived. A concatenated overall model cannot be estimated directly; rather, model parameters are derived by the substitution and replacement of terms of various submodels that are estimated separately. Hence, the attributes varied in each subexperiment are not related directly to the final response of interest, be it preference or choice.

Secondly, the values of remaining decision constructs are not specified in each separate subexperiment because one assumes that they have no systematic effect on the evaluations of a particular decision construct. Consequently, respondents may have to assume or infer values for the other decision constructs (e.g., Johnson, 1987). Hence, attribute effects are only tested over a limited range of values of other decision constructs, and there is no control over a respondent's inferences about the values of other decision constructs.

Thirdly, the validity of the bridging experiment poses problems because respondents have to evaluate or choose profiles described in terms of their profile ratings in the subexperiments. The difficulty of this task is unclear. Also it is not clear whether the resulting attribute evaluations reflect the respondent's real preferences. For example, profiles in bridging tasks typically contain only numerical scores, which may encourage respondents to average out their evaluations of the decision constructs. Hence, the parameters estimated from bridging experiments may not be valid.

Fourthly, the original HII approach does not test the assumed hierarchical decision structure. Instead, one must assume that the hierarchical structure is correct to concatenate the subexperiments.

Fifthly, although the bridging experiment can be designed as a choice experiment, Louviere's original approach does not allow subexperiments to be framed as choice experiments. Because there may be situations in which one wishes to model choice behavior, there may be a need to design subexperiments as choice experiments.

Finally, interactions between variables that define different constructs cannot be estimated, nor can interactions between attributes and decision constructs.

As an alternative to the original HII approach, Oppewal et al. (1994) proposed and tested an HII approach based on integrated subexperiments. Like the subexperiments of the conventional HII approach, the profiles of the subexperiments in the integrated approach include attributes that define a particular decision construct. However, in addition to these attribute profiles, summarizing hypothetical evaluation scores of the other decision constructs are added to the profile. As with the bridging experiments of the conventional HII approach, these hypothetical evaluation scores are typically expressed as ratings on a rating scale. The resulting profiles describe all major aspects of a choice alternative, in terms of a series of attributes plus hypothetical construct ratings. The required subexperiments are schematically depicted in Figure 2.

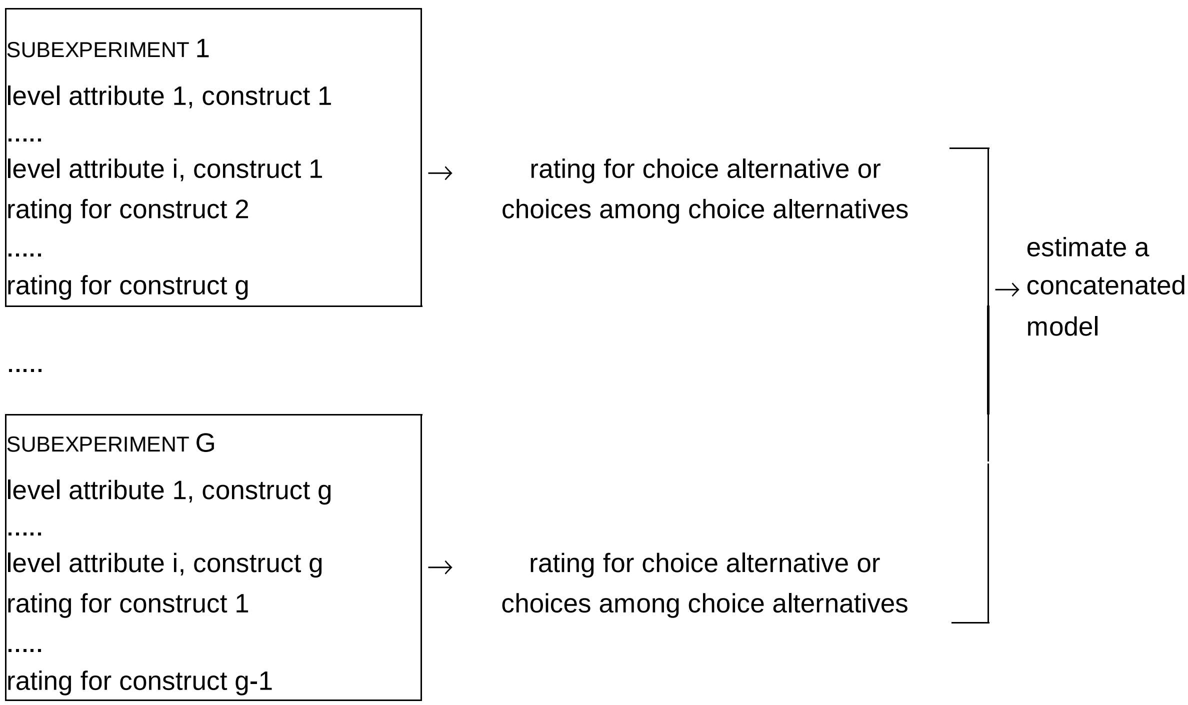

Figure 2 Scheme of experiments underlying HII with integrated subexperiments

The HII approach with integrated subexperiments overcomes most of the limitations of the original HII approach. Firstly, the profile evaluations can be regarded as overall evaluations because all the decision constructs are described in the profiles, either as combinations of attribute levels or as summarizing ratings of constructs. Consequently, all separate subexperiments can be concatenated to estimate a single model. Secondly, respondents do not have to make inferences about the omitted decision construct levels, because all decision constructs are specified in each profile. This gives enhanced control over the respondents' inferences, which potentially increases the reliability of the estimated parameters. Thirdly, the need for a bridging experiment no longer exists, because the profiles describe the complete choice alternative. The resulting profiles consist of a mix of attribute level descriptions and the numerical ratings of decision constructs which reduces the probability that respondents adopt simplifying response heuristics, like averaging. Fourthly, the validity of the assumed HII decision-making process can be tested (Oppewal et al., 1994 and van de Vijvere et al., 1998). These validity tests will be discussed in the fourth section. Fifthly, the subexperiments can be designed as choice experiments. Finally, the interactions between attributes and decision constructs can be estimated, although it is still not possible to estimate the interactions between variables that define different constructs. Thus, the HII approach with integrated subexperiments avoids the limitations identified for the original HII approach, whilst preserving its benefits.

2.3 Combined conventional HII and integrated choice experiments

In the transportation literature several studies have been published that combine elements of conventional HII and integrated choice experiments (Hensher, 1991, Norogono and Young, 2003, Bos et al, 2004 and Molin and van Gelder, 2008). In this HII variant, subexperiments are constructed in the same way as in the conventional HII approach, thus varying only the attributes that define a single decision construct. The construction of the bridging experiment however differs in the sense that in addition to the experimentally varied construct evaluations, other attributes of interest are included. The result is that the bridging experiment is framed as an integrated choice experiment in the sense that it combines summarizing construct evaluations and regular attributes.

Bos et al. (2004) initiated the application of this variant by referring to their experiences with an earlier experiment that varied multi-modal mode alternatives (van der Heijden and Molin, 2002). In that experiment, different time-related attributes were varied, such as waiting time and travel time, but respondents indicated that they summed the different time components and responded to total travel time. Indeed, model estimation indicated that the coefficients for the different time components did not differ significantly. Assuming that respondents would also sum different cost components and would only respond to total costs, Bos et al. decided not to treat time and cost as decision constructs and therefore did not construct subexperiments for those attributes. Instead they included time and costs as regular attributes in the bridging experiment in addition to the decision construct evaluations that did refer to the subexperiments. Molin and van Gelder (2008) and Norojono and Young (2003) also included a regular time and cost attribute in addition to the construct evaluations in the bridging experiment, and Hensher included a cost and an interchange attribute in the bridging experiment.

Although all applications of the combined HII approach were in the transportation field, this HII variant is in fact a general approach that can be applied in any situation in which, in a stated preference experiment, the researcher wishes to include one or more complex attributes of which the levels express summarizing subjective evaluations, like, for example, perception scores on comfort, convenience, and safety. If the interest is to understand how the subjective evaluations are related to a bundle of underlying physical attributes, one could construct a subexperiment in which the underlying physical (objective) attributes are experimentally varied and the subjective evaluations of the complex attribute are measured. Such a subexperiment links subjective evaluations to physical attributes, thus providing managerially relevant information on how subjective evaluations may be changed by changing the levels of physical attributes.

3 AN OVERVIEW OF HII APPLICATIONS

To give the reader an idea about the way HII has been applied in various studies and to illustrate how HII reduces the respondents' task, this section provides an overview of HII applications, summarized in Figure 3. To the best of the authors' knowledge, this figure includes all published HII studies with the following two exceptions. The first exception concerns three papers that applied a modified version of HII (Úlengin, 1998, Úlengin et al. 2001 and 2002). In these studies none of the attributes in the bridging experiment are expressed in terms of the rating scales used in the subexperiments and therefore cannot be directly linked to the construct evaluations as observed in the subexperiments. Hence, the modified HII approach is not based on the assumptions underlying HII as discussed above and is therefore not included in this review. A second exception is formed by studies that apply HII to study group (cq. family) decision making (Timmermans et al. 1992, Borgers and Timmermans, 1993, Dellaert et al., 1998, and Molin et al., 1999, 2000 and 2003). These studies include a summarizing construct evaluation in the bridging experiment or integrated experiments separately for every group member, which allows, for example, the researcher to study the influence members have on the group decision. As the application of HII to study group decision making requires additional assumptions, this is beyond the scope of this paper. Hence, in this overview we focus on HII applications for individual decision making.

The third column of the overview indicates that most of the studies (6) applied the conventional HII variant, 3 studies applied the integrated choice variant and 4 studies applied the combined approach. As one may argue that the combined variant is more similar to the conventional approach than to the integrated variant, we can conclude that the conventional variant is the most popular approach. The fourth column includes a description of the decision constructs that were used, for illustrative purposes, to give the reader an idea of the type of decision constructs that have been used.

The last two columns of Figure 3 illustrate how HII may reduce the respondents' tasks. As has already been argued, the partition of the total number of attributes across different subexperiments limits the number of attributes per profile and consequently reduces the possibility of information overload. The fourth column indicates that the number of attributes in most of the subexperiments is indeed quite limited. In the large majority of the applications the number of attributes in each experiment is between 3 and 7. As in the integrated variant, in addition to the attributes defining a decision construct, summarizing construct evaluations for all other decision constructs need to be included. The integrated variant has a slightly lower potential for reducing information overload than the conventional variant. Column 4 indeed illustrates that overall more attributes are included in the experiments of the integrated variant than in experiments of the conventional type.

Also, as argued before, HII has the potential to reduce the number of profiles which results in a smaller respondent burden. Louviere and Geath (1988) suggested that this can be achieved by assigning only a single subexperiment together with the bridging experiment to each respondent. However, the same authors also suggest that as each subexperiment of the HII is different, individuals can probably be asked to respond to more than a single task. In any case, the completion of a complete subexperiment has the advantage that if ratings are used as the response format, one can model the construct evaluations at the individual level. Column 5 of Figure 3 indicates that in five of the applications respondents are indeed requested to complete only a single subexperiment and the bridging experiment. Some of these applications also required respondents to complete a few profiles of each of the other subexperiments in order to familiarize themselves with the attributes used in those experiments so that they have a better understanding of the meaning of all decision constructs. In contrast, three applications required the respondents to complete all the experiments, which therefore did not reduce the respondents' task at all. Finally, four of the applications requested the respondent to complete an equal part of all the constructed experiments, ranging from evaluating only a single profile to evaluating half of the profiles in each subexperiment.

The study reported by Oppewal and Klabbers (2003) is the only one that actually measured the task load of HII-based and of full profile experiments and compared the results. The amount of time taken to complete the first experiment that was completed was measured. A full profile task involving the rating of 27 profiles each including 13 attributes was compared with the completion of 4 rating-based subexperiments and a bridging experiment each involving 9 profiles. They found that a full profile experiment took less time to complete than all the HII subexperiments that applied the same order of the attributes. However, another partition of attributes across decision constructs resulting in an easier attribute order took the respondents less time to complete than the full profile experiment. Hence, although all HII experiments in total involve the rating of more profiles than the full profile design (each including less attributes), this result indicates that it does not take more time to complete HII experiments than the full profile experiment and that the task load depends on the logic of the division of the attributes across the decision constructs as perceived by the respondents which may result in tasks that are easier to process and consequently take less time to complete.

In addition, some of the applications asked the respondents to evaluate several full profiles that included all the attributes in order to compare the full profile and HII approach. Hence, it should be noted that several of the HII applications in our overview are specifically conducted for methodological purposes to test validity aspects. Hence, these HII applications demand more from the respondents than HII applications set up with only a substantial focus in mind. Nevertheless, Figure 3 illustrates the possibilities for reducing potential respondent burden by distributing the total task across respondents.

author	**subject**	**variant**	**experiments (number of attributes / number of profiles / CS = choice sets)**	**task per respondent**
Louviere, 1984	telephone services	conventional HII, rating bridging	costs (3 / 8) services (3 / 8) bridging(2 / 9)	all = rating 27 profiles *(+ 16 full profiles)**
Louviere and Gaeth 1988	supermarket choice	conventional HII, rating bridging	price (4 / 9) quality (3 / 16) selection (4 / 16) convenience (11 / 16) bridging (4 / 25)	all = rating 83 profiles
Louviere and Timmermans, 1990a & 1992	recreational destination choice	conventional HII, choice bridging	environment/accessibility (5 / 16) facilities (7 / 16) maintenance (4 / 16) social use (3 / 16) bridging (4 / 9 / 16 CS)	1 subexp + 4 profiles of each other subexp. = 28 ratings + 16 choices
Timmermans, 1988 and Louviere and Timmermans, 1990b	residential choice	conventional HII, choice bridging	house (6 / 16) residential environment (6 / 18) relative location (6 / 18) social and economic ties (4 / 8) bridging (4 / 9 / 16 CS)	1 subexp + 4 profiles of each other subexp. = max. 30 ratings + 16 choices
Oppewal and Klabbers, 2003	residential choice	conventional HII, choice bridging	size of living quarters (3) size of sleeping quarters (3) room positions (4) sunlight (3) - (13 / 81**) bridging (4 / 9)	9 full profile + 27 subexp + 9 bridging exp = 56 ratings + 8 choices
Chiang et al. 2003	Intercity mode choice	conventional HII, choice bridging	service quality (6 / 18) transfer quality (4 / 18) information quality (3 / 18) environment quality (3 / 18) bridging (4 / 18(?) / 9(?) CS)	1 profile from each subexp. = 4 ratings and 1 choice
Oppewal et al. 1994 & 1995	shopping centre choice	integrated HII, choice tasks	location convenience and accessibility (8 / 128 / 48 CS) appearance, layout and furnishing (13 / 128 / 48 CS)	9 choices

			selection of stores for food and packaged goods (12 / 256 / 96 CS) selection of stores for clothing and shoes (12 / 256 / 96 CS)	
Van de Vijvere et al, 1998	residential choice	integrated HII, choice tasks	house (7 / 32 / 16 CS) residential environment (5 / 32 / 16 CS) relative location (8 / 32 / 16 CS) full profiles (14 / 64 / 32 CS)	1 subexp. + 2 choice set of full profile = 18 choices
Vriens et al. 1995 & Oppewal and Vriens, 2000	bank preferences	integrated HII, rating tasks	accessibility (7 / 27) competence personnel (5 / 27) accuracy and friendliness (9 / 27) tangibles (7 / 27)	1 subexp. / 2 ratings per profile = rating 27 profiles
Hensher, 1991	mode choice (bus preferences)	combined conventional and integrated HII, rating bridging	trip quality (4 / 8) wait quality (3 / 8) vehicle quality (5 / 8) information quality (5 / 8) bridging (6 / 8)	4 profiles from each experiment = 20 ratings
Norojono and Young, 2003	freight mode choice	combined conventional and integrated HII¸ choice bridging	quality (4) reliability (3) bridging (4)	?
Bos et al, 2004 & 2006	destination Park and Ride choice	combined conventional and integrated HII	P&R facilities (7 / 18) connecting public transport (4 / 9) bridging (4 / 18 CS)	all = 27 profiles and 18 choices
Molin and van Gelder, 2008	origin Park and Ride choice	combined conventional and integrated HII	public transport access node (6 / 18) P&R facilities (7 / 18) connecting public transport (4 / 9) bridging (5 / 18 CS)	1 subexp = max. 18 profiles + bridging = 18 choices

* HII and full profile results were compared.
** the profiles for all subexperiments are simultaneously abstracted from a single experimental design.

Figure 3 An overview of HII applications

4 VALIDITY

4.1 Validity tests for conventional HII

In this section, we review the validity issues that have been addressed in the studies included in our overview. A first validity test that was conducted in several studies involves comparing the results based on the HII experiments with the results of a single full profile task including all attributes. It should be noted however, as Louviere and Gaeth (1988) pointed out, that the results of the full profile experiment are not necessarily more realistic. If indeed individuals as assumed follow a hierarchical strategy, than the results based on the HII experiments are probably more realistic. Hence, it may be argued that by increasing the number of attributes it becomes more likely that a hierarchical process can be followed.

A comparison of HII with full profile results was first conducted by Louviere (1984) in an application based on rating tasks only and that included a relatively small number of 8 attributes (see Figure 3). He found that the scale values of both methods were very strongly related by a linear transformation, a permissible transformation with interval scales. In an application that included many more attributes (22), Louviere and Gaeth (1988) found stronger differences in coefficients estimated by both methods. However, using the HII results to predict the responses observed in the full profile experiment, they concluded from the graph that plotted both results that the means were monotonically, but not linearly related. Hence, they concluded that, if both models were used in a choice predictor to predict choices based on the 'highest predicted response equals first choice', both methods would produce similar aggregate results. Finally, Oppewal and Klabbers (2003) compared two different partitions of attributes across subexperiments with a full profile approach in a study including 13 attributes. They found that the part worth utilities based on the full profile experiment have high correlations with those based on both HII experiments (0.85 and 0.95 respectively). Furthermore, they found higher hit rates for both HII models than for the full profile model (0.81 and 0.81 as opposed to 0.77), although the differences were not statistically significant.

The study by Louviere and Timmermans (1992) is the only conventional HII study that tested the external validity of HII models. They applied the conventional HII method with a choice bridging experiment in a recreational choice study involving 4 decision constructs. They tested how well the single choice model derived from combining the parameter estimates obtained from the five different experiments could predict reported choices. The model was used to predict the choice probabilities for each of 40 real-choice recreational areas in their study area and compared this with the observed shares based on the reported visits to these areas of the same respondents. They found that the product-moment correlation between predicted and observed shares of visits to each recreation area was 0.68, and the relationship was reasonably linear. This close relation between predicted and observed shares provides support for the predictive validity of the HII method.

Molin and Timmermans (2003) examined whether individuals follow a hierarchical process to arrive at overall residential preferences. Their analysis was based on revealed preference data that involved descriptions of respondents' current house and residential environment in a number of attributes and their evaluations for house, residential environment and overall housing situation. They hypothesized that according to HII theory, the housing evaluation should only be affected by housing attributes; residential environment evaluation only by residential environment attributes; and finally, the overall evaluation should only be influenced directly by the housing and residential environment evaluations and thus not directly be influenced by any of the attributes. The path model they estimated based on the assumed structure fitted the data well. From this analysis it can be concluded that individuals indeed followed a hierarchical process to arrive at their overall preferences, at least in the case of residential preferences. Testing whether a hierarchical process is followed cannot be tested within the conventional HII approach but it is possible within the integrated variant, which is discussed next.

4.2 Validity test for integrated HII

Oppewal et al. (1994) proposed and applied four validity tests for the integrated HII variant. We briefly summarize these tests here and refer to Oppewal et al. (1994) for more details. The first three validity tests involve regression analysis and require additional measurements: in each subexperiment the attribute combinations should be evaluated on the same scale as is used to express the experimentally varied decision construct evaluations.

A first validity test involves regressing the observed construct ratings against the attributes defining a single decision construct and the hypothetical construct ratings. If the construct is well defined, all attributes should have significant coefficients. Furthermore, to test whether the evaluation of a particular construct is independent of the hypothetical ratings of the other decision constructs, none of the coefficients estimated for the hypothetical construct ratings should turn out to be statistically significant in this construct-regression model.

A second validity test examines whether each attribute is well presented by a decision construct. This test involves comparing the significance of an attribute both in the construct-regression and the overall concatenated model. If an attribute is significant in the overall concatenated model but not in the construct-regression model, the attribute is not well represented by the decision construct.

A third validity test involves a test of process equality across subexperiments. According to the assumptions underlying HII, the construct evaluation based on the profile of attributes defining a construct should have the same effect on the overall evaluation as a hypothetical construct rating. To test this, the construct-regression model for a particular decision construct is used to predict the decision construct ratings for each profile in a particular subexperiment. Then a concatenated model is estimated across all subexperiments including only the hypothetical construct ratings and the predicted construct ratings. The predicted construct ratings are then tested to see whether each construct affects the overall evaluation to the same extent as the hypothetical construct ratings.

A fourth validity test can be performed without measuring the construct ratings. It involves testing whether parameters estimated for the same hypothetically defined constructs differ between different subexperiments. In theory, if each subexperiment represents the same choice process, the parameters of decision constructs should be the same except for sampling error and differences in error variability across the different subexperiments. Therefore, one expects no significant construct effect differences between subexperiments that both included the construct as a hypothetical rating. It should be noted that this test requires at least three decision constructs to be distinguished in order for the decision constructs to appear in at least two subexperiments as hypothetical ratings.

Oppewal et al. (1994) conducted all these validity tests in their empirical application on shopping centre choice, distinguishing 4 decision constructs (see Figure 3). Reporting the results of two subexperiments, they found that all attributes were significant both in the construct-regression models and in the overall concatenated choice models. Furthermore, they found that, except for a single quadratic effect, none of the hypothetical construct ratings had a significant effect on the construct-regression models. Hence, the results on the first outlined validity tests were in line with the expectations based on HII theory.

With respect to the third validity test, they found that the construct ratings predicted from the attribute combinations generally had a larger effect on overall choice than the hypothetical construct ratings. With respect to the fourth validity test, they found that hypothetical construct ratings have different impacts in different subexperiments. These results suggest that construct effects were not equal across subexperiments, even after taking scale difference between the different subexperiments into account (Swait and Louviere, 1993). This suggests caution in using the concatenated choice model for predictive purposes because the model may be context sensitive. However, the authors suggest that it could be that context effects “average out” across subexperiments, but this could not be directly tested. Instead, the predictive validity was tested, which was based on comparing the predicted choices for 30 shopping destinations with the observed choices for those destinations observed for the same sample. They found a correlation of .42, which was substantially higher than predictions based on any restricted model. From this result they concluded that inclusion of more attributes increases the predictive validity of experimentally based models.

Van de Vijvere et al. (1998) conducted the fourth validity test in an integrated HII application on residential choices. In this application they included 14 attributes partitioned across 3 decision constructs (see Figure 3). In contrast to the findings reported by Oppewal et al (1994), van de Vijvere et al. could not reject the hypothesis of equal construct parameters across different subexperiments. In addition, they also compared the utility function based on the integrated choice experiments with the utility function based on a full profile experiment. Again, they could not reject the null hypothesis of equal parameter estimates. Furthermore, as they corrected for differences in scale value, they found that the scale factor for the full-profile experiment was relatively small, which implies that the error in the full-profile task is larger than in the HII tasks.

CONCLUSIONS

In this paper, we have reviewed the background and quintessence of variations in Hierarchical Information Integration. Several examples were discussed. The distinct feature of methods of Hierarchical Information Integration is to break down the large number of potential influential attributes into smaller subsets. Because subexperiments are constructed for each construct separately, subjects receive less information when asked to make judgments, implying that the experiments are less demanding. On the other hand, because subjects are, in addition, requested to complete a bridging experiment, some of this gain is lost in the sense that the total number of profiles or choice sets to evaluate becomes higher again and perhaps even higher than the number when using a conventional fractional factorial design approach.

The studies examining and comparing the validity of Hierarchical Information Integration and conventional conjoint measurement suggest that the derived utility scales tend to be strongly related, even if a substantial number of attributes or choice sets is involved. These findings suggest that researchers have a choice as to what design strategy to use for choice problems of this size. Only if the total number of attributes becomes very high, conventional design strategies no long seem feasible and Hierarchical Information Integration becomes a feasible option.

REFERENCES

Anderson, N.H. (1974), Information Integration Theory: A Brief Summary, in: D. Kranz, R. Atkinson, D. Luce, P. Suppes (eds.), *Contemporary Developments in Mathematical Psychology*, W.H. Freeman, Sans Fransisco, 236-305.

Anderson, N.H. (1981), *Foundations of Information Integration Theory*, Academic Press, New York.

Anderson, N.H. (1982), *Methods of Information Integration Theory*, Academic Press, London.

Arentze, T., A. Borgers, H. Timmermans and R. DelMistro (2003), Transport Stated Choice Responses: Effects of Task Complexity, Presentation Format and Literacy, *Transportation Research Part E*, 39, 229-244.

Borgers, A., and H. Timmermans (1993), Transport Facilities and Residential Choice Behavior: a Model of Multi-Person Choice Processes, *Papers in Regional Science*, 72, 45-61.

Bos, Ilona D.M., Rob E.C.M. Van der Heijden, Eric J.E. Molin and Harry J.P. Timmermans (2004), The Choice of Park & Ride Facilities: An Analysis Using a Context-Dependent Hierarchical Choice Experiment, *Environment and Planning A*, 36(9), 1673 – 1686.

Bos, D.M and E.J.E. Molin (2006), Is there a 'Stick' Bonus? A Stated Choice Model for P&R Patronage incorporating Cross-Effects, *European Journal of Transport and Infrastructure Research, 6(3), 275-290.*

Caussade, S., J. de D. Ortuzár, L.I. Rizzi and D.A. Hensher (2005), Assessing the Influence of design dimensions on stated choice experiment estimates, *Transportation Research Part B*, 39, 621-640.

Chiang, Y.S., J.L. Lu and H.C. Chang (2003), Modeling the effect of destination attributes on the intercity Travelers' mode choice behavior in Taiwan area, *Proceedings of the Eastern Asia Society for Transportation Studies*, Vol.4, October.

Dellaert, B.G.C., M. Prodigalidad and J.J.Louviere (1998), Using conjoint analysis to study family travel preference structures: a comparison of day trips and 1-week holidays, *Tourism Analysis*, 2, 67-75.

Johnson, R.D. (1987), Making Judgments when Information is Missing: Inferences, Biases, and Framing Effects, *Acta Psychologica*, 66, 69-82.

Hensher, D.A. (1991) Hierarchical stated response designs and estimation in the context of bus use preferences. *Logistics and Transportation Reviews* 26(4): 299-323.

Hensher, D.A. (2006), How do respondents process stated choice experiment? Attribute consideration under varying information load, *Journal of Applied Econometrics*, 21, 861-878.

Heijden, R.E.C.M. van der, and E.J.E. Molin (2002), Locating P&R facilities based on Travel Behaviour: A Dutch case study, in: *Proceedings of the 7-th Urban Transport Conference*, Sevilla, Spain.

Louviere, J.J. (1984), Hierarchical Information Integration: A New Method for the Design and Analysis of Complex Multiattribute Judgement Problems, In: T.C. Kinnear (ed.), *Advances in Consumer Research*, vol. 11, Association for Consumer Research, Provo, Utah, 148-155.

Louviere, J. J., and G. J. Gaeth (1987), Decomposing the determinants of retail facility choice using the method of hierarchical information integration: A supermarket illustration. *Journal of Retailing*, *63*, 25-48.

Louviere, J.J., and H.J.P. Timmermans (1990a), Hierarchical Information Integration Applied to Residential Choice Behavior, *Geographical Analysis*, 22, 127-145.

Louviere, J.J., and H.J.P. Timmermans (1990b), Using hierarchical information integration to model consumer responses to possible planning actions: Recreation destination choice illustration, *Environment and Planning A*, 22, 291-309.

Louviere J.J., and H.J.P. Timmermans (1992), Testing the external validity of hierarchical conjoint analysis models of recreational destination choice, *Leisure Science*, 14 179 – 194.

Molin, E.J.E. (1999), Conjoint Modeling Approaches for Residential Group Preferences, Faculteit Bouwkunde, Technische Universiteit Eindhoven, Bouwstenen 53 *(dissertation).*

Molin, E.J.E. and M. van Gelder (2008), Freeway access to public transport: A Hierarchical Multimodal Choice Model, paper presented at TRB's annual meeting, Washington D.C. (CD-ROM).

Molin, E.J.E., H. Oppewal and H.J.P. Timmermans (1999), Group-based versus Individual-based Conjoint Preference Models of Residential Preferences: A Comparative Test, *Environment and Planning A*, 31(11), 1935-1947.

Molin, E.J.E., H. Oppewal and H.J.P. Timmermans (2000), A Comparison of Full Profile and Hierarchical Information Integration Conjoint Methods to Modeling Group Preferences, *Marketing Letters*, 11(2), 165-175.

Molin, E.J.E., H. Oppewal and H.J.P. Timmermans (2003), Conjoint Modeling of Residential Group Preferences: A Comparison of Hierarchical Information Integration Approaches, *Journal of Geographical Systems*, 4(4), 343-358.

Molin, E.J.E. and H.J.P. Timmermans (2003), Testing Hierarchical Information Integration Theory: the causal structure of residential satisfaction, *Environment and Planning A*, 35(1), 43-58.

Norojono O, W. Young, (2003), A stated preference freight mode choice model, *Transportation Planning and Technology*, 26, 195 – 212.

Oppewal, Harmen, M. Klabbers (2003), Compromising between information completeness and task simplicity: a comparison of self-explicated, hierarchical information integration, and full-profile conjoint methods, in: Advances in Consumer Research, volume 30, eds. Punam Anand Keller and Dennis W. Rook, Valdosta, GA : Association for Consumer Research, 298-304.

Oppewal, H., J.J. Louviere, and H.J.P. Timmermans (1994), Modeling Hierarchical Conjoint Processes with Integrated Choice Experiments, *Journal of Marketing Research*, 31, 92-105.

Oppewal, H. and M. Vriens (2000), Measuring perceived service quality using integrated conjoint experiments, *International Journal of Bank Marketing*, 18, 154-169.

Pullman, Madeleine, Kimberly J. Dodson and William L. Moore (1999), A Comparison of Conjoint Methods When There Are Many Attributes, *Marketing Letters* 10 (2), 123-138.

Swait, J. and J.J. Louviere (1993), The role of the Scale Parameter in the Estimation and Use of Generalized Extreme Value Models, *Journal of Marketing Research*, 30, 305 – 314.

Timmermans, H.J.P. (1989), Een Decompositioneel Hiërarchisch Model voor Woningkeuze: Theorie en Illustratie, In: S. Musterd (ed.), Methoden voor Woning- en Woonmilieubehoeften Onderzoek, Siswo, Amsterdam, 46-72 *(in Dutch)*.

Timmermans, H.J.P., A.W.J. Borgers, J. van Dijk, H. Oppewal (1992), Residential Choice Behaviour of Dual-Earner Households: A Decompositional Joint Choice Model, *Environment and Planning A*, 24, 517-533.

Ülengin, B. (1998), Using Hierarchical Information Integration to examine customer preferences in banking, *International Journal of Bank Marketing*, 16, 202-210.

Ülengin Burç, Füsun Ülengin and Ümit Güvenç (2002), Living environment preferences of the Inhabitants of istanbul: a modified Hierarchical information integration model, *Social Indicators Research*, 57, 13–41.

Ülengin, B., F. Ülengin and Ü. Güvenç (2001), A multidimensional approach to urban quality of life: The case of Istanbul, *European Journal of Operation Research* 130, 361–374.

Van de Vyvere, Y., Oppewal, H. and Timmermans, H. (1998), The validity of hierarchical information integration choice experiments to model residential preference and choice, *Geographical Analysis*, 30(3), 254-272.

Vriens, M., H. Oppewal, J.A.Linde, and H. Bremer (1995), Simultane analyse van preferenties en percepties bij diensten: Een toepassing op bankdiensten, *Marketing*, 613-625 *(in Dutch)*.