# Modeling importance assessment processes in non-routine decision problems

PROEFSCHRIFT

Ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof.dr. F.A. van Vught,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op woensdag 26 november 2003 te 13.15 uur

door

Johannes Martinus Gerardus Heerkens

Geboren op 16 september 1958

te Gouda

Dit proefschrift is goedgekeurd door:

| | |
|---|---|
| Prof.dr. A. van Harten | (promotor) |
| Dr. B.I.J.M. van der Heijden | (assistent-promotor) |
| Dr. M.J. van Riemsdijk | (assistent-promotor) |

# Modeling importance assessment processes in non-routine decision problems

Promotiecommissie (dissertation committee):

| | |
|---|---|
| Prof. P. Lawrence | Loughborough University |
| Prof.dr. W. van Rossum | University of Twente |
| Prof.dr. E.R. Seijdel | University of Twente |
| Prof.dr.ir. M.J.F. Wouters | University of Twente |
| Dr.M.D.T.de Jong | University of Twente |

Layout: Bernadette Pol
Printing: Printpartners Ipskamp, Enschede

Heerkens, J.M.G.
Title: Modeling importance assessment processes in non-routine decision problems
Thesis University of Twente, Enschede, The Netherlands
including summary in Dutch

ISBN: 90-365-1992-6

## Preface

Is writing a dissertation a sensible thing to do? If you are of the reckless, adventurous kind, and like starting something not knowing when – and if - it's going to end and whether anyone is going to care about the end result, a Ph.D. project is for you. On the other hand, if you are inclined to the meticulous, bothered by details that the rest of the world happily ignores, and like writing whatever you are writing at least three times, then you would not want to do without at least one Ph.D. project every four years. If you like venturing into the realm of Newton and Weber, exploring distant horizons and changing the way Men looks upon the world, you'll love writing a thesis – at least the first few weeks of it. But if you believe that progress comes from small steps forward (or sideward or backward), noticed by no one but yourself, a Ph.D. project will give you much joy – that is, the last few months of it.

I am glad I got the chance not only to do a Ph.D. project, but also to choose my own research subject. In today's highly structured research environment, few get this chance. I chose the subject of importance assessment because I wanted to explore distant horizons, but also because careful, meticulous, deliberate research in a new area requires an amount of time that I was not sure to have available in anything but a Ph.D. project. The horizon was further away than I expected, so my work in the end constitutes only a small step forward. Therefore, I had reason to enjoy both the first few weeks and the last few months of the journey. And in between? It varied. But fortunately I had a companion to share the wonders of the journey, and to fight the dragons that we occasionally met. If it were not for Bernadette and the faith she had in me, we would probably not be here today. And the layout of the thesis is her work.

There were other companions, some travelling a large part of the way with me, some appearing now and then. My supervisor, Aart van Harten, was not always sure the seeds I sawed would grow as fast as he would like, but eventually stimulated me by his conviction of a happy ending. He gave me the freedom to explore my own chosen path. Aart also gave ideas from a much more exact perspective, a valuable extension of my own thinking. The other members of the Ph. D. committee, Beate van der Heijden and Maarten van Riemsdijk, contributed their knowledge on expertise and rationality, respectively. Maarten helped me to focus my ideas in the beginning of the project. Beate played a major role in the concluding phases of my work. She strongly supported the innovative aspects she saw in my research, and had enough confidence in it to participate in writing two of the articles that form the basis of the thesis. I am glad to see Peter Lawrence, my friend since I met him at almost my first day at the faculty, in the reading commission, together with Wouter van Rossum, who was always convinced that the result would be worthwhile. Erwin Seydels positive comments on the methodology used gave me much confidence. Recently, he encouraged me to continue my research after completion of this thesis. Marc Wouters joint the reading committee with his much more quantitative orientation on experimental research. He put much effort in the final draft. Menno de Jong never got tired of talking about the chapters I kept sending him, and I learned a lot from him about think-aloud research. This is a moment to remember Freerk Lootsma, one of the two persons to whom I dedicate this work. He would have been here

today as a member of the Committee had he not been taken from us, still full of ideas and encouraging words for all who knew him.

Many students also contributed to my work. Joost Beijaard and Wilfried Krukkert performed a large part of the quantitative analysis, coding the protocols and identifying points of interest in the results. Joost, and Isabel Zilvold, conducted literature studies to fill in various aspects of what turned out to be a much broader subject than I had anticipated. Then there were the student-assistants who supervised the think-aloud sessions and trial sessions. And, of course, there were the subjects who had to perform an unfamiliar task with the burden of having to think aloud. Their motivation and ability yielded rich sources of data. My regular student-assistants in the past years, Dieuwke Boxman, Judith Nijeboer, Astrid Pijpker, Kim Brons and Nina Wessels, gave critical reflexions on the research and provided valuable administrative assistance, putting method in my madness.

I also am indebted to many people who, without any formal obligation, helped and inspired me in the past seven years. Jan Gutteling searched for literature for me, laying to rest any fears that what I wanted to do was already pursued elsewhere. Ad Pruijn made me think about quantitative follow-up research. John Boele introduced me at TNO where I had the chance to present some of my results and establish valuable contacts. Harry Janssen's originality turned out to extend to importance assessments. With Sirp de Boer, Harm-Jan Steenhuis, Aard Groen and Roel Wieringa I had many discussions about methodology, in the case of Roel resulting in articles written together. Klaasjan Visscher provided philosophical and humorous input. He made the two go together quite naturally. Frans Houweling helped for the data analysis using SPSS, which was fortunate since computers never seem to do as I tell them. With Piet Verschuren I had many enlightening discussions on methodology. Celeste Wilderom, Claudia van der Heijde, Jan Telgen, Gerard Wassink and Johan de Knijff all at some point in time gave me useful hints about literature, new angles to look at the problem of importance assessment and the nun-usefulness of statistics given my small number of subjects.

Then there were my fellow teachers on methodology of the Dutch Association for Research on Management NOBO (now NOBEM), in particular its inspiring coordinator, the late Jan Geersing. During our periodical discussions, my ideas on management research were slowly forged into the shape of the research in this thesis. The numerous people I interviewed for my secondary job as an aerospace journalist had the same influence on me. They made me aware of the enormous complexity of judging the quality of aircraft, in which weighting attributes – the subject of this thesis – plays a vital role.

Some close friends were a welcome audience for free and general talks about my research, sometimes leading to quite unexpected questions and suggestions. I especially want to mention Reinder and Ellen Houwen, Klaas Aikes and Edith de Rooy, Kris Lulofs and Ellen Wildschut, Henri and Corinne Holtkamp and Rina and Thijs Weststrate. Rina's subtle touch kept me on course on more than one occasion.

Although I like the individualistic nature of research, a pleasant vakgroep makes a good working environment. I am fortunate with the colleagues I have, and enjoy taking part in vakgroep activities like cart racing and the newly reborn rally. Leo van der Wegen and Luitzen de Boer, whose research interests come closest to mine, were always alert for useful literature and explanation about the behavioral aspects of decision-making.

I thank all of you for the help, knowledge, wisdom and inspiration during the past years. And I dedicate this book to Freerk Lootsma, but above all, to Bernadette, for her selfless love and support.

Hans Heerkens

Enschede, October 2003

# Contents

Contents

# Chapter 1: Introduction[1]

## 1 The problem

Decision-making is a vital activity within organizations. Day after day, managers have to choose between alternative courses of action. Whether these choices concern which new computer system to buy, whether to submit a bid to a request for proposals, or whether to cancel a meeting with a prospective supplier, managers constantly make choices that affect their fortunes and those of their organizations to a greater or lesser extent.

Decisions present themselves as choices between alternatives. But what drives these choices? Why does a manager choose computer system A and not B? Perhaps because system A has a bigger storage capacity or is more user-friendly. Or simply because it is cheaper. The characteristics of the computer system (storage capacity, user-friendliness, price) are called 'attributes' in decision theory, and they describe the alternatives to choose from. If we come to think about why we make a certain choice, we look at how well each alternative scores on each attribute, and most often we also develop arguments – reasons convincing to, at least, ourselves – that indicate which attributes are important to us.

This thesis is about importance assessment, the process by which attributes are identified and weights for attributes are generated. It is about the structure of the, often unconscious, way of thinking that leads people to decide what is important and what is not. It is not about what people find important in a particular situation, or why they consider certain attributes more important than others. The aim is to develop a general model that describes how people think when they have to decide what is important to them while taking a decision.

We do not aim to generalize about all possible decisions facing man; it is limited to certain types of decisions taken in organizations by laymen (non-experts), in particular those with a non-routine character.. Why this is an interesting case to study will become clear in this introduction. The exact nature of the problem to be researched is explained, and the choices made in order to arrive at the problem statement are discussed.

The problem statement of this research is:

> *What is the structure of the thinking process by which layman actors involved in non-routine decision-making processes assess the importance of attributes of the various alternatives under consideration?*

[1] In order to ensure readability, we generally use the male form of addressing throughout this thesis. Obviously, 'he' is to be read as 'he/she'.

*In particular:*

*1A: In terms of what mental activities can the importance assessment process be described?*

*1B: What characterizes laymen's performance of importance assessment processes?*

*2: What quantitative and/or qualitative relationships exist between these activities?*

*3: To what extent, and how, do actors limit the cognitive workload performing the activities of an importance assessment?*

Although not part of the problem statement, a spin-off of our research is a general research method for studying problems for which only a very rudimentary body of knowledge is available to start with, but which nonetheless yields qualitative as well as quantitative results. This method is described in Chapter 2 of this thesis.

The reasons for concretizing the problem statement in the particular sub-problems formulated above are given in Section 4 of this introduction, where the problem statement is related to the fields of decision theory, problem solving (design methodology), bounded rationality and human choice strategies. The argumentation behind focusing on the problem statement in terms of scientific and practical relevance is addressed in Sections 2 and 3.

In order to fix the position of the importance assessment process in the decision process as a whole, let's have a look at Figure 1.

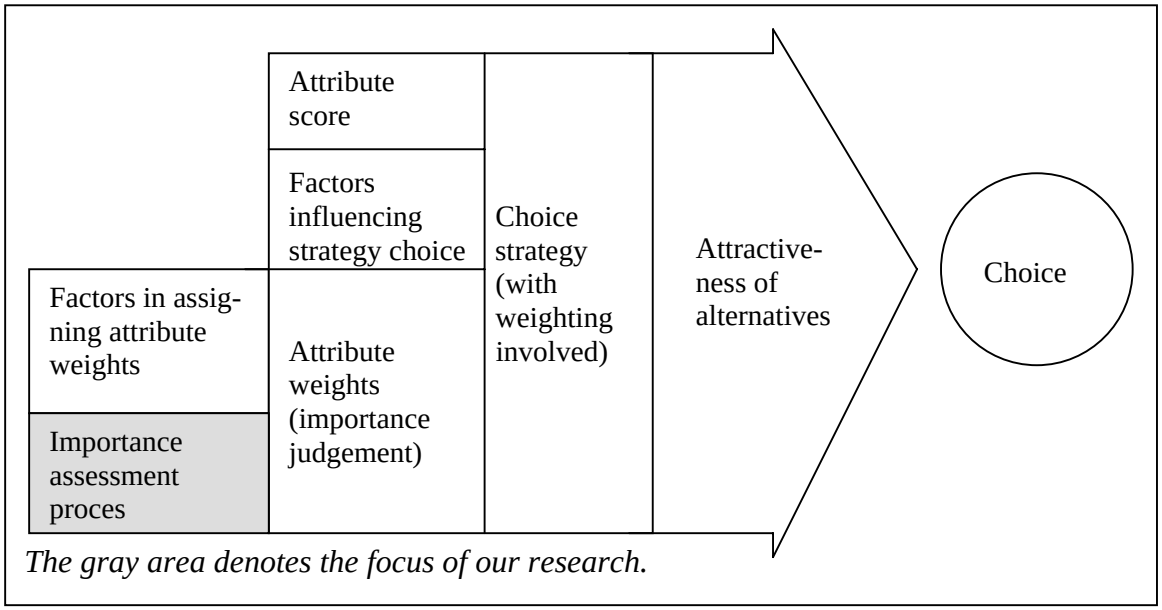



*Figure 1: The focus of our research*

Figure 1 should be read as follows: A decision is a choice from alternatives. Therefore, an actor involved in making a decision uses a choice strategy. With this strategy, scores on attributes (characteristics) of each alternative, and in some cases the weights of each attribute, are converted into a choice. The importance assessment process we study in this thesis is only relevant for those choice strategies in which weights are used. The attractiveness of alternatives is based on their attributes. The higher the score on an attribute, the more attractive an alternative becomes. An actor wanting to make a choice needs to assess the scores of all alternatives on all attributes. We do not go into this in this thesis. Attributes differ in importance. Important attributes get higher weights than attributes that are not so important. So, the actor needs to set weights for each of the attributes; he needs to make an *importance juegement*, also called a *weight judgment*. The actor may have weights readily available, for example because he has used the same weights in the past. But it may be necessary for the actor to think long and hard in order to establish what the weights should be. We call the thinking processes with the aim of establishing weight values the *weight assessment process* or the *importance assessment process*. The importance assessment process is the subject of this thesis. We are not concerned with choice strategies, with the way scores on attributes are assessed, or with the values of the weights eventually given.

Let us now discuss the various elements of the problem statement by means of Koopmans' (1980) classification of decision research. Koopman uses three elements in his classification: the distinction between descriptive versus normative (prescriptive) models, the level of analysis (individual, group or organization) and the distinction between routine and non-routine decisions. The reasons for focusing the problem statement are sometimes of a methodological nature. They are addressed here only as far as necessary to understand how the limitations of the problem statement came about. Methodological aspects are more elaborately discussed in Chapter 2 of this thesis.

Koopmans' (1980) classification was applied as follows:

**The thesis is descriptive**

The aim of this research is to describe importance assessment processes and to generate a general model of these processes, not to make normative statements about them. So, no judgment is given as to whether an importance assessment is made 'correctly', or to prescribe how these assessments should be made. The most that is done in the sense of normative models is that some suggestions are given as to which elements of the results of this research might be used in the design of future instruments aimed at supporting importance assessments. But this is more a spin-off from the research than a predetermined aim.
There are several reasons for choosing a descriptive instead of a normative approach. Firstly, no descriptive model of importance assessment processes yet exists, so it seems proper to first make a descriptive model that can then be used as a basis for normative models, for example, by observing the effectiveness of importance assessment processes under various circumstances. Secondly, it is very difficult to establish the quality of an importance assessment process. Importance judgments are up to a certain degree based

on personal values. How can these be judged right or wrong in any objective fashion? Of course, one can measure the satisfaction of an actor with the alternative chosen based on the importance assessment. But the quality of a choice may only become apparent after the alternative has been in use for some time, and then it is difficult to pinpoint a lack of quality to the importance assessment involved. Even if the decision-maker admits that, with the benefit of hindsight, the original importance judgment was wrong, this need not imply that he would have been able to make a different judgment with the information available at the time of the decision. Neither are the implications for the process underlying the wrong judgment clear. It is, of course, possible to let the decision-maker evaluate his importance assessment, perhaps in a peer group. But that still says little about the quality of the eventually chosen alternative as perceived once the alternative is implemented.

It may not be impossible to establish some quality measure for importance judgments, but the difficulties to be expected, combined with the fact that there is no descriptive model for importance assessment processes, prompted the decision to aim for a descriptive instead of a normative model. Also, the chosen research design (a think-aloud study with a limited number of subjects) does not lend itself well for designing normative models. In other words, once the optimal research design for generating an empirical model was chosen, it became practically impossible to extend the research to designing a normative model.

**The thesis is about individual actors**

The second element in Koopmans (1980) typology is the level of analysis, which defines the actors mentioned in the problem statement. It is the explicit aim of this thesis to focus purely on individual importance assessments. The actors concerned can be decision-makers, advisors or other stakeholders. So, the influence of individual actors or groups on the decision-maker is to be excluded. The reasons for choosing the individual perspective are the following:

1: Even when decisions in organizations are often taken or at least influenced by collectives of actors, actors involved in such decisions may well enter the decision-making process with their individual ideas about elements of the decision like importance judgments, ideas about possible alternative solutions and/or the attractiveness of these solutions. If we assume that these individual ideas influence the group decisions, and there are indications that this can be the case (see, for example, Hollingshead, 1996; and Wei *et al, 2000*), then it follows that a better understanding of a collective decision process may be gained by looking at the ideas with which individual actors enter the process. Sharma & Pillai (1996) argue that there is a correspondence between the nature of the decision processes in an organization (entrepreneurial, bureaucratic or planning oriented) and the nature of the decision process of actors within that organization. They argue that companies wanting to do business with an organization should adapt to the information needs of that organization and its relevant actors. Furthermore, the entrepreneurial style, where one or a few actors take the important decisions, is by no means limited to one-man

companies. Sharma and Pillai cite Ford, IBM and Microsoft as examples of organizations in which, in their early stages, the entrepreneurial decision making style was prevalent;

2: There is an extensive body of research on decision-making by individuals and on the circumstances that influence their importance judgments. One only has to think about biographies of historically important figures. This thesis forms an extension of that research;

3: Weight elicitation methods like conjoint measuring and AHP (Analytic Hierarchy Process) are primarily directed at individual actors. They can be used for consensus building in groups by giving insight in the weights chosen by individual group members, but they are primarily aimed at measuring weights at the individual level. Also, Keeney (1992, 1994) proposes an elaborate method of assessing preferences of top-level decision makers involved in strategic decisions. The implications are that individual judgments do count and that we may make use from research on individual decision-making for, for example, operationalizing certain concepts. Hence, it is both relevant and feasible to study individual importance assessment processes;

4: A methodological reason for choosing the individual level is that it seems very difficult to differentiate between group dynamics (group pressure, conflict, groupthink as described by Janis, (1971, 1972) and Janis & Mann (1977) and individual cognitive processes. By excluding the influences of interaction between actors on the way in which those actors assess importance, the complexity of the research for this thesis is reduced. Arriving at a workable model of importance assessment processes may prove a complex undertaking even without adding variables pertaining to group processes, given the lack of knowledge available about these processes. The choice for think-aloud sessions in order to study the importance assessment process under controlled conditions made a focus on the individual logical. In future research, groups can of course be addressed.

**This thesis is about non-routine decisions**

In this thesis, attention will only be paid to decisions that are to a greater or lesser extent non-routine. That is: they have a certain ‘newness’ and ‘open-endedness’ (Koopman, 1980). The road from problem to solution is to a greater or lesser extent vague. Conversely, a routine decision, according to Koopman, is not singular (so it occurs more than once). Organizations are able to develop specific procedures to handle them. Koopmans distinction runs more or less parallel to Simons (1977) programmed versus unprogrammed decisions and with Rittel & Webber’s (1974) ‘wicked problems’. In this thesis the level of ‘routineness’ of a decision is held constant, so there is no need to dwell on the operationalization of ‘routine’. The core characteristic of a non-routine decision in this thesis is that it is singular, in Koopmans terms, to the decision-maker. That is: in terms of the definition of ‘decision’ by Van der Pligt & Koele (1993), either the possible actions (or available objects) or the information about these actions or objects are unique for a certain decision. Also, the goals to be achieved may be unique.

There are three reasons for focusing on non-routine decisions, namely:

1: Non-routine decisions have a high probability of new importance judgments having to be considered, independent of whether those judgments in the end differ from judgments made during previous decisions. Obviously, if no importance judgments need to be made, there are no importance assessment processes to study;

2: There is a danger in trying to measure the way decision-makers make importance judgments during routine decisions – if those decisions require making importance judgments at all. Decision-makers might not so much make an importance judgment, but use judgments formed in previous decisions, even without knowing it. Or they might start with their earlier judgment and try to find reasons for adopting it in the decision at hand. Either way, previous importance judgments contaminate the importance assessment process;

3: The added value of insight in importance assessment processes is greatest in the case of non-routine decisions. When routine decisions are in order, the importance judgments are already there and need only be elicited. Methods exist for this, but they help actors in the importance assessment process only to a limited extent. Elicitation methods like conjoint measuring and AHP make decision-makers give importance judgments, but do not necessarily encourage them to deliberate about them. So, they are best used for routine decisions where decision-makers already have arrived at importance judgments, either implicitly or explicitly. Keeney (1992) proposes his value-focused approach for strategic decisions, which will often be non-routine, but even this is based on eliciting importance judgments that decision-makers already implicitly have, instead of addressing the importance assessment process itself. Therefore, by focusing on non-routine decision processes, this research can yield recommendations for improving the quality of importance judgments where they are the most difficult to make.

There is another focus in this thesis, mirroring the degree of 'newness' of problems. This is the level of expertise of the actors concerned. The dimensions of expertise are discussed more extensively in Chapter 3, but the essence is that an expert has, relative to a layman, a high level of knowledge, skills, metacognition, social recognition of his expertise and ability to develop his expertise further (Van der Heijden, 1998). Within the context of this research, expertise may concern the content of a decision or the importance assessment process. So, someone wanting to buy a new car and needs to set weights for attributes like safety, comfort, ease of maintenance and the like, may be an expert on cars (starting out with, for example, knowledge about how various cars score on relevant attributes) or on the way he could arrive at weights (regardless of the subject at hand).
Being an expert does not mean that a decision is routine. An actor can be an expert on cars without ever having had to buy one. Thus, an expert may know which cars are available, how well they score on certain attributes, how attributes are causally related to each other (for example, how the quality of the suspension of a car may influence both its safety and comfort), and he may have a preference for a certain type of car. But

weighting attributes may still be a non-routine task for him, if only because he has never consciously done it. For someone knowledgeable about and skilled in weighting but ignorant about cars, the weighting obviously is a non-routine problem also. Only for experts with extensive knowledge about and skills in weighting and assessing cars could the weighting of car attributes resemble a routine problem. But even this does not need to be the case, for even if an expert has all content and weighting knowledge available, he need not have processed this knowledge and skills into a method readily available for dealing with a particular weighting problem. On the other hand, it is difficult to see how a problem can be routine for a layman in the field.

In sum, the terms ‘routine’ and ‘non-routine’ apply to decisions and the terms ‘layman’ and ‘expert’ apply to actors involved in making decisions. A non-routine decision is a new decision for which no standard methods or procedures are available to the actor concerned. A routine problem is one that has been has been dealt with by the actor before and certain methods and procedures are available to handle it. An expert is an actor who has knowledge and skills available pertaining to (elements of) a decision, like content knowledge about the decision context or about weighting of attributes in general, or both. A layman lacks this knowledge and skills. For a layman in the field concerned, a problem is always non-routine. Examples are given in Figure 2.

| | **Routine problem** | **Non-routine problem** |
|---|---|---|
| **Layman** | No realistic example available | Newly appointed Director of Operations of a transport company, formerly Director of Administration, preparing the formulation of requirements for a fleet of minibuses |
| **Expert** | Purchasing official of a fashion house with 15 years of experience setting the requirements for the new line of business suits every three months | Director of a transport company setting requirements for a new fleet of minibuses, for the first time participating in a fuel cell experiment |

*Figure 2: Examples of (non-)routine problems handled by laymen and experts.*

It may not always be easy to separate the characteristics of the actor from those of the problem. For example, a layman may become an expert in a certain field by solving many problems in that field. But this likely means that these problems become routine problems, so the characteristics of both the actor and the problem change. In this thesis we will not concern ourselves with these issues. We confine ourselves to problems that are non-routine for the actors that are confronted with them (for example because they occur infrequently and under continually changing circumstances). So, even if the actors we study could develop expertise in dealing with these problems, the problems would never become routine. We confine ourselves to the right upper cell of Figure 2: layman confronted with non-routine problems.

In reality, most actors making importance assessments will be to a greater or lesser extent experts and not laymen in the field concerned. They usually will have made similar importance assessments earlier in their careers. Therefore, if we focus on non-routine decisions we should also devote attention to the role of expertise, in order to relate our research on laymen to the real world in which importance assessments are made. So, we will not just develop a model of the importance assessment process in non-routine decisions, but we also explore the role of the expertise of the (layman) actor concerned (Chapter 3 of this thesis).

Koopmans (1980) classification aims in focusing our research. But there are some limitations unrelated to Koopmans' classification that should be mentioned. They will be further explored in Section 4.1.2 of this introduction and in Section 3 of Chapter 2. The limitations are:

1: This research is confined to situations in which weights are established as part of the decision process. Some decision strategies do not require weights, but they are of no concern to us;

2: Only decisions in which weights are determined *before* alternatives are evaluated are studied;

3: We do not look at purely intuitive or emotional decisions, although the role of intuition is recognized, as is the role of emotions.

The structure of the remainder of this chapter is as follows. In section 2, we introduce a fictional real-life problem that might occur to a manager of a private company. From this practical problem, we illustrate and clarify, in broad terms, the research problem addressed in this thesis. In section 3, the scientific and practical relevance of our research problem is addressed in general terms. In section 4, the problem is further specified and the boundaries of the problem are explored. Section 5, in which the structure of the thesis is explained, concludes this chapter.

## 2 Importance assessments: an illustration

George Jones owns a firm that transports passengers to and from an airport some 150 km away. His clients are, for various reasons, unable or unwilling to travel by car or by public transport. Some years ago, George bought a fleet of Volkswagen minibuses that needs to be replaced in the coming year. The minibuses available on the market differ from each other in many respects, such as ease of maintenance, running cost, safety features, passenger capacity, passenger comfort and the like. As stated before, we call these characteristics of the minibuses *attributes*. As George wants to standardize on one type of minibus, he has to decide which type to acquire. In order to do this, he needs to decide how important the various attributes of the minibuses are to him. Is ease of maintenance more important than passenger comfort? If so, and if George has to choose between minibus A that is somewhat easier to maintain and bus B that is a bit more

comfortable, he'll probably choose bus A. But what if bus A is somewhat easier to maintain, but bus B is a lot more comfortable? Then George has to decide how much more important ease of maintenance is than comfort.
The task facing George is a difficult one, for one simple reason. The attributes whose importance has to be assessed cannot be directly compared to each other. George probably has good arguments why he thinks certain attributes are important, very important, rather important, not important and so on. He may, following his business instincts, decide that preserving his present position in the market for business travelers may be difficult to sustain due to the advent of high-speed trains, so comfort is of the utmost importance. Or he could decide that if the supplier of the minibuses can take care of all second-line maintenance, the ease of maintenance need not be a factor in his choice. If one of George's minibuses gets an accident and one or more passengers die, George faces the specter of a tarnished image and possibly huge damage claims. This makes safety important.

But these arguments explain why certain attributes may be important, not why they are more important than other attributes.

It would be convenient if the attributes under consideration could be expressed in a common denominator. Money immediately comes to mind. For example, if George can calculate the impact of one unit of extra safety on total sales – which he very likely can't - and he can do the same with one extra unit of comfort, then he could simply calculate which of the alternative minibuses would maximize sales or profit. But in practice, neither safety nor comfort can be completely expressed in money. An obvious reason is that preferences of potential clients accident are not exactly known. But also, there may be other, for example moral, considerations. Would George willingly lower safety standards if it were to save him money? Would his professional pride not urge him to emphasize client satisfaction even if it does not always bring direct monetary rewards? George may have other goals than straightforward profit, like long-term client loyalty and satisfying his own ambitions as an entrepreneur.

Furthermore, George is not just spending his own money. He may have to motivate his choice to employees of the bank where he applies for a loan. His shareholders or his employees may also be interested in his motives. George is not buying a car or a television set for his own use, in which case nobody else cares which attributes he considers important or not. His decision will have to be motivated – to people who have the same difficulty of assessing the importance of attributes.

The challenge that George faces is all the greater since the decision he has to make is far from routine. The decision can be called strategic and will shape the future of the company for years to come. The acquisition of a new fleet of minibuses is likely to occur only every five years at most. So, George never gets proficient in it. And in each case, technological and market advances may call for a reappraisal of the decision procedure of which assessing the importance of attributes is a part. So the chance that there is a standardized procedure that guarantees an optimal decision or even makes it possible to reliably assess the importance of attributes is negligible.

All in all, George faces the task of assessing the importance of attributes that influence the choices that he has to make. It is clear that this is a difficult task and although there are methods for eliciting the importance that George attaches to an attribute once he has made up his mind, it is by no means clear what goes through George's mind when he is trying to figure out how important the various attributes are to him. This thesis is about the thinking process of George and others faced with the same predicament.

In the remainder of this chapter, the motivation for our research is described in a more general setting than just George's. It is explained how our research fills gaps in knowledge in the literature and how the problem is approached in this thesis. But first we should get a definition issue out of the way.

In this study, the importance an actor attaches to an attribute is called the *importance judgment*. Examples of importance judgments are: 'the safety of a minibus is more important than its comfort' and 'fuel consumption gets a weight of .3 on a scale from 0 to 1'. The thinking of an actor about how important an attribute is to him (and that precedes the importance judgment) is called *importance assessment*. Hence importance assessment refers to the cognitive process. The result of the importance assessment process is the importance judgment. This study is not concerned with importance judgments but only with importance assessments. So, for example, the question how important George thinks safety is relative to comfort is of no interest in this study. Even the arguments for his importance judgment are not relevant. The focus of our attention is the structure of George's thinking (the cognitive activities) during the importance assessment process.

## 3 Why do we want to know how importance assessments work?

The main motive for undertaking this study is that hardly any scientific research has as yet been done on the cognitive processes involved in importance assessment. This will become clear in section 4. But apart from the scientific challenge of exploring uncharted territory, there are several other reasons for research on importance assessment processes. The first reason is a very practical one, implied in the case of the minibus company. Knowing how importance assessment processes are conducted may yield ideas how to improve them or predict their outcome better, either objectively or in the minds of decision-makers. If a decision-maker feels that the importance assessment has been conducted in a proper way, he or she may have more confidence in the resulting importance judgment. Also, when actors have more insight in their own importance assessment processes, this may help them to communicate their motives to others and to understand what their fellow-actors are trying to say. Goldstein & Mitzel (1992) see statements about importance as meta-cognitive judgments as 'means for communicating one's decision-making strategy to others' (p 383). It is entirely possible that insight in one's own reasoning when performing importance assessments may serve the same functions. Note that this does not necessarily mean that it will be easier to reach a decision for a group of decision-makers who have to choose an alternative together. If decision-makers have high confidence in their importance judgment they may be less inclined to compromise. As Keeney & Raiffa (1976) note: 'In some circumstances, the

more confusion that abounds, the easier it is to establish a compromise' (p10). But they add: 'Still, we think that familiarization can sometimes facilitate reconciliation'. This optimistic view is echoed in this thesis.

Actors within an organization may also be interested in the weight assessment process in order to influence it, or to be able to determine their negotiation strategy during collective decisions. Actors and organizations may want to optimize the decision process environment, for example: by ensuring that actors have the optimal quantity and quality of resources such as information. Stakeholders outside the organization may want to have insight in the behavior of the organization (i.e. its actors) as accurately as possible, for example to assess the soundness of their investments.

The second reason for doing research on importance assessment is linked closely to a problem often found in practical life: the problem of integrating information pertaining to different disciplines. This problem is well known in management. It also should be a central issue in research on management (Heerkens, 1999). According to many authors, a key characteristic of research on management is that it is multi- or even interdisciplinary (Easterby-Smith *et al*, 1993, Van Riemsdijk, 1999). Research on management leans heavily on knowledge from economics, sociology, marketing and the like. But it is, however, far from clear how these disciplines are integrated, both in the body of knowledge on management research and in the heads of managers (Heerkens, 1999). Integrating the various pieces of information needed to take a decision implies, amongst other things, importance judgments and hence importance assessments. Knowledge about importance assessments may help to develop instruments for managers to integrate information needed for decisions. Developing these instruments would be an excellent goal for management scientists and would do eminent justice to the multi- and interdisciplinary pretences of the field of research on management (Heerkens, 1999). It would also enhance the identity of research on management and its unique contribution to science and society. Furthermore, it would help the practitioners of management research to further develop a body of basic knowledge as opposed to the large body of knowledge available at present that gives research on management, in the opinion of many authors, its decidedly applied character.

Another reason for the relevance of studying importance assessment becomes clear when looking at the predicament of the manager of the minibus company described in Section 2. Seen from the perspective of the modeling of problem-solving activities (see for an overview Hicks, 1995), the manager has a problem and, having identified several potential solutions, now has to choose between them. This is an example of phase 3 of Simons (1960) problem-solving model. But before the choice can be made, the importance of the various pros and cons has to be assessed. This thesis aims to shed light on this activity of problem solving. On the basis of its results, both normative and empirical problem-solving models may be further developed, especially the activity of comparing alternatives, which is not well described in problem-solving models at present. Given the proliferation of decision-process and problem-solving models, both empirical and normative, in which judgment of alternatives has a place (see for an overview Hicks, 1995), it is logical that there should be interest from researchers in fields like psychology

and decision theory in the area of importance assessment in managerial decisions.
The last reason for this research is that it is meant as a basis for future comparative research, in areas such as the role of expertise in managerial decisions. Once a generic importance assessment process model is established, research can be done on how importance assessment by experts differs from that by laymen. Groundwork for this is laid in this thesis. This can be a valuable addition to research mentioned earlier. Boerlijst & Van der Heijden (1998), Boerlijst, Munnichs & van der Heijden (1998) and Van der Heijden (2000) have done extensive research on the development of expertise by employees and career development. The concept of importance assessment processes may help to explain further the changes observed with employees as they get older or more experienced. For example: more experienced employees may devote more or less effort to certain phases of the importance assessment process than younger or less experienced colleagues. Or they may be more (or less) consistent in the way they perform importance assessments with various types of problems. The role of information processing may also differ: older people may be less receptive to new information and hence may emphasize those phases of importance assessment that do not require the acquisition or processing of new information.

It was stated that little research has been done on importance assessment processes. This issue is addressed in the next section.

# 4 Theoretical context

In section 4.1, a brief overview is given about the presently available knowledge of the assessment of importance of attributes. This leads to the description, in section 4.2, of the knowledge that is lacking. In 4.3, the limitations following from the positioning of our research in the available body of knowledge are described.

## 4.1 Previous research on importance assessment processes

As this study is about importance assessment, it is logical to look first at the concept of ‘importance’. Then, we briefly review the existing body of knowledge about importance assessment processes.

### 4.1.1 What is ‘importance’?

The reader is assumed to be familiar at a basic level with decision theory. Therefore, in this introduction, we will only cover one aspect of decision theory, namely importance. We will hardly touch upon, for example, utility functions in which importance judgments can be inserted to compute the attractiveness of alternatives. Importance judgments form one of the inputs of utility functions on the basis of which the attractiveness of alternatives can be calculated. But we are not interested in the use of importance judgments, only in their generation.

When the manager introduced in Section 2 buys a minibus, what does he mean when he says that safety is more important than comfort? When he says: “I bought the Opel minibus. It may not be as comfortable as the Volkswagen, but it seems safer to me”, does he mean that he thinks safety is more important than comfort? Or is the difference in safety between the Opel and the Volkswagen as perceived by the manager outweighed by the difference in comfort? Perhaps the manager considers comfort more important in the sense that he devotes more time assessing the comfort of the candidate minibuses than their safety, while in the end deciding on the basis of safety. Often, ‘importance’ is defined loosely, if at all. Jaccard, Brinberg & Ackerman (1986, for example, state that ‘an attribute is said to be important if a change in the individual’s perception of that product attribute leads to a change in the attitude towards that product’. This is a very general definition; what does ‘attitude’ mean in terms of the decision process and can attributes have different levels of importance? Alpert (1971) notes that salience (the inclination of actors to name certain attributes when queried about an object), importance (attributes that are essential for an object but are shared by all alternatives and hence do not determine choice) and determinance (influence on choice) are often confused. Goldstein & Mitzel (1992) argue that there is no consensus about the meaning of ‘importance’, while suggesting that actors might perceive a distinction between ‘importance’ related to the outcome of a decision process (the influence of attributes on the choice) versus ‘importance’ related to the decision process itself (the attention that should be paid to evaluating certain attributes). Wilkie & Pessemer (1973) argue that ‘importance’ can be ‘an ambiguous term which might reflect either prominence or value’. Other authors give rather precise definitions. For example, Von Nitzsch & Weber (1993) state: ‘They (the scaling constants indicating importance) represent the tradeoffs between units of different conditional value functions’. In other words: the relative importance of two attributes A and B is the amount of attribute A that an actor is prepared to give up in order to obtain one unit of attribute B. In the many experiments in which subjects have to judge the importance of attributes, some of which are addressed later in this chapter, always the same implicit definition is used: importance is the relative influence of the attribute concerned on the attractiveness (in the eyes of the decision-maker) of each of the alternatives to be chosen from (see, for example, Fischer, 1995). This expression of importance is called ‘weight’ in decision theory, and there are reasons to believe that this is the meaning of ‘importance’ that subjects in experiments usually work with (Goldstein & Mitzel, 1992). So, ‘weight’ has, in this thesis, the same meaning as ‘importance’.

What does an actor do, once the weights of the relevant attributes are known, to determine which alternative to choose? Intuitively, the more important an attribute is and the higher the score of an alternative on that attribute, the more attractive that alternative is. The safer a certain minibus is, the more attractive it is to the manager of the minibus company described in Section 2, and the more important safety is to him, the faster the attractiveness of a minibus increases with each ‘unit’ of extra safety. In utility theory, this is formalized in the following way: (Keeney & Raiffa, 1976):

$$U_i = \sum_{j=1}^{N} A_{ij} W_j$$

In words: the score A on each attribute of an alternative is multiplied by the weight W of that attribute (the weight is the same across all alternatives) and the results, called 'partial utilities' are added to get the total utility, or attractiveness, of an alternative. The alternative with the highest attractiveness should logically be chosen.

So, in this thesis, we see importance of an attribute as its relative contribution to the total utility of alternatives, and hence its relative effect on the choice of an alternative. But this definition is only used to set up the theoretical framework for our research. The subjects we studied in the course of our research were allowed to use their own definition of 'importance'.

But there is more to importance assessments than the formula shows. It is possible that some or all attributes are split (decomposed) in sub-attributes. For example, the attribute 'comfort' of a minibus may be split in 'dimensions of the cabin', legroom', 'quality of the seats', 'noise level' and so on.
This implies that importance assessments may be conducted in four ways:

1: Giving weights to attributes without considering sub-attributes;

2: Giving weights to sub-attributes first and then aggregate these weights to attribute weights;

3: Weighting sub-attributes without giving weights to attributes;

4: Weighting attributes first and then deriving weights for the sub-attributes.

In our research we study the weighting process of attributes. This means that methods 3 and 4 should not be relevant, although the subjects we study may use them nonetheless if they think they are useful. Starting with method 2 but then, for whatever reason, not converting sub-attribute weights into attribute weights effectively means that method 3 is applied.

Method 1 is straightforward, given the purpose of our research, but method 2 is not. After all, which method should be used for converting sub-attribute weights into attribute weights? Is a method used at all? One should keep in mind that weights have only a relative meaning. Hence, the bottom line of aggregating sub-attribute weights is that we can assume that the sum of the weights is 1 and if not, then a correction factor should be applied reflecting the weight of the attribute (aggregated sub-attributes) relative to the weights of the other attributes to be weighted. Borscherding, Schmeer & Weber (1995) show that the weight of a characteristic of an alternative can depend on whether it is treated as an attribute or as a sub-attribute. Weber, Eisenfúhr & Von Winterfelt (1988) showed that aggregated weights of sub-attributes do not always add up to the weights elicited for the attributes; the so-called splitting bias. Although these results are somewhat contested (Póyhónen & Hámáláinen, 1998), there is ample reason to study in our research how actors handle the relationships between sub-attributes and attributes.

One final remark about the definition of 'importance' should be made. It is certainly possible to provide a more exact definition than has been done in this section. We chose not to, for two reasons. Firstly, we do not know how actors involved in decision processes actually think of importance, so we wanted the subjects in our research to be free to give their own meaning to 'importance'. Our definition is very general and is likely to give subjects as much freedom as possible. Goldstein (1990) has investigated what concepts of 'weight' actors can have. But this was done in a specific, non-organizational context and with experiments that might have influenced the concept of 'importance' his subjects used, so it seems appropriate not to commit to too specific a definition of 'importance' before more insight in the weight assessment process is obtained. Of course, one could argue that taking utility theory as a frame of reference for this thesis might also unduly limit the scope of our definition of 'importance', but it does so at a very general level.

A second reason, albeit a minor one, of not defining 'importance' more specifically, is that many available definitions are linked to specific ways of measuring weights. For example, the above definition of Von Nitzsch & Weber is closely associated with conjoint or 'structural' measuring methods (see, for example, Danaher, 1997; Harte & Koele, 1995; Jaccard, Brinberg & Ackerman, 1986). This does not mean that the definition's usefulness is limited to these measuring methods, but the association is there, and we want to avoid it.

Now we have a general idea of the concept of 'importance'. The next logical step is to look at previous research done in the area of importance assessment and to identify what knowledge our research adds.

### 4.1.2 Perspectives on importance assessment processes

In this section, links are established between our problem statement and four related fields of research: decision theory, problem-solving and design methodology, bounded rationality and human choice strategies. A fifth area: research focused on specific decision contexts, is also addressed to complete the picture of decision research and to accentuate the relevance of our own study.

**Decision theory**

An important area of research in decision theory linked to importance assessment is the study of factors that influence the assessment process. Notable examples are: perception of risk and attitudes towards risk (Kahnemann & Tversky's (2000) Prospect Theory), the perspective of the decision-maker (Kray & Gonzalez, 1999; Kray, 2000), information presentation and usage (Guo, 2001; Russo, Medvec & Meloy, 1996), the concept of attribute weights (Keeney & Raiffa, 1976), factors that cause biases in weighting (Borcherding, Schmeer & Weber, 1995; Fischer, 1995; Póyhónen & Hámáláinen, 1998), the influence of unimportant or irrelevant attributes on choice (Goldstein & Busemeyer, 1992; Hsee, 1995), group decision-making, particularly the relationship between

individual and collective preferences (for example Hollingshead, 1996; Wei *et al*, 2000), the internal and predictive validity of various methods for measuring importance judgments (Harte & Koele, 1995; Jaccard, Brinberg & Ackerman, 1986), and the influence of regret aversion (Zeelenberg *et al*, 1996). Marketing research has focused on socioeconomic and demographic factors influencing the perceived importance of attributes, and Meehl (1954) studied the degree to which limited numbers of relevant illness symptoms occurring with patients influence doctors' diagnosis. Further contributions regarding the explanatory power of regression models have been realized by Dawes (1979) and others. Keeney's (1992, 1994) Value-Focused Thinking approach shows how various instruments assessing value preferences of decision-makers can be used to optimize non-routine decisions within an organizational context.

The above-mentioned research provides us with some building blocks for modeling importance assessment processes, like the concept of 'importance', and the relationship between attribute scores, weights and attractiveness of alternatives. It is likely that these concepts will figure in one way or another in a model of the importance assessment process (sub-problem 1A of our problem statement) and the way this model manifests itself in practical situations in terms of the role of expertise (sub-problem 1B of the problem statement). But this process itself is not studied in the research discussed above. The above-mentioned research either takes the weights actors assign for granted and concentrates on the decisions made on the basis of these weights (preferences), is concerned with eliciting the weights with sufficient validity (necessary for linking preferences with choices) or looks at the factors influencing the weights, like perceptions of risk. The importance assessment process, i.e. the way actors think while weighting, is addressed merely incidentally. To the best of our knowledge, no attempt has been made to model the importance assessment process as a whole. So, our research covers the process of generating weights, while the above-mentioned research concerns the weights when they are, or have been, assigned. This is shown in Figure 3.

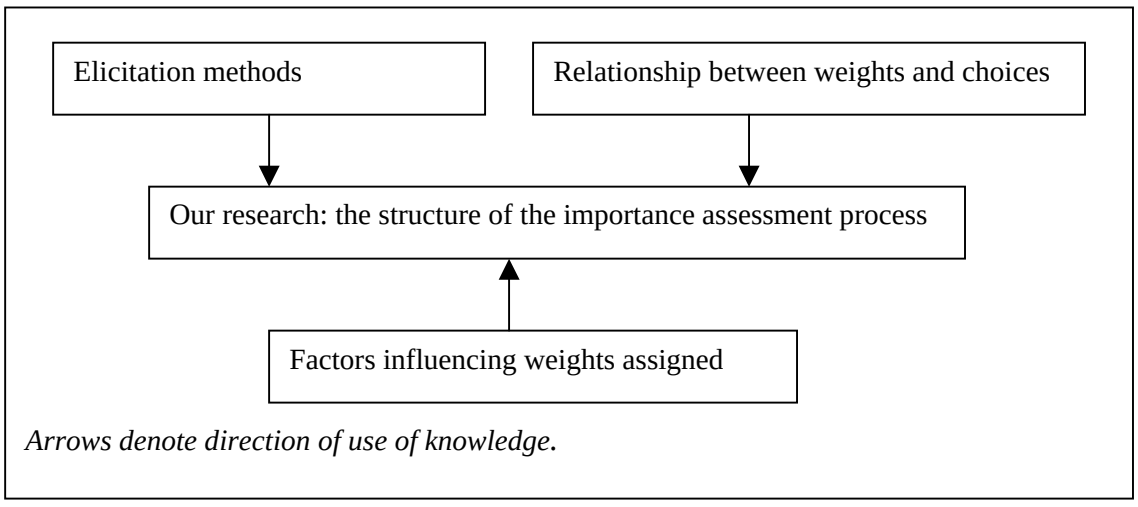



*Figure 3: The place of our research in decision theory.*

**Problem solving and design methodology**

Much research has been done on logical problem solving (see for an overview Ericsson & Simon, 1993), general mental operations involved in problem-solving (Newell & Simon, 1972), problem analysis (see, for example, Checkland & Scholes, 1990), problem solving by designers (see, for example, Hamel, 1990), the extent to which phased problem-solving models are followed in practice (Lipschitz & Bar-Ilan, 1996), the influence of arousal and its various accompanying emotions on task performance (Kaufmann, 1999) and difference in task performance between laymen and experts, (Chi, Glaser & Farr, 1988; Ericsson & Smith, 1991; Van der Heijden, 1998; Van der Heijden, 2000). So, research on problem-solving and design methodology is aimed at studying problem-solving and design processes in general, or specific processes like problem-solving by experts. Mental actions that actors perform during weight assessment processes have, to our knowledge, not yet been studied.

In many problem-solving models, we find a phase of generating alternative solutions and judging them. This phase often involves some sort of weighting, either crude or sophisticated. It would seem logical that the importance assessment would have been studied in this context. Regrettably, descriptive and prescriptive research in this area is aimed at eliciting weights, and at providing formats for presenting weights and alternatives in an efficient way. Also, much attention is paid to methods for the generation of alternatives, like brainstorming and morphological methods. The structure of mental activities of importance processes is not covered.

Research on problem-solving and design methodology provides phase models that could be taken as a basis for a model of the importance assessment process. Many models distinguish, for example, some sort of problem structuring phase and a solution phase, or a problem analysis phase, a solution generation phase and a choice phase (see for an overview Hicks, 1995). These models can be of help in placing the activities identified in importance assessment processes on the basis of decision theory in a sequential order. They may also give insight in how these activities are related, for example in the sense that some activities are building-blocks for subsequent activities (sub-problem 2 of our problem statement).

The place of our research in the field of problem-solving and design methodology is shown in Figure 4.

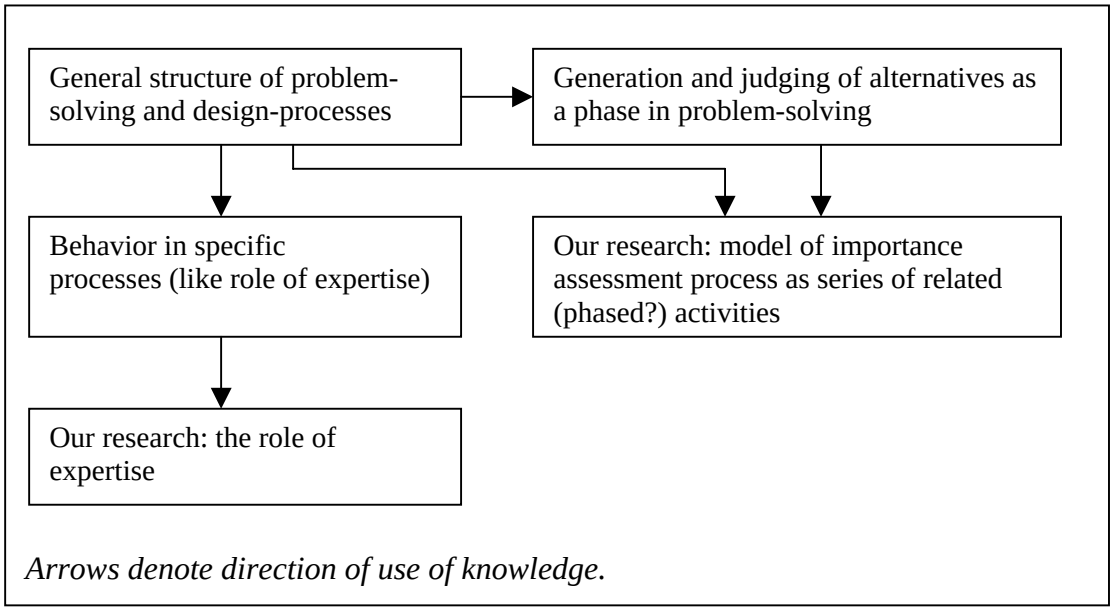



*Figure 4: The place of our research in theory on problem-solving and design methodology.*

**Bounded rationality**

This area is closely related to the previous one, but we treat it separately for it does not concern structuring the problem-solving or design process as a whole, but handling the various activities within a problem. Importance assessment problems, like many other problems, are inherently complex, especially when they are non-routine. The various alternatives open to a decision-maker, like the types of minibuses on the market in the case outlined in Section 2, may have many attributes, and each attribute may have an infinite number of possible weight values. Taken all these attributes and weight values may well be beyond the intellectual ability of many decision-makers, and they have to reduce the complexity of the importance assessment process by taking only a subset of the available courses of action into consideration. This way of complexity reduction is called bounded rationality (see, for example, Simon, 1976). Bounded rationality is more of a general concept, an idea, than a theory from which specific hypotheses can be derived (Aumann, 1997). Bounded rationality has been modeled at a general level, pertaining to decision-making in general instead of to specific activities relevant for importance assessments (see, for example, Munier *et al*, 1999; Rubinstein, 1998). Much research has been done on the rationality of human behavior and especially on reasons for lack of rationality (see, for example, Simon, 1976, Arthur, 1994, Kahneman, 1994, Loewenstein, 1996, Simonson, 1990, Shafir, Simonson & Tversky, 1993, Levine, Halberstadt & Goldstone, 1996, Kaufmann, 1999, Shafir & LeBoeuf, (2002); Schwartz, 2000, Toda, 1980 and Damasio, 1995). However, much of this research is not linked to complexity reduction (bounded rationality) but to irrationality and its explanations (emotions, actors' lack of insight in their own preferences and so on). While taking for granted that actors are often boundedly rational, little is known about the way in which they are boundedly rational. What does 'not taking all possible courses of action into

consideration' actually mean in the context of specific types of problems? Given that importance assessment processes have hardly been studied, it should come as no surprise that this question cannot be answered (sub-problem 3 of our problem statement). Still, the answer is important for understanding how importance assessment processes work in practice since, as stated before, importance assessment problems are so complex that actors have no choice but to reduce complexity. The way they handle the activities identified under sub-problems 1 and 2 of our problem statement may be influenced by this complexity reduction.

Figure 5 shows the relation of our research with research on bounded rationality.

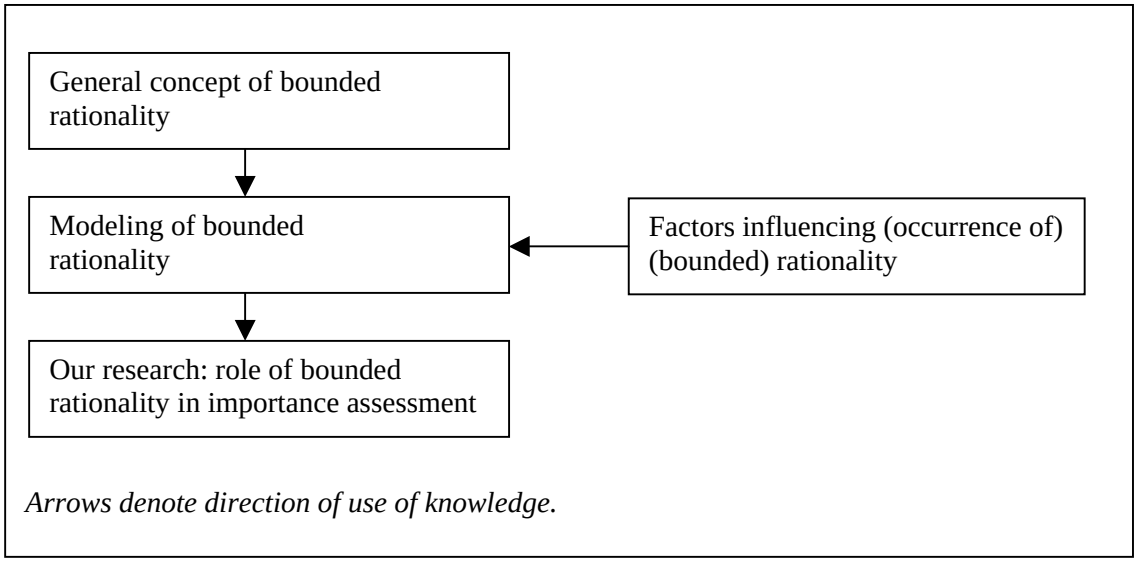



*Figure 5: The place of our research in theory on bounded rationality.*

**Human choice strategies**

This perspective is largely an extension of the perspectives discussed above. The role of choice strategies is also shown in Figure 1. With the building blocks of decision theory, problem solving and (bounded) rationality, decision strategies used in practice by, for example, consumers are studied. These strategies may consist of several techniques or types of behavior also found in decision theory. Decision-making means choosing, so it it appropriate that, in this thesis, we address possible choice strategies, in some of which weights play an important role. In this subsection we will address:

- the main types of decision strategies available to actors;
- the main factors that determine which strategy is likely to be chosen in a particular instant;
- the main factors that influence the values of the weight used in choice strategies.

On the basis of these three points, the boundaries of our research will be further explored.

*Decision strategies*

The most important choice strategies are the following (Bettman, Luce & Payne, 1998; Van den Brink, 1993):

1: The weighted adding strategy. An actor assesses the scores on all relevant attributes of the alternatives and multiplies these scores with their respective weights. This is in line with the normative utility model often referred to in decision theory;

2: The lexicographic strategy: The alternative with the highest score on the most important attribute is chosen. In case of a tie, the alternative with the highest score on the second most important attribute is selected, and so on;

3: The satisficing strategy. Alternatives are evaluated sequentially, and the first alternative that attains the desired level of utility (attractiveness) is chosen. Further alternatives are not considered;

4: Elimination-By-Aspects (EBA). Alternatives are eliminated in stages. First, all alternatives that do not attain a minimum level of utility on the most important attribute are eliminated. Then, all remaining alternatives that fail to attain a minimum level of utility on the second most attribute are left out. And so on, until only one alternative is left;

5: The equal weights strategy. Every attribute receives an identical weight, effectively eliminating all weighting that has to be done in the case of the weighted adding strategy;

6: The majority of confirming dimensions strategy. Alternatives are compared in pairs, attribute by attribute. The alternative with the highest number of attributes with the highest score wins. This is done until one alternative is left;

7: Counting the numbers of good and/or bad features of each alternative. Several algorithms for determining the winning alternative are possible: the alternative with the highest number of good features, the lowest number of bad features, the best ratio of good and bad features (features can also be neutral) and so on.

Each of these strategies may comprise information search strategies. Information may be sought about, for example, the scores of alternatives on certain attributes. Information search strategies that can be seen as generating input for the choice strategy (see, for example, Guo, 2001; Russo, Medvec & Meloy, 1996; Saad & Russo, 1996) are not extensively discussed. They fall outside the scope of our research.

Combinations of strategies are possible. An example is the Image Theory (Beach, 1993, Beach, Puto, Heckler, Nylar & Marble, 1996). According to this theory, actors may first eliminate alternatives that do not attain a minimum utility, in terms of confirmation to an

image of what an acceptable alternative should look like. Then, the remaining alternatives are judged by the weighted adding method. Not only is this method a combination of strategies 1 and 4, it is also an example of a combination of a noncompensatory strategy (4) and a compensatory strategy (1). With a noncompensatory strategy, a high score on one attribute does not compensate for a low score on another attribute. With strategy 4, an alternative that does not attain a minimum score on a predetermined attribute is eliminated, even with excellent scores on all other attributes. Strategy 1, on the other hand, is compensatory. An alternative with a low score on attribute 1 and a high score on attribute 2 can be just as attractive as an alternative with moderate scores on both attributes.
It could be that parallels to the above-mentioned strategies can be found in importance assessment processes. Although we do not focus our research on these parallels, we will identify them if appropriate.

It is clear that not all strategies require weighting. The satisficing strategy, for example, does not. This research looks at situations where weighting does take place. This assumes that a choice strategy is used for which weighting is required. But it is, of course, possible that, after attributes have been weighted, for whatever reason a strategy is chosen that requires no weighting. Then the weight judgments are not used after all. For our research it makes no difference whether the weights assigned are actually used in a choice strategy.

*Factors determining the choice strategy employed*

If actors have several possible strategies at their disposal, a logical question is which factors determine the choice of a strategy. Bettman, Luce & Payne (1998) assume that actors have a repertoire of strategies and choose the one that best fits the decision under consideration. This assumes a level of rationality. Based on an extensive literature study, they propose the following factors, (although firm empirical evidence sometimes has yet to be obtained).

1: Relative importance of the various possible goals of the strategy: maximizing accuracy and potential for justification of the choice, and minimizing cognitive effort and the experience of negative emotion. Note that these are generic ‘procedural’ goals, not the reasons why a choice needs to be made in the first place. Bettman *et al*. propose that the importance of the various goals is influenced by characteristics of the problem, including the importance and irreversibility of the choice and the timeliness and ambiguity of the feedback available on the performance on each goal. For example, irreversible decisions may make accuracy relatively important. If an actor attaches much value to accuracy, the weighted adding strategy is the most suitable. If effort is to be minimized, perhaps because of a time constraint, EBA or even a random choice becomes more likely. Fennema & Kleinmuntz (1995) observe that actors often do not estimate the required effort and obtained accuracy correctly before starting a decision process. This need not undermine the relationship between goals of effort and accuracy on the choice strategy used, but cases doubt on explanations for this relationship from the perspective of rationality.

2: Individual characteristics, like experience and training. For example, the more expertise a actor has with regard to the alternatives under consideration, the easier it may be to assign weights to attributes, opening the possibility for strategies in which weights are used.

3: Emotion. When decisions induce emotions (like when personal taste or well-being is involved), actors are, according to Bettman, Luce & Payne (1998) more likely to value accuracy (in order to avoid negative emotions such as post-decision regret). They are also likely to use strategies that avoid direct trade-offs between attributes (which inevitably means giving something up). Examples are strategies where scores on attributes of different alternatives are compared (instead of scores on different attributes for the same alternative), like EBA.

4: Environmental factors, like the amount of information available. For example, the weighted adding strategy requires detailed information on all attributes, whereas the satisficing strategy can be executed with much more global information. If actors select information that does not reflect their underlying values, the use of this information may decrease accuracy. Also, an externally imposed need to justify one's choice to others can be seen as an external factor (as opposed to the need for juastification an actor may feel himself). Actors may cope by taking the (perceived) preferences of those to whom they are accountable into consideration. They may also place increased emphasis on accuracy, or they may opt for alternatives where the advantages are relatively easy to spot (self-explaining), even if they need not the best alternatives possible. So, a lexicographic strategy may be in order. Time pressure may lead to more use of attribute-based strategies and more selectivity in the use of information. Hagafors & Brehmer (1983) state that the need for justification may lead to an analytical mode in judgment behavior. In terms of the strategies mentioned above, this would seem to be a weighted adding strategy. De Hoog & Van der Witzenboer (1986), however, do not observe this relationship. Interestingly, they also conclude that the need for justification does not influence the number of attributes taken into consideration. Note that the (ezxternally imposed) need for justification may alter the relative importance of goals addressed in point 1.

5: The complexity of the problem. As problems get more complex in the sense that the number of alternatives to be chosen form increases, noncompensatory and attribute-based strategies such as EBA appear relatively attractive. Increases in the number of attributes generally do not lead to strategy changes, Bettman *et al*. propose. Negatively correlated attributes (if one attribute scores high, the other scores low) would lead to alternative-based strategies (like weighted adding) in less emotional tasks. If alternatives are noncomparable (are described by different sets of attributes), actors may use more abstract attributes, like 'enjoyment' that are applicable to all alternatives. If alternatives have to be compared judging) alternative-based strategies are used relatively more often than when one alternative has to be chosen over all others (choice). In the latter case, attribute-based strategies are relatively popular.

In our research the above factors are not varied, so their influence on the importance assessment process is not established. The setting of our research can be characterized as follows: we do not control the importance of goals to the subjects (factor 1), but they have to be able to explain their work to others. We confine ourselves to laymen (factor 2), and do not focus on emotional decisions, although some attributes may have an emotional component to the actors concerned (factor 3). We study importance assessment processes within an organizational context (factor 4) for decisions which are non-routine, so: complex (factor 5).

*Factors influencing weight values*

It is clear that the key elements of strategies outlined above are (perceptions of) attribute scores and (in some cases) weights as input variables and the resulting choices as output variables. We do not concern ourselves with attributre scores in this research. So, now that we know which are the main choice strategies available, it is time to look at possible factors that influence the values of the weights used by the actors (the weight values for short) in those strategies in which weights are used. We exclude factors that influence the process of setting weights during group decision making, which falls outside the scope of our research. Our focus is not on arguments for weight values in specific decision contexts ('The safety of a minibus is important because an accident can have huge financial consequences') but on factors that apply to decisions in general. From the literature, several classes of factors can be determined:

1: Personal characteristics of the actor. One of the main functions of market research is to identify preferences of specific groups, not only for preferences for specific products (older people generally seem to like pop music less than youngsters) but also in terms of attribute weights. Someone wanting to buy a family saloon may deem top speed less important than the prospective buyer of a sports car, although, of course, it is also possible that the former sets a lower minimum required value (cutoff point) for speed than the latter when using, for example, a satisficing strategy. Attitude towards risk also influences attribute weights (Keeney, 1992). For example, someone wanting to avoid as much as possible the running cost of his car going up because of fuel prices increases may deem fuel burn per kilometer more important and the price of the car less important than someone not concerned with the risk of increasing running costs. Expertise also can play a role. Devine & Kozlowski (1995) found that the level of expertise increases accuracy of decisions in some circumstances (in well-structured decisions, but not in ill-structured decisions);

2: The framing of the decision. Kahneman & Tversky (2000) observed that it makes a difference whether possible outcomes of choices under uncertainty are formulated as gains of as losses. Generally, a gain, for example, (the chance of) receiving a certain amount of money is valued less than (the chance of) losing the same amount. Highhouse, Paese & Leatherberry (1996) found framing effects go beyond semantic manipulation (for example, *presenting* a strategic business issue as either an opportunity or a threat). The perception of actors is at least partly independent from the semantic presentation of information, as far as framing effects are concerned.

Bazerman & Loewenstein (1992) found that the importance of perceived personal benefits may decrease if a reward is presented as relative to rewards that others receive. In the latter case fairness increases in importance. It has been observed that actors desire a higher price for a good when they have the opportunity to sell it then what they want to pay when they buy it (Kahneman, Knetsch & Thaler, 1990). This 'endowment effect' may be linked to the perceived attribute scores of the good dependent on whether actors posess it or not. But change of specific attribute weights influencing utility also seems a possible cause. Goldstein (1990) found that importance judgments may depend on the stimuli in a specific situation and not merely on an actor's global concepts of attribute weight;

3: The presentation and context of the attributes. Weber, Eisenfúhr & Von Winterfeld (1988) state that if attributes are decomposed (split) in several sub-attributes, the weight of the main attribute is lower than the sum of the weights of the constituting sub-attributes. This notion, however, is not universally accepted (Póyhónen & Hámáláinen, 1998). Fisher, Damodaran, Laskey & Lincoln (1987) observed a more or less similar phenomenin, namely that proxy attributes (indicators of attributes that may not be measurable themselves) may under some circumstances be heavily overweioghted, leading to sub-optimal choices. Hsee, Loewenstein, Blount & Bazerman (1999) and Hsee, 1996)) show that the evaluability of attributes influences choice, more readily evaluable attributes can become more important. Trivial attributes may assume disproportional importance if they are, for example, perceived of indicators for the attractiveness of an alternative as a whole (Brown & Carpenter, 2000). So, the color of a car may make that car attractive, although it is an irrelevant attribute (usually, you can buy a car in any color you want). Mellers & Cooke (1994) found that the range of values with which attributes were presented altered their weight. This is in itself in line with normative utility models, but Von Nitsch & Weber (1993) found that adjustment of weights for ranges by their subjects was less than required by normative theory;

4: The person on whose behalf the decision is to be taken. Kray (2000) and Kray & Gonzalez (1999) observed that if actors have to advice someone else, they weigh attributes differently than when the decision pertains to themselves;

5: The choice of strategy may influence the weight values used. A weighted adding strategy implies the use of rather precise weights, like 0.4 on a scale from 0 to 1. EBA merely requires a rank order of importance and a lexicographic strategy only requires the first element of that ranking order (the most important attribute. It is, of course, possible that, instead of the strategy determining the weight values used, the weight values an actor is able to use (together with the information available on attribute scores) determines which strategy is practical. For example: if precise weight values (and attribute scores) are available, it is possible, and desirable from a normative point of view, to use the weighted adding strategy;

6: The phase of the decision process can influence attribute weights. Svenson (1992, 1996) proposes that weights set before a choice is made may be modified after the

choice in order to differentiate the chosen alternative from alternatives not chosen. The intention is to increase the level of relative attractiveness of the chosen alternative. Phillips (2002) adds to this that actors may also re-define attributes (potentially modifying their weights) after having made a choice in order to justify their decision;

7: Reasoning may affect attribute weights. In an experiment by Levine, Halberstadt & Goldstone (1996) subjects who had to reason about their preferences were less consistent in the weighting of attributes than those who could express preferences without deliberating about their motives for these preferences;

8: The type of task may influence weights. For example, according to Tversky, Slowic & Sattath (1988), in a choice task, the more prominent attributes loom larger than in a matching task (making alternative equally attractive by altering attribute scores);

9: Situational factors can have a large impact on choice. Loewenstein (1996) emphasizes the importance of factors like hunger or pain that can make actors make choices that they full well know are sub-optimal in the longer run. This means that some choices are incompatible with the weights of the attributes in order;

10: The time perspective. Kahneman (1994) discusses a series of experiments from which it becomes clear that actors have difficulty in predicting their future preferences. While Kahneman implies this to be the result of changing perceptrions of atrtribute scores over the period between choice and consumption, it seems possible that it could also be a consequenxce of changing weights.

Many of the phenomena described above, like the endowment effect, are violations of normative utility theory that states that weights should be, amongst other things, independent of inconsequential variations in the formulation of attributes and of the procedure to elicit choices (Kahneman, 1994). Actors may, in practice, choose alternatives that they should not, given the attribute scores and weights that they themselves set. The relationship between weights and choices is thus, in practice, not deterministic; actors do not always follow normative models for determining the utility of alternatives or, when they do, they do not always choose the alternative with the highest utility. This behavior is often called irrational' because it deviates from normative decision theory. In Chapters 4 and 5, rationality in importance assessment processes is addressed.

The above factors were not varied and we think they had no disturbing influence on our research. We were not concerned about purely personal preferences (factors 1 and 4) but on importance assessment within an organizational context. The framing and the presentation of the attributes (factors 2 and 3) were as 'value-free' as possible, not even definitions were given, nor was, for example, an assessment of the environment of the organization concerned. Factor 4 was fixed, as already mentioned, and choice strategies were irrelevant (factor 5) because the weighting was done before a choice strategy was applied. We confine ourselves to the initial phases of decision processes: weighting by

individual actors before evaluation of alternatives (factor 6). Reasoning may have influenced the eventual weight judgment (factor 7), but that was of no concern. The general structure of the importance assessment process is unlikely to be influenced by reasoning (see Chapter 2). The weighting we focus on is done for a choice task, but since it takes place isolated from evaluation of alternatives, the influence of the type of task (factor 8) is unknown. Anyway, the type of task is representative for organizational decisions. Factors like hunger or pain were not an issue (factor 9), and we did not assume severe time pressure. In organizational decisions, uncertainty about the future plays an important role, and this was acknowledged in our research.

We can draw the following conclusions:

1: Weights sometimes, but not always, play a role in choice problems. Examples of 'weightless' strategies are satisficing and the majority of confirming dimensions strategy. This means that there need not necessarily be weight assessment processes during decisions. We confine our research to situations where they do: major decisions in organizations where explicit weighting is required and at least the initial weighting is separated from the evaluation of alternatives (see below);

2: The relationships between weights and choices points at the possibility, also suggested by Bettman, Luce & Payne (1998) that the setting of weights and the evaluation and choice of alternatives (or there elimination in, for example, an EBA strategy) can go together in an iterative process. It is then to be expected that weights will often evolve at the same time when at least some of the available alternatives are evaluated. This research, however, is confined to situations where the setting of weights is done before alternatives are evaluated, as may well be done in organizational decision processes. So, the interaction between weights and choices will not be covered any further;

3: Factors that may increase or decrease weights that actors apply can be identified, but in the research discussed above no clear reference was made to the general structure of the importance assessment process. Actors are, in some strategies, assumed to weigh attributes, but how they arrive at those weights does not become clear from these strategies. Factors were given that may influence weights, but through which thinking processes these changes come about is unclear. Much of this research was done by inferring weights from choices subjects made, while we want to study the situation where the setting of weights is separated from making choices (see point 2). We study the structure of the thinking process in which actors arrive at weights, and it may be possible to establish relationships between the factors mentioned above and the (structure of) the thinking process. We will not do that in this research, however, since the focus is on identifying the general structure of the thinking process. We study the importance assessment process in one, ideal typical setting (see Chapter 2). Our emphasis is on internal validity of our results (the model of the importance assessment process) in the setting studied, which should reveal as many aspects of the model as possible. When the generic model is complete, it will be possible to use it in order to assess the influence of the factors discussed above on the importance

assessment process. The research method chosen for this (see Chapter 2) does not lend itself for establishing complex relationships with many variables involved that would have to be controlled;

4: Some strategies require weighting only at the ordinal level, while others require quantitative (interval or ratio) weights. It is thus relevant to know whether actors, when free to set their own weights (so, not limited by a type of weights imposed on them) have a preference for ordinal or quantitative weights. It is, after all, possible that their choice strategy is not only determined by the factors discussed above, but also by the type of weights actors generate. In most research mentioned in this section, the type of weights subjects are to generate or use is fixed. For example: subjects are required to give weights on a ten-point interval scale. Another method widely used is to let subjects make choices and then elicit the weights they have apparently been using. In this case, also, the researcher controls the nature of the weights, again mostly using an interval scale. Our research differs in that the subjects were left completely free in the type of weights they gave, at least initially. Only when subjects gave qualitative (ordinal) weights as their final weight judgment, they were then asked to try and give qualitative (interval or ratio) weights. And if they could or would not do so, this was accepted. So, our research shows, unlike most of the research cited in this section, what kinds of weights actors give when left completely free. If these weights differ significantly from the weights normally present in research, this may lead to a new look at the results of that earlier research.

The conditions described in points 1 and 2 are unlikely to reflect a common state of affairs with, for example, consumers' decisions. Whenever choices are difficult, because emotions are involved, there are a large number of attributes, there is time pressure and so on, strategies that do not involve weighting (like satisficing) or require weighting only at the ordinal level (like EBA) are likely to be used, at least by consumers. With non-routine problems, the type of problems studied in this research, we would likewise expect strategies that require little weighting to be used, at least if we assume that actors within an organizational context behave like consumers. But in many formal decision processes, for example acquisition decisions by organizations, methods used for evaluating the attractiveness of alternatives often involve weights, and the weighting of attributes may be done before alternatives are formulated (so as not to 'contaminate' the weighting process with information about the available alternatives). Some of the methods for determining weights may use hypothetical alternatives (Harte & Koele, 1995; Jaccard, Brinberg & Ackerman, 1986; Keeney, 1992), and the alternatives available in reality may not be known. So, our research pertains only to decision contexts where weights play a role and where the weighting of attributes is done independently of the evaluation of alternatives. It is to be expected that the applicability of the results of our research is limited mainly to relatively important, organizational decisions. Day-to-day consumers' decisions fall outside the scope of our research.
Thus, on the one hand, studying importance assessment processes is relevant for organizational decision making, but on the other hand, the use of strategies which involve weighting attributes may be relatively rare for the setting we want to study: non-routine problems, which can be assumed to be seen as complex by the decision-makers

concerned. Hence, we have to somehow assure that the subjects studied in our research actually perform weight assessment. How we do this is explained in Chapter 2.
Note that our research does not exclude intuitive attribute weighting or choice between alternatives (see, for example, Burke & Miller, 1999, Covin, Slevin & Heeley, 2001, Holder, 1995and Holloman, 1992). Intuition is not explicitly addressed in our research (cf. Bettman, Luce & Payne, 1998), but seems in line with their notion of strategies evolving during the choice process. This so-called incrementalism is addressed in Chapter 4.

The relationships between the perspective of human choice strategies and our research is shown in Figure 6.

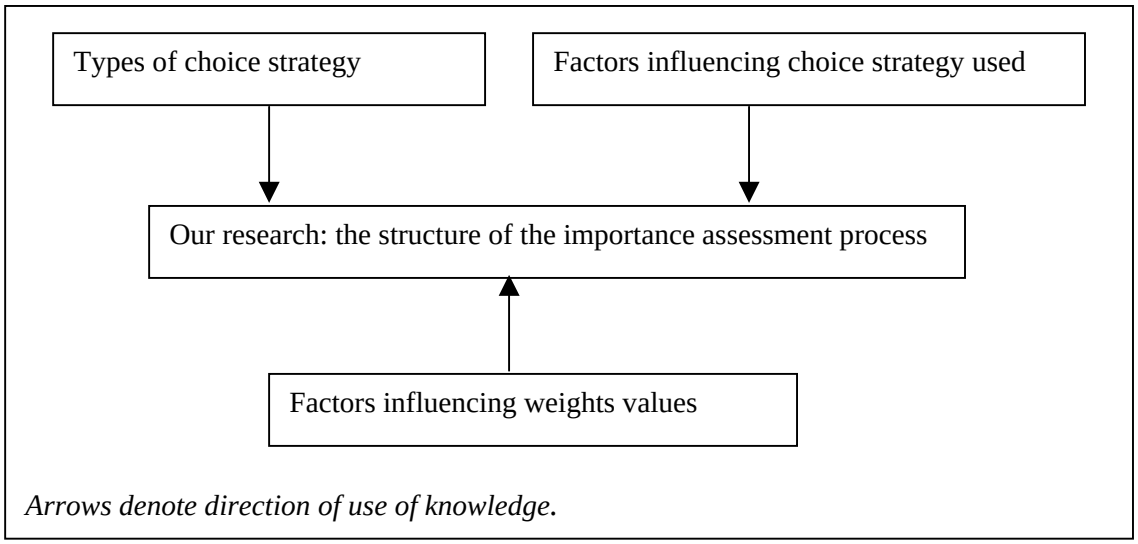



*Figure 6: The place of our research in the perspective of human choice strategies.*

**Descriptive and prescriptive decision studies for specific contexts**

Various methods are aimed at assisting decision-makers in identifying relevant attributes and give pointers how to assess their importance (for an overview, see Johnson & Scholes, 1999). This has been done even for areas that can be characterized as "non-routine" in nature. For example an extensive body of research on location decisions discusses how important decision-makers consider certain factors. Such factors include the fiscal climate, available infrastructure, quality and quantity of labor, distance to suppliers and markets and political stability when choosing a location for their organizations (for a recent overview, see Reijs *et al*, 2001). Also, many case studies have been done on complex, multi-actor decision processes in which arguments for decisions contain, explicitly or implicitly, importance judgments. Examples include the development, especially the specification phase, of new combat aircraft (Aronstein, Hirschberg & Piccirillo, 1998)), political decision-making concerning new military

technology (Enserink, 1993) and the strategies of the Entente and Central Powers during the first month of the First World War (Tuckman, 1962). Narrative descriptions, not written from a scientific perspective, include the design and development of aircraft (Irving, 1993; Kelly, 1990) and strategic decisions made during the Gulf War in 1991 (Gordon & Trainor, 1995).

There is another area not directly concerned with research in importance assessment or judgment, but all the more with its applications: strategic management in the sense of the taking of decisions that concern the long-term future of an organization. It often involves various management functions like finance, marketing, human resource management and research and development.. There are many models for the mapping of attributes relevant for strategic decisions (see for an overview Johnson & Scholes, 1999). Many of these models leave the importance judgments to the actor employing them. Some give simple methods for generating importance judgments or even state a measure of importance. For example the 'impact' of attributes on the achievement of organizational goals or on the added value produced can be a measure of importance while performing a SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis or Strategic Group Analysis (Johnson & Scholes, 1999). Some instruments even give the importance that certain attributes should have in particular situations (Van Beek, 1996). Generally, however, these models are either so general that they still leave the importance judgment in a specific decision process to the decision-maker, or they specify importance judgments for general situations that, in practice, will seldom present themselves in the way theory prescribes. In the end, it is to a large degree up to the decision-maker to arrive at sound importance judgments and to find his way through the importance assessment process. The assessment process as such is not the focus of this type of literature.

The above-mentioned research is not used directly in this thesis, but illustrates the role of importance assessments in practical situations. These situations may be better understood if we can reconstruct the way the actors involved make importance assessments. In Section 3, it was stated that our research might lead to the development of instruments aimed at improving the quality of importance assessments. They may lead to better decisions in situations as described in the literature mentioned above.

## 4.2 Taking it all together: gaps in our knowledge of importance assessment processes

So, we know quite a lot about how to measure importance judgments once they are there, and about environmental, organizational and individual factors affecting these judgments (decision theory). We also know quite a lot about problem-solving and choice strategies and are able to either tell decision-makers how important various attributes are in certain situations from a theoretical perspective or, in general terms, which aspects decision-makers should take in consideration while making importance assessments (problem-solving en design methodology). We can recommend choice strategies for certain decision contexts. And we know how bounded rationality can reduce the complexity of tasks. But we don't know what goes on in the minds of decision-makers during the

importance assessment process. We don't know the general structure of their thinking, we don't know how they structure the various mental activities and we don't know how they handle the inherent complexity of an importance assessment task that makes complexity reduction a necessity, thus perhaps affecting the general structure of the process. Yet there is an important reason not to take for granted that weight assessment is fundamentally similar to problem-solving in general.

Many problem-solving models focus on dealing with logical problems, for example mathematical operations, an area that has been widely researched (see for examples Ericsson & Simon, 1993). But performing an importance assessment inevitably means that personal values are involved. Weights represent opinions, not facts. They express the value actors attach to certain characteristics of alternatives, not the characteristics themselves. It is not to be expected that actors can express their preferences by exhaustive logic. They are sure to be able to give arguments for their importance judgments, but these arguments will likely be only a partial explanation. The manager of the minibus company described in Section 2 can undoubtedly give various arguments why the safety of his minibuses is more important than their comfort: accidents cost a lot of money, may deter future clients, even defects that do not lead to accidents may lead to non-availability of the minibuses, and so on. But he will also be able to given several arguments why comfort is more important than safety: clients take safety for granted and choose their means of travel on the basis of comfort, it gives clients satisfaction, making them come back next time, the competitor also emphasizes comfort and so on. Then the question remains: which attribute is more important at the end of the day? At the end of all arguments and deliberations, personal values come into play. It is clear that importance assessment may, in terms of mental activities, differ from the parts of a problem that can be solved by logic.

### 4.3 Limitations

In this section, the limitations of this research and their general consequences for the external validity of the research are discussed. The methodological reasons for these limitations are discussed in Chapter 2. However, the general philosophy deserves attention here. Our aim was to focus on internal validity. Because, as indicated in Section 4.2, little research has been done on importance assessment processes, we chose to do an in-depth, mainly qualitative, analysis of these processes. Because of the time-consuming nature of the qualitative analysis, we studied a small sample of subjects, accepting low statistical validity (but with explicit criteria for accepting or rejection expectations). Since we dit not quite know what results to expect, we felt that statistical validity was less important then depth. Because we wanted to study importance assessment processes 'uncontaminated' by factors such as group dynamics or previous importance judgments, we studied students with no previous experience with the task at hand, in a laboratory context.
What does this mean for the validity of our research?

1: The small sample. In-depth research with small samples is not unknown (a good example is Hamel, 1990). The disadvantage as far as internal validity goes is lessened as long as the results apply to a large percentage of the subjects, so that clear trends are visible. By using multiple indicators for our variables we tried to eliminate this disadvantage as much as possible, but it remains there. The lack of statistical validity is there and can't be remedied. Selected results from our study could perhaps be tested on a larger scale in order to assess statistical validity.
The small sample had another consequence. It was too small to be divided into one or more experimental and control groups. This means we could not study the behavior of our subjects under various circumstances and that we could not check and refine conclusions in subsequent experiments. This, in turn, means that our results cannot readily be related to, for example, organizational practice without further study.

2: Students as subjects. Students cannot tell us exactly how actors make importance assessments in practice. But because they are not 'contaminated' by past experiences, they can show us which thinking processes are involved when performing importance assessments in complete freedom. Our students had the intelligence to make importance judgments and may, during their later careers, be faced with the task they had to perform in our research. We feel that we get maximum richness of data by using students. Furthermore, because students can be seen as laymen (for the characteristics of laymen see, for example, Chi, Glaser & Farr, 1988 and Van der Heijden, 1998), we can, based on our results, formulate propositions about the behavior of experts that can be tested in future research.

3: Laboratory context. This characteristic of our research also makes it risky to draw conclusions about the way importance assessment processes are performed in practice. With the thinking processes during importance assessments in an 'undisturbed' context, in future research these processes can be observed as they take place in the real world, to see if and in which form the thinking processes addressed in our research can be identified, and what the influences of phenomena like dynamic processes are.

## 5 The structure of this thesis

This thesis is structured in the form of articles, each representing a chapter of the thesis:

*Chapter 2:*
The methodology of the research is described. Basically, the rough data from a think-aloud experiment are converted into a structured description of the problem space of importance assessment processes, as a preparation for Chapter 3, 4, and 5. This concerns the areas of methodology and decision theory.

The methodology was not straightforward and can be seen as a result of the Ph.D. project in its own right. We propose a methodology that combines the flexibility and richness of explanatory research with the potential to test expectations (broadly formulated

hypotheses) that we find in more formalized quantitative research. Our work is based on the think-aloud method as described by Ericsson & Simon (1993) and others, and on the Grounded Theory approach (Strauss & Corbin, 1998). In this chapter, the validity of our research is also addressed. We also define very broadly the various elements that can be used as building blocks for a model of the importance assessment process. These building blocks are not yet related to each other or placed in a sequence. That is done in Chapter 3.

*Chapter 3:*
The phase model of the importance assessment process and the characterization of the way laymen perform importance assessment processes will be discussed. This concerns the areas of decision theory and expertise.

The model of the importance assessment process was the primary result of our study. It describes the sequence of phases that can be identified in the importance assessment process. The phases are placed in a sequential order, but the relationships between the phases are not yet thoroughly explored. This is done in Chapter 4. For reasons explained in Chapter 2, the subjects studied in the experiment that was conducted were laymen. They were university students who had the general intellectual abilities to perform importance assessments of the type we wanted to study, but they were not experts in the field under consideration. In real life, even in the case of non-routine decisions, importance assessments are generally made by actors who are, to a greater or lesser extent, familiar with the decision context. It was, therefore, logical to explore the importance assessment process in terms of expertise.

Two specific areas of the importance assessment processes are discussed in the last two chapters: the relationships between the phases of the model described in chapter 3, and how bounded rationality manifests itself.

*Chapter 4:*
The way actors structure the importance assessment process is studied. Some quantitative and qualitative relationships between the phases of the model developed in Chapter 3 are described in this chapter. This concerns the area of problem-solving and design methodology.

In this chapter we explore the extent to which the subjects who were studied in this research actually followed our model. Explicit attention is paid to the way in which the phases (clusters of activities) of the model are empirically and logically related to each other, thus aiming to validate the use of a, theoretically logical and intellectually convenient, phased model. The main question here is: do actors going through a certain phase of the model build on the results of previous phases, and if so, in what way, or are the phases addressed independently of each other? If the latter is the case, a logical question is: which phases are really relevant for the importance assessment process. Could some of the phases we identified perhaps be eliminated without harming the importance assessment process?

*Chapter 5:*
Complexity reduction in importance assessment processes is the next issue. This concerns the area of bounded rationality.
Importance assessments are very complex mental activates. Since this study deals with non-routine, and therefore complex, decisions, it could be expected that bounded rationality as a driver of complexity reduction could well be an important tool to describe the way importance assessment processes are conducted. This issue is covered here. The model developed in Chapter 3 gives us some clues as to how bounded rationality may manifest itself in importance assessment processes, in the sense that some aspects are identified by which the number of alternative courses of action to be taken into consideration by the actor can be reduced. We do not study whether actors are boundedly rational; with the complexity of importance assessments we take that for granted. We want to find out in what ways actors are boundedly rational. As will be shown, the obtained knowledge provides clues for developing instruments that may improve the quality of importance assessment processes.

In the final chapter (*Chapter 6*) of this thesis, some concluding remarks are made and suggestions for future research are given.

Chapters 2 to 5 are based on articles published in or submitted to journals. Therefore, they can in principle be read independently of each other. Common issues like the research method are briefly addressed in all chapters concerned. This induces some duplication, but increases readability.

Many issues of the importance assessment process are not covered in this thesis, although our rough data from our experiment would have permitted us to do so. These details pertain to specific elements of our model and the importance process can be understood in broad terms without them. Examples of issues not discussed are: the role of information-processing, the role of framing in terms of (perceived) preferences of stakeholders and a more detailed comparison of working with qualitative (ordinal) versus quantitative weights. These issues will be covered in future articles.

## Literature


Alpert, M.I. (1971). Identification of determinant attributes: a comparison of methods. *Journal of marketing research*, volume 8, 184-191.

Aronstein, D.C, Hirschberg, M.J, & Piccirillo, A.C. (1998). *Advanced Tactical Fighter to F-22 Raptor*. Reston: American Institute of Aeronautics and Astronautics Inc.

Arthur, W.B. Inductive reasoning and bounded rationality, *Complexity in Economic Theory*, Volume 84, 406-411.

Aumann, R. (1997). Rationality and bounded rationality. *Games and Economic Behavior, volume 21*, 2-14.

Bazerman, M.H., Loewenstein, G.F. (1992), Reversal of preference in allocation decisions: judging the alternative vs. choosing among alternatives, *Administrative science quartily*, Volume 37, 220-241.

Beach, L.R. (1993), Broadening the definition of decision making: The role of pre choice screening of options, *American Psychological science*, volume 4, 215-220.

Beach, L.R. Puto, C.P, Heckler, S.E, Naylor, G & Marble, T.A (1996). Differential versus unit weighting of violations, framing, and the role of probability in Image Theory's compatibility test. *Organizational behavior and Human Decision Processes, volume 65*, 77-82.

Bettman, J.R, Luce, M.F. & Payne, J.M. (1998). Constructive consumer choice processes. *Journal of consumer research*. Volume 25, 187-217.

Boerlijst, J.G, Munninchs, J.M.A & Van der Heijden, B.I.J.M. (1998). The older worker in the organization. In: Drenth, PhD, Tierry, H.K. & De Wolff, Ch. J. (Eds,), *Handbook of Work and Organization Psychology*, Vol 2, 182-213. London: Psychology Press.

Boerlijst, J.G. & Van der Heijden, B.I.J.M. (1998). The older employee in the organization; a managerial and a development problem. In: Rosenfeld, R.H. & Wilson. D.C. (Eds.). *Managing organizations: text, readings & cases*. Maidenhead: McGraw-Hill.

Borcherding, K, Schmeer, S. & Weber, M. (1995). Biases in multi-attribute weight elicitation. *Contributions to decision making-1*, Caverni, J.P, Bar-Hillel, M. & Jungemann, H. (Eds.). Amsterdam [etc.]: Elsevier Science BV.

Brink, W.P. van den (1993).Beslissingsmodellen. In: Koele, P. & van der Pligt, J. (Ed.). *Beslissen en beoordelen.* Meppel: Boom.

Brown, C.L. & Carpenter, G.S. (2000), Why is the trivial important? A reasons-based account for the effects of trivial attributes on choice, *Journal of consumer research*, Volume 26, 372-385.

Burke, L.A. & Miller, M.K. (1999) Taking the mystery out of intuitive decision making, *Academy of management executive*, volume 13, 91-99.

Checkland, P, & Scholes, J. (1990). *Soft Systems Methodology in action*. Chichester [etc]: John Wiley & Sons.

Chi, M.T.H, Glaser, R. & Farr, M.J. (Ed.) (1988). *The nature of expertise.* Hillstate: Lawrence Erlbaum Associates.

Covin, J.G., Slevin, D.P. & Heeley, M.B. (2001) Strategic decision making in an intuitive vs. technocratic mode: structural and environmental considerations, *Journal of Business research,*Volume 52, 51-67.

Damasio, A.R. (1995) *De vergissing van Descartes – gevoel, verstand en het menselijk brein* (translated from English into Dutch by Liesbeth Teixeira de Mattos. Amsterdam: Wereldbibliotheek.

Danaher, P.J. (1997), Using conjoint analysis to determine the relative importance of service attributes measured in customer satisfaction surveys, *Journal of retailing*, Volume 72, 235-260.

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

Devine, D.J, & Kozlowsky, S.W.J. (1995). Domain-specific knowledge and task characteristics in decision making. *Organizational Behavior and Human Decision Processes, volume 64,* 294-306.

Easterby-Smith, M, Thorpe, R. & Lowe, A. (1993). *Management research, an introduction*. London [etc]: Sage Publications.

Enserink, (1993). *Influencing military technological innovation*. Enschede: PhD thesis

Ericsson, K.A. & Simon, H.A. (1993). *Protocol analysis: verbal reports as data* (2nd ed.). Cambridge, MA [etc.]: Bradford Books/MIT Press.

Ericsson, K.A. & Smith, J. (1991). *Towards a general theory of expertise*. Cambridge: Cambridge University Press.

Fennema, M.G. & Kleinmuntz, DS. (1995). Anticipations of effort and accuracy in multiattribute choice. *Organizational Behavior and Human Decision Processes, volume 63,* 21-32.

Fischer, G.W., Damodaran N., Laskey, K.B. & Lincoln, D. (1987), Preferences for proxy attributes: The overweighing bias, Management science, Volume 33, 198-214.

Fisher, G.W (1995). Range sensitivity of attribute weights in multiattribute value models. *Organizational Behavior and Human Decision Processes, volume 62,* 252-266.

Goldstein, W.M. & Busemeyer, J.R. (1992). The effect of irrelevant variables on decision-making: Criterion shifts in preferential choice, *Organizational behavior and human decision processes,* Volume 52, 425-454.

Goldstein, W.M. & Mitzel, H.C. (1992). The relative importance of relative importance: Inferring other people's preferences from relative importance ratings and previous decisions, *Organizational behavior and human decision processes,* Volume51, 382-415.

Goldstein, W.M. (1990). Judgment of relative importance in decision making; global and local interpretations of subjective weight. *Organizational Behavior and Human Decision Processes, volume 47,* 313-336.

Gordon, M.R, & Trainor, B.E. (1995). *The Generals' War; The inside story of the conflict in the Gulf*. Boston [etc]: Little, Brown & Company.

Guo, C. (2001) A review on consumer external search: amount and determinants, *Journal of Business And psychology,* Volume 15, 505-519.

Hagafors, R & Brehmer, B. (1983) Does having to justify one's judgments change the nature of the judgment process?, *Organizational behavior and human performance,* Volume 31, 223-232.

Hamel, R. (2990). *Over het denken van de architect: Een cognitief psychologische beschrijving van het ontwerpproces bij architecten*. Amsterdam: AHA Books.

Harte, J.M & Koele, P. (1995). A comparison of different methods for the elicitation of attribute weights: structural modeling, process tracing and self-reports. *Organizational Behavior and Human Decision Processes, volume 64,* 49-64.

Heerkens, J.M.G. (1999). Bedrijfskundige methodologie; steen der wijzen of steen des aanstoots?. In: Riemsdijk, M.J. van (ed.) *Dilemma's in de bedrijfskundige wetenschap*. Assen: Van Gorcum & Comp. bv.

Hicks, M.J. (1995). *Problem solving in business and management; hard, soft and creative approaches.* London [etc.]: International Thomson Business Press.

Highhouse, S., Paese, P.W. & Leatherberry, T. (1996) Contrast effects on strategic-issue framing, *Organizational behavior and human decision processes,* Volume 65, 95-105.

Holder, B.J. (1995) Intuitive decision making, *CMA Magazine*. Volume 69, 6.

Hollingshead, A.B. (1996), The rank-order effect in group decision-making, *Academic Press,* 181-193.

Holloman, C.R. (1992). Using both head and heart in managerial decision making. *Industrial Management*. Volume 34, 7-10.

Hoog, R de. & Wittenboer, G van den. (1986) Decision justification, information structure and the choice of decision rules, *New differences in Research on decision making,* 191-204.

Hsee, C.K. (1995), Elastic justification: How tempting but task irrelevant factors influence decisions, *Organizational behavior and human decision processes,* Volume 62, 330-337.
Hsee, C.K. (1996), The evaluability hypothesis; an explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational behavior and human decision processes*, Volume 67, 247-257.
Hsee, C.K., Blount, S., Loewenstein, G.F. & Bazerman, M.H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis, *Psychological bulletin*, Volume 125, 576-590.
Irwing, C. (1993). *Widebody; the making of the Boeing 747*. Reading: Cox & Wyman Ltd.
Jaccard, J, Brinberg, D & Ackerman, L.J (1986). Assessing attribute importance. *Journal of Cvonsumer Research, volume 12*, 463-467.
Janis, I.L. & Mann L. (1977). *Decision making: a psychological analysis of conflict, choice and commitment*. New York: Free Press.
Janis, I.L. (1971) Groupthink. In: Straw, B (ed.) *Psychological dimensions of organizational behavior*. New York: McMillan.
Janis, I.L. (1972). *Victims of groupthink; a psychological study of foreign policy decisions and fiascoes*. Boston: Houghton, Mifflin.
Johnson, G, & Scholes, K. (1999). *Exploring corporate strategy, fifth edition*. Harlow: Pearson Education Ltd.
Kahneman, D, Knetch, J.L. & Thaler, R,H. (1990). Experimental tests of the endowment effectand the Coase theorem. *Journal of political economy*, volume 96, 1325-1348.
Kahneman, D. (1994) New challenges to rationality assumption, *Journal of institutional and theoretical economics*, Volume 150, 18-36.
Kahnemann, D. & Tversky, A. (2000) (ed.). *Choices, values and frames*. Cambridge: Cambridge university Press.
Kaufman, B.E. (1999). Emotional arousal as a source of bounded rationality. *Journal of Economic Behavior and Organization, volume 38*, 135-144.
Keeney, R.L & Raiffa, H. (1976). *Decisions with multiple objectives, preferences and value tradeoffs.* New York [etc.]: John Wiley & Sons.
Keeney, R.L (1992). *Value-focused thinking.* Cambridge MA [etc.]: Harvard University Press.
Keeney, R.L (1994) Creativity in decision making with value focused thinking, *Sloan management review,* 3341.
Kelly, O. (1990). *Hornet; the inside story of the F/A-18*. Schrewsbury: Airlife Publishing Ltd.
Koopman, P.L. (1980). *Besluitvorming in organisaties*. Assen: Van Gorcum.
Kray, L. & Gonzalez, R. (1999). Differential weighting in choice versus advice: I'll do this, you'll do that. *Journal of Behavioral Decision Making, volume 12*, 207-217.
Kray, L.J. (2000). Contingent weighting in self-other decision making. *Organizational behavior and human decision processes, volume 83*, 82-106.
Levine, G.M., Halberstadt, J.B., Goldtsone, R. (1996) Reasoning and the weighting of attributes in attribute judgments, *Journal of personality and social Psychology*, Volume 70, 230-240.
Lipschitz, R, & Bar-Ilan, O. (1996). How problems are solved; reconsidering the phase theorem. *Organizational Behavior and Human Decision Processes, volume 65,* 48-60.
Loewenstein, G. (1996) Out of control: Visceral influences on behavior, *Organizational behavior and human decision processes,* Volume 65, 272-292.
Meehl, P.E./ (1954). *Clinical versus statistical prediction*. Minneapolis: University of Minnesota Press.
Mellers, B.A. & Cooke, A.D.J. (1994) Trade offs depend on attribute range, *Journal of experimental psychology*, Volume 20, 1055-1067.
Munier, B et al. (1999) Bounded rationality modeling, *Marketing letters*, Volume 10, 233-248.
Newell, A. & Simon, H.A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ [etc]: Prentice-Hall.

Nitzch, R von & Weber, M. (1993), The effect of attribute ranges on weights in multiattribute utility measurements, *Management science,* volume 39, 937-943.
Phillips, F. (2002), The distortion of criteria after decision-making, *Organizational behavior and human decision processes,* Volume 88, 769-784.
Póyhónen, M & Hámáláinen, R. (1998). Notes on the weighting biases in value trees. *Journal of Behavioral Decision Making, volume11*, 139-150.
Reijs, T, Kuipers, B, Cornelissen, C, Harten, J, Schuur, P, Boer, L. de, Wegen, L. van der, Heerkens, H & Sijstma, F. (2001). *Core centra en vestigingsplaatsen*. Delft: TNO.
Rittel, H.W.J. & Webber, M.M. (1973). Dilemmas in a general theory of planning. *Policy Sciences,* Volume 4, 155-169.
Rubinstein, A. (1998). *Modeling bounded rationality*. Cambridge (MA), London: The MIT Press.
Russo, J.E, Medvec, V.H, Meloy, M.G. (1996). The distortion of information during decisions. *Organizationam Behavior and Human Decision Processes, volume 66,* 102-110.
Saad, G. & Russo, J.E. (1996). Stopping criteria in sequential choice. *Organizational Behavior and Human Decision Processes, volume 67,* 258-270.
Schwartz, N. (2000). Emotion, cognition and decision making. *Cognition and Emotion, volume 14,* 433-440.
Shafir, E. & Leboeuf, R.A. (2002). Rationality. *Annual Revenue Psychology, volume 53,* 491-517.
*Shafir, E., Simonson, I & Tversky, A (1993), Reason-bases choice.* Cognition, *Volume 43, 11-36.*
Sharma, A & Pillai, R. (1996) Customers' decision making styles and their preference for sales strategies: conceptual examination and an empirical study, *Journal of personal selling & sales management,* Volume 1, 21-33.
Simon, H.A. (1960). The executive as decision maker. In: *The new Science of Management Decision*. New York [etc]: Harper and Row.
Simon, H.A. (1976). *Administrative behavior.* New York: The Free Press, London: Collier.McMillan Publishing.
Simon, H.A. (1977). *The new signs of management decision.* Englewood Cliffs, New Jersey: Prentice Hall.
Simonsos, I (1990). The effect of purchase quantity and timing on variety-seeking behavior. *Journal of marketing research,* volume 27, 150-162.
Strauss, A. & Corbin, J. (1998). *Basics of qualitative research; Grounded Theory, procedures and techniques.* Thousand Oaks CA [etc.]. Sage Publications.
Svenson, O (1996). Decision making and the search for fundamental psychological regularities: what can be learned from a process perspective? *Organizational Behavior and Human Fecision Processes, volume 65,* 252-267.
Svenson, O. (1992), Differentiation and consolidation theory of human decision-making: a frame of reference for the study of pre- and post-decision process, *Acta psychologica,* Volume 80, 143-168.
Toda, M (1980). Emotions and decision making. *Acta Psychologica, volume 45,* 133-155.
Tuckman, B. (1962). *The guns of August*. New York: McMillan.
Tversky, A., Slovic, P. & Sattath, S. (1988), Contingent weighting in judgment ad choice, *Annual psychological review,* Volume, 95, 371-384.
Van Beek, T. (1996). *Innoveren in investeren*. Ph.D. thesis, University of Eindhoven.
Van der Heijden, B.I.J.M. (1998). *The measurement and development of professional expertise throughout the career; a retrospective study among higher level Dutch professionals*. Enschede: PhD thesis.
Van der Heijden, B.I.J.M. (2000). The development and psychometric evaluation of a multi-dimensional measurement instrument of professional expertise. *High Ability Studies. The Journal of the European Council for High Ability, 11*(1), 9-39.
Van der Pligt, J. & Koele, P. (1993). Beslissen en beoordelen vanuit psychologisch perspectief. In: Koele, P & van der Pligt, J. *Beslissen en beoordelen*. Meppel: Boom.

Van Riemsdijk, M.J. (1999). Dilemma's in de bedrijfskundige wetenschap; enkele achtergronden. In: Van Riemsdijk, M.J. *Dilemma's in de bedrijfskundige wetenschap*. Assen: van Gorcum.

Von Nitzch, R & Weber, M. (1993), The effect of attribute ranges on weights in multiattribute utility measurements, *Management science*, volume 39, 937-943.

Weber, M, Eisenfúhr, F. & Winterfeldt, D. von (1988). The effect of splitting attributes on weights in multi-attribute utility measurement. *Management Science, volume 34*, 431-445.

Wei, Q., Yan, H., Ma, J. & Fan, Z. (2000) A compromise weight for multi-criteria group decision making with individual preference, *Journal of the operational research*, volume 51, 625-634.

Wilkie, W.L. & Pessemier, E.A. (1973) Issues in Marketing Use of multi-attribute attitude models, *Journal of marketing research*, Volume 10, 428-441.

Zeelenberg, M, Beattie, J, van der Pligt, J. & de Vries, N.K. (1996). Consequences of regret aversion: effects of expected feedback on risky decision making. *Organizational Behavior and Human Decision Processes, volume 65*, 148-158.

# Chapter 2: On a tool for analyzing cognitive processes using exploratory think-aloud experiments[2]


## Abstract

In this chapter, we develop a method that enables cognitive processes to be analysed quantitatively having only a rudimentary conceptual framework in place at the start of the research. With current methods, the general structure of the cognitive process in question has to be known before quantitative analyses can be preformed. With the method we developed, we studied the cognitive process of weighting the importance of attributes in a purchasing decision.

Our method consists of seven steps:

I: Formulating the problem statement and determining the research method (in our case: the think-aloud method);

II: Conducting an experiment;

III: Design of the data-collection method;

IV: Based on a very general analysis of the think-aloud protocols, choosing analytical concepts for identifying key features in the protocols;

V: Concretising the features in specific categories and establishing relationships between categories are established. This leads to a qualitative model of the cognitive process concerned;

VI: Quantitatively analysing the model, i.e. making frequency distributions, establishing correlation values and the like, of well-defined categories of statements as found in the think-aloud protocols;

VII: Assessing the external validity.

The key contribution is the tool that we use for the preliminary analysis of the unstructrured data in step IV, using a very loose framework based on the theory on problem solving.

Our method provides a straightforward model of the weighting process with high explanatory power. We formulated both the general structure of the method and its


[2] Heerkens, J.M.G., B.I.J.M. van der Heijden (2003). *On a tool for analyzing cognitive processes using exploratory think-aloud experiments*. International conference ‘ Innovating HRM?’, November 2003, University of Twente, Enschede (accepted).

validity requirements clearly enough to make it adaptable for research on other cognitive processes. Some salient validity issues are discussed.

## 1 Introduction

In psychology, the think-aloud protocol is an important method for information gathering on cognitive activities (Ericsson & Simon, 1993). This method can be used for both modeling cognitive processes and for testing hypotheses (Van Someren et al., 1994). Yet, making sense of tens or hundreds of pages of think-aloud protocols is extremely difficult when an a priori theory or a certain research perspective that can be used as a basis for structuring the protocols is lacking.

However, from a practical point of view, the think-aloud methodology can be very useful for theory building, since cognitive processes can be studied without having to specify the variables concerned in advance. The power of the think-aloud approach is that the research question can be rather open. Only a situation in which the required cognitive processes can take place needs to be established, based on general ideas of the phenomena to be studied. Of course, the question how to make sense of the rich but unstructured pool of data is far from trivial and this is an important issue that we want to address in this chapter. We want to emphasize that a tool for preliminary analysis of the unstructured data is an essential ingredient in getting useful scientific results. The type of tool that one uses has an immediate impact on the sort of results that one can obtain. Often, purely qualitative descriptions do not provide answers that are precise enough to provide a sensible contribution to the scientific body of knowledge. Quantitative analysis, on the other hand, is often impossible when an analytical framework is lacking. Still, in this chapter, we want to show that useful compromises are possible, if an adequate preliminary analysis tool is part of the research methodology.

An example may illustrate this. In a study we aim to find out how people come to a judgment on the importance of certain aspects of a good they want to buy. From decision theory, we know elements that may be included in the judgment process. It is possible to express these elements in a mathematical formula for utility as a basis for rational behavior. We also know which deviations from rational behavior people are likely to make and which conditions influence their judgment, e.g. the number of items to choose from, or the amount of information that is available. But we do not know how people handle the elements that are included in the judgment process, i.e. the structure of the thinking process. We may know a little bit about the way these elements may change during the thinking process, but knowledge on the kind of thoughts that induce these changes is lacking. In this sense this is explorative research.

It is possible to ask people about the way they make importance judgments. We may interview them and ask them to memorize a decision they made in the past. But if we don’t know what we are looking for, what do we have to ask? We can ask open questions. But if people don’t know how they think and why, what are they going to answer? They are likely to rationalize their thoughts in order to be able to communicate

them to the researcher. Even experts in the field concerned may not be able to explain their way of thinking.
Thus, if we use methods like questionnaires or in-depth interviews, we may end up with results that are either very general, or not reliable, or both. The think-aloud method has been proven to be capable of revealing peoples' thoughts in full richness in a valid way (Ericsson & Simon, 1993) and to enable precise measuring. However, for precise results, as stated, we have to know in advance what we want to measure in terms of operationalized concepts. But in early stages of exploration of a field such operationalizations simply are not available. This makes the choice of a tool for preliminary analysis of think-aloud protocols rather tricky. Any tool is implicitly based on some model of reality. In order to be adequate for explorative research our tool should enable us to start from a *sufficiently general* model, incorporating notions on importance assessment mentioned above. The emphasis on generality reduces the risk of overlooking interesting phenomena due to "tunnel vision". In a later stage, the descriptive power of the general model should allow us to obtain results that provide more precise insights in cognitive processes.

Hence, this chapter focuses on the way in which think-aloud protocols can be analyzed using only a very general structure beforehand while still yielding data that can be analyzed in a quantitative way. The aim is to combine the richness of verbal protocols with the scientific rigor that is common in quantitative analysis. We want to apply this technique for problems for which no prior analytical framework exists. It is to be a generic scheme, adaptable to the research of a range of cognitive processes.

So, the aim of this chapter is:

> *To present a generic research design scheme while using the think-aloud method as the principal data gathering instrument, concerning problems pertaining to cognitive processes that are only to a limited extent embedded in an accepted body of knowledge, culminating in results suitable for qualitative and quantitative analysis*

Our research design scheme is based on our experience in a research project concerning a specific type of cognitive processes: the weighting of attributes that describe alternatives in a decision process. Our starting point was not a methodological problem as such but the need to make a design for our research on weighting processes. We feel that the resulting generalized design as presented in this chapter is applicable to research on cognitive processes in general under the conditions described in the above problem statement. Especially the role of a tool for the preliminary analysis of think-aloud protocols is not well recognized in our opinion. The research design scheme is described in section two. We also discuss some main validity issues that have to be taken into account. In the next sections the research design scheme is illustrated for our research project on weighting of attributes. In section three, the think-aloud methodology that has been used in that study is described. Section four deals with designing a very loose framework for the analysis of the protocols, based on Simon's (1979) problem-solving model. Section five addresses the way the think-aloud protocols were analyzed

qualitatively, whereas the quantitative analysis is discussed in section six. The external validity of the results is addressed in section seven. Finally, section eight concerns a discussion of our results and some recommendations for further research.

**Scientific relevance**

Our research design scheme is relevant for cognitive processes in general and management research in particular.
Management research is seen as an interdisciplinary science (Easterby-Smith et al., 1993; Van Riemsdijk, 1999). Still, it is unclear how managers deal with interdisciplinary problems and how they mentally integrate chunks of incompatible information. We can measure input (information) and output (decisions), but the process that lies in between is a black box. Our method might be helpful in understanding the behavior of individual managers engaged in decision-making.

The link between organizational phenomena and cognitive processes of individual members of working organizations is not often studied, perhaps as the research methods for each part of the link – surveys, case-studies and simulations on the one hand and psychological tests, experiments, think-aloud sessions and measuring brain activity on the other hand - are often incompatible. In order to decrease this knowledge gap, the approach described in this chapter is meant to allow the study of individual cognitive processes in organizational settings.
Our approach was derived with regard for, and hence is specifically geared towards, cognitive processes in judging the importance of attributes when subjects have to choose between alternatives in a decision task. However, there does not seem to be any reason why the approach cannot be used to study cognitive processes pertaining to other tasks. That is to say, the research methodology that has been used seems suitable for cognitive processes in general.

## 2 A research design scheme for think-aloud experiments and aspects of 'design for validity'

We start with the various phases (in Roman script and capitals) that constitute the research design scheme. The main validity issues for each of the phases (in numbers) are described, in general terms. In subsequent sections, the phases and validity issues are dealt with more thoroughly as applied to our weight assessment research project. This section is concluded with a discussion of the new elements in our scheme compared with literature.

### *I. Formulating the explorative problem statement and the choice for a think-aloud experiment*

Cognitive processes are notoriously difficult to observe and measure, because the results of such processes are much more visible than the processes themselves. After all, a large part of any person's cognitive activities consists of automated subconscious thoughts.

Verbalization has its limits.
This is reflected in the validity issues specific for this kind of research:

I.1 Verbalization can lead to sufficiently rich and relevant information;
I.2 Automated behavior can be suppressed in a sufficient way.

Of course, in the data gathering phase recorded verbalization of the think-aloud experiment can be combined with other sources of information in view of specific requirements of the research question, see phase III.

II The design of the experiment:

*A. The choice of the experimental setting;*

*B. The choice of the subjects;*

*C. The design of the assignment*

As for the experimental setting in think-aloud sessions, the usual option is a laboratory experiment with individual subjects with which the cognitive processes can be studied in isolation from external influences and group interactions. This is in line with an overwhelming body of literature on psychological and decision research and it needs, in our view, no further elaboration.
We summarize related validity issues as follows:

II.A.1 Minimizing process distortions.

As for the choice of the subjects there is a direct relation with the assumptions underlying the research question and also, the external validity of results that one strives for. The following validity issues prevail:

II.B.1 Qualification for the task to be solved in the assignment;
II.B.2 The degree of experience with previous (analogous) tasks.

These qualifications usually refer to certain requirements on expertise necessary for the task or inherent to a role given to the subject. In experiments where the research question is about routine tasks, the task in the assignment has a certain familiarity to the subjects. However, cognitive processes for non-routine tasks are even more interesting and challenging from a methodological point of view. Also, they pose less risk of automated behavior. The familiarity of the subjects with the task should be closely controlled.
As for the structure of the assignment there is a strong relation of validity with avoidance of automated and/or enforced behavior. The amount of information and the way it is provided is crucial in this respect. Subjects performing non-routine tasks may need a certain amount of information, but care should be taken not to overwhelm them. Furthermore, commitment with the task plays a role.
Altogether this leads us to the following validity issues:

II.C.1 The artificiality of the task / role should not lead to a-typical or automated behavior or non-commitment;
II.C.2 Stress and time pressure control, in order to prevent induced automated behavior;
II.C.3 Control of the supplied information.

***III: Data collection design***

In think-aloud experiments audio recording of the sessions is usually the basic material. In addition, a method that answers to specific data requirements needed for the research question can be chosen.
Such additional data collection instruments can take various shapes, like observation protocols, notes made by the subjects themselves, notes made by the experiment supervisor (e.g. if some form of interviewing is used), recording of computer actions or video recordings. Altogether there are several validity issues concerning the interaction of the subject with the recording medium or an interviewer. These can be summarized as follows:

III.1 Control of interference with the medium and / or experiment supervisor.

***IV: The design of a tool for preliminary analysis.***

The input of this phase is general knowledge about areas related to the research question, plus the as yet unstudied protocols. The more the research has an explorative character, the less the variables and relationships to be studied are defined. The output is a model of the cognitive processes to be studied, culminating in a coding scheme in which variables and concepts are available that can be used as the starting point for the qualitative analysis. The coding scheme will, at this stage, merely refer to identifying potentially relevant variables and concepts, not to establishing relationships between them or measuring quantities or variables. In this sense the analysis is preliminary to the following phases.

Especially in exploratory research, a basic framework, a model for analysis is needed to make sense of the think-aloud protocols. This model provides the elements to look for in the think-aloud protocols. In exploratory research sufficient generality of the model is a main criterion. If not, interesting phenomenae may be ignored from the outset. Generality has repercussions in terms of lack of detail. The model is not necessarily complete and it does not necessarily relate the various elements to each other. Furthermore, the model should be capable of handling the dynamic perspective, inherent for a process description, up to a certain degree. In case of a static perspective various static elements (say, input and output elements of reasoning) can be captured, but not, to any degree, the activities that lead to changes. Of course evolution of cognitions is a characteristic of cognitive processes in think-aloud experiments.

This framework should form the basis for further qualitative analysis and also it should preferably yield some basic quantitative information, like, say, how many subjects performed specific kinds of activities.
Validity issues in this respect can be summarized as follows:

IV.1: Construct validity.
This issue refers to picking the right elements from a body of knowledge concerning the research question and framing them in a sufficiently general model. This is a challenge if there is little or no previous research available to use as a guideline.

Then, picking the right elements is an art; if the researcher doesn't have the talent no interesting or relevant picture in terms of results will emerge;

IV.2: Defining the various elements in the general model.
This should be done in such a way that they can be unambiguously identified in the protocols. If the problem given to the subjects is unstructured and the assignment does not prescribe a solution method the subjects have a lot of freedom in the execution of the task. The more freedom for the subjects, the more difficult unambiguous identification of the model elements is likely to become. Therefore, it is non-trivial to define the elements of the model in such a way that they are general enough to discover regularities across the protocols, and yet so specific that they can give meaning to the text of the protocols;

IV.3: Coder consistency.
In line with the previous point, coder consistency is a challenge as well. Generality in the model may easily reflect itself in difficulties to define variables and indicators so clearly that a coder would give the same code to identical (but sometimes differently formulated) variables over time. The elements in the general model should be straightforward enough for operationalization to ensure correct interpretation by the coder. The analysis tool has to lead to identical labels to similar phenomena, even if they manifest themselves in different ways across and within protocols. This does not mean that two or more coders should code the same phrase identically; the classic meaning of coder consistency. In this phase, the coding is done by one person, so that he or she can develop experience and a 'feeling' for the protocols at hand. So, coder consistency here means that the coder remains consistent during the coding process. Naturally, as experience and insight in the protocols is gained, consistency may suffer. This has to be addressed by, for example, going through each protocol several times. In phase VI coder consistency will be addressed again, then in the more common meaning of inter-coder reliability;

IV.4: Cognitive modeling restrictions.
The analysis tool should not restrict the interpretation of the protocols in the sense that relevant information in the protocols would go unnoticed in the qualitative analysis. There is the danger that the researcher focuses too much on labeling identifiable elements in the analysis tool with clear, but too restrictive, meanings, thus leaving out some other interesting aspects.

***V: The qualitative analysis***

In exploratory research this can be done according to the Grounded Theory approach (Strauss & Corbin, 1998). This is a method for building rather than testing theories.
The following steps can be distinguished:

1: Labeling phenomena. This entails deriving general concepts from a body of qualitative data and hence is closely related to phase IV;

2: Discovering categories, that is, grouping various phenomena on the basis of some similarities;

3: Naming categories, that is, finding names that best describe the uniqueness of the categories;

4: Developing categories in terms of their properties and dimensions;

5: Establishing relationships between the various categories.

Basically, this method involves studying the think-aloud protocols a number of times. One of the aims is to refine the preliminary analysis tool until sufficient confidence in its validity issues is built up. Next, the essence of science is addressed: to identify relevant phenomena that have not been identified in the general model of the previous phase, and to establish some (qualitative, logical) relationships between those phenomena. Again, but with more structure than before, the resulting model should be able to deal with a dynamic perspective of the activities the subjects performed during the cognitive processes (e.g. their sequencing) and relationships instead of describing merely static behavior. Typically, the dynamic perspective can reflect itself in recognizing phases in the cognitive process under study consisting of clusters of task related activities and establishing relationships between these phases.

Validity issues are partly the same as for the previous phase (designing the preliminary analysis tool), but in addition:

V.1: Objectivity in establishing relationships between phenomena.
In the Grounded Theory approach, there are no fixed criteria in order to establish whether relationships exist. The method is meant for exploring relationships in a qualitative way. The relationships found need to be tested in quantitative research.

The input of this phase is the content of the protocols from the experiment (see phases II and III) after applying the preliminary general coding scheme (see phase IV). The output is a qualitative model of the cognitive process to be studied from a dynamic perspective.

***VI: The quantitative analysis.***

The aim of the quantitative analysis is to quantify some of the results of the qualitative analysis. For example, if the qualitative analysis leads to recognition of phases consisting of clusters of activities, then in the quantitative analysis, we want to find out how much effort the subjects devote to each cluster. Besides, we want to check the main results of the qualitative analysis in terms of significance. The input of this phase is the qualitative model of phase V, on the basis of which a formal coding scheme for measurements in a dynamic perspective is based. . The output consists of results on categories of phenomena structured and stored in such a way that statistical operations on them are possible.

Whether it is possible to establish statistical relationships depends on the number of subjects in the think-aloud experiment. But, even if the number of respondents is (far) too small for statistically sound conclusions, quantitative analysis is worthwhile, since it provides more precise information on what to expect in a generalized context. In this chapter we shall not elaborate on the statistical validity issues. They are well-covered in a wealth of literature on empirical research (see for an overview (Cooper & Schindler, 2003). Additional validity issues are:

VI.1: Content validity of the named categories of phenomena.
This concerns the relationship between what we measure and what we assume to measure given the name and inherent interpretation of the category. This is an issue for, despite the work done during the qualitative analysis, further operationalization of coding categories was sometimes necessary and protocols were so diverse that the relationship between phrases in the protocols and the categories they seemed to pertain to were not always clear;

VI.2: Inter-coder reliability.
The issue is: to what extent did different coders give the same codes to a category of phenomena. This is always an issue with coding, but in case of an unstructured nature of the assignment the subjects have to fulfill, it calls for extra attention.

***VII: Assessment of the external validity of the final results.***

Up to now the focus was on the internal validity issues in the think-aloud experiment. Of course there is a claim for a more general scope of validity of the final results than the experiment itself. This gives rise to the following validity issue:

VII.1: Well-founded external validity.
Of course, this issue should be anticipated already in the earlier phases of the research design discussed above. But, as a final step after execution of the research design, reflection on external validity is necessary. It is not uncommon, that this leads to further interesting research question.

**What is new about our research method**

Our method is innovative in three respects. Firstly, it provides a structured way to obtain exact, even quantitative information about cognitive processes from think-aloud protocols in case of explorative research without a specific analytical framework being available beforehand. There are numerous generic methods available (see, for example, Chi, Glaser & Farr, 1988; Eisenhardt, 1989; Strauss & Corbin, 1998; Van Someren et al., 1994) However, they either take certain general models as input (van Someren *et al*, 1994), are more directed towards analysis of interview protocols (Strauss & Corbin, 1998), or they are not directed specifically at analyzing cognitive processes but merely measuring their results (Chi, Glaser & Farr, 1988; Eisenhardt, 1989). Secondly, it provides a way for analyzing protocols from experiments where subjects have complicated tasks for which no pre-defined solution procedures exist. In real life, we often encounter such non-routine

tasks. Accordingly, a method to analyze them in-depth is valuable. Many think-aloud experiments concern more or less structured tasks (see for an overview Ericsson & Simon, 1993). The third innovative aspect is the emphasis on a general model as a means for preliminary analysis of the protocols in the think-aloud experiment. This general model is not our definitive analysis model, but it provides a general idea of elements to look for in the protocols, a sort of 'checklist'. The use of such a list of general elements to look for proved to be a valuable asset compared with starting from a completely blank sheet (if this were possible at all), if an elaborated model is not available.

Our method is a combination of existing methods (with an emphasis on the interface between the think-aloud method and the Grounded Theory, the tool for preliminary analysis of the protocols), rather than a completely new method. Still, it opens up areas for which, as far as we know, a sound and practical research method was not available until now.

This concludes the general description of our research design. In the following sections, the way the research design scheme can be used in our research on attribute weighting, how that research project was conducted and the way the validity issues were handled are discussed more thoroughly.

## 3 The think-aloud method in attribute weighting

### 3.1 Pros and cons of the think-aloud method

Let us now illustrate the logic behind the research design scheme in more depth with our research project on attribute weighting. The (summarized) problem statement of the research can be formulated as follows:

> *How do individual actors within an organizational context arrive at importance judgments when involved in the initial phase of buying a capital good, in cases where this is a non-routine decision for the actors?*

The first phase of the research design scheme asks for reflection on whether a think-aloud experiment fits well with this research question (research design scheme, phase I). Section 3.1 starts with an outline on the reasons for choosing the think-aloud method as the general format for our study. In section 3.2, the way in which the think-aloud experiment was designed is discussed with regard to the assignment, the choice of the experimental setting and the choice of the subjects (research design scheme, phase II). Also, some details on our data collection are given. (research design scheme, phase III) in section 3.3. Section 3.4 addresses the related validity issues of the think-aloud method for this specific research project.

***Requirements on the research method in case of our attribute weighting project and the choice for a think-aloud experiment (phase I)***

From the problem statement of our research some requirements (denoted as R) for the research method can be derived. In our case, the research method should be capable of:

R1: Being used to study individual actors. It does not have to be suitable for studying groups;

R2: Being used in either a 'real-life' or a simulated organizational context. It does not have to be suitable for studying subjects in a private environment;

R3: Being used to study evolution during cognitive processes, not just the verbal or behavioral results of these processes. This, of course, leaves open the possibility that verbal or behavioral phenomena serve as indicators for the thought processes;

R4: Being used in a situation where the thinking and behavior of the subjects is not regulated by formal procedures or constraints, as may be the case in especially the preliminary phases of a non-routine decision process. Subjects can use any concept of 'importance' they desire and perform the importance assessment in any way they find appropriate. Since no specific concept exists concerning the way actors conduct importance assessments, this means that it is unknown in advance which variables are relevant. So, we need a research method that keeps our options open as far as the variables to be studied are concerned;

R5: Being used with subjects who have little or no experience in the task at hand. Our research concerns non-routine decisions, in which no previous importance judgments are readily available. This implies that making a new judgment is both necessary and relatively challenging. It is therefore doubtful that the subjects will use standard methods for making importance assessments, even if such methods would be available. The consequence is the same as in the previous point: the research method has to be able to capture a potentially wide array of variables and structures of assessment processes.

As will be explained shortly, the think-aloud method fulfills all the above requirements. In its simplest form, the think aloud-method is nothing more than letting an individual subject verbalize every thought at the time and in the form in which it occurs to him or her (Davison *et al.*, 1997; Ericsson & Simon, 1993; Séguinot, 1996; Van Someren et al., 1994;). The verbalizations are recorded on tape and typed out partly or completely for analysis. The typed-out version of the verbalizations is called the think-aloud protocol. Cognitive processes take place in a person's mind and hence cannot be observed directly. Certain behavior, for example the choice of a good from several alternatives or a stated importance judgment, can be seen as results of cognitive processes. But these processes themselves remain largely invisible, i.e. a 'black box' in terms of Séguinot (1996). Since our research aims to analyze the reasoning that takes place within the 'black box', it is logical to choose the think-aloud method as a means of gathering data.

Let us now consider the issue of verbalization in the think-aloud method (I.1). In comparison to another, very popular, verbal method, retrospective reporting, the think-aloud method offers several advantages. The first is that the subjects do not have to have insight in their own mental processes, since they don't need to explain their thoughts. Furthermore, subjects can be given a task to perform and asked to think aloud while doing it, so their is no need for prior experience with the task because no reporting on previous experiences is sought. Thirdly, it is difficult for the subjects (although perhaps not impossible) to rationalize their thought processes, i.e. to make them look more rational than they actually are. Fourthly, the think-aloud method enables to analyze from a dynamic perspective in detail how subjects behave when going through a problem-solving process (Ericsson & Simon, 1993), in this case an importance assessment process.

Yet, the think-aloud methodology does not come without a price. It has occasionally been used to analyze complex cognitive processes (Davison et al., 1996), but Ericsson & Simon (1993) give many examples of think-aloud studies concerning the solving of highly structured problems, like mathematical problems or variants of the Tower of Hanoi problem. These problems are characterized by the fact that the number of possible actions is limited (for example, in mathematical problems there may be actions like adding, subtracting and the like), and that the solution can be judged to be right or wrong. In our research, there is an infinite number of possible actions and there is no 'right' solution (see section 4). This makes both the coding and the interpretation of the data difficult.

Processing and analyzing data from think-aloud studies is extremely labor-intensive. This is addressed in the literature (Whitney & Budd, 1996) and definitely proved to be the case in this research. The number of subjects that can be handled is therefore limited if an in-depth analysis is required.

Another drawback of the think-aloud method is that it is applicable only in certain situations. It is not usable for analyzing group processes, since thinking aloud and communicating with others don't go together very well. This is one of the reasons why this research focuses on individual importance assessment processes, the other being that we were unable to isolate determinants for individual reasoning from the influence of group processes. Also, the think-aloud method can only be used for analyzing processes at the moment they occur. There are other pros and cons of the think-aloud method that have been discussed extensively elsewhere (Ericsson & Smith, 1991), but these have no consequences that are specific for this research and correspondingly will not be addressed.

When the characteristics of the think-aloud method are compared to the requirements R1-R5 mentioned before and also to the validity issues of the research design scheme I.1-2 as formulated in the previous section, the following observations can be made. The method focuses on individuals (R1). The assignment that subjects have to fulfill can either take place within an organization (for example, letting a manager think aloud about a decision to be made) or the organizational context can be incorporated in the assignment (R2). Although the thoughts expressed by a subject are only those thoughts that are verbalized, the general structure of the thought process is likely to remain intact as long as the method is used properly (Ericsson & Simon, 1993), and there is no better method available to study thought processes (R3, I.1). Procedures and other constraints can be

controlled by the design of the assignment (R4, I.2) and the method can be used with inexperienced subjects (R5).
The assessment is that even though the think-aloud method has some drawbacks, these can be overcome. The detailed insight into cognitive processes that we strive for in the research project on attribute weighting and the possibility to give tasks to subjects under controlled conditions far outweigh the drawbacks, which in any case could be neutralized to a significant extent.

## 3.2 The think-aloud experiment as applied in our study

### 3.2.1 The setting (phase II.A)

In our experiment we chose for a laboratory setting at our university. Even though an organizational context is part of the assignment, the fact that we worked with student was decisive in this respect.

### 3.2.2 The subjects (phase II.B)

The 18 subjects in our study were seventeen third and fourth year undergraduate students and one MBA-student from the University of Twente, Enschede, The Netherlands. All but one of the undergraduate students studied management science at the School of Business, Public Administration and Technology. The MBA student had a technical MSc. It is a natural assumption that this kind of students has affinity and commitment to the assignment as described (II.C.1) and it is clear that these subjects should be qualified for the task in the assignment (II.B.1). Based on studies on alumni, one can assume that some subjects are going to attain positions wherein acquisition decisions are to be made, over the course of their future career. They were also assumed to be sufficiently able to take the view of a person within an organization having to perform an importance assessment., since they had conducted several practical case projects during their study. Interviews that were conducted after completion of the assignment indicated that this assumption was by and large correct. By using students and not acquisition experts, we aimed to avoid ‘automatic’ thought processes (Ericsson & Hastie 1994) that do not enter working memory and hence are not verbalized. Our second motive for choosing students and not experts is expressed by Harris (1992), cited by Séguinot (1996), in the context of think-aloud studies: “To study advanced forms of a skill before understanding how beginners do it is to build the house before digging the foundations”. Hence in this sense experience was clearly controlled (II.B.2).

### 3.2.3 The assignment (phase II.C)

During the think-aloud sessions, each subject had to execute an assignment. The subjects had to imagine that they worked for a travel company that transported passengers to

Amsterdam Schiphol Airport, some 100 miles away from the city where the students' university is located (University of Twente, Enschede). The students were asked to advise the management regarding the acquisition of a new fleet of minibuses to replace the current one. The advice they had to give was an importance judgment of safety versus passenger comfort. It was stressed that the subjects were allowed to perform the assessment in any way they liked and that there were no limits as to the kind of reasoning that was allowed. Since all subjects but one had a background in business studies we can assume that artificiality of this sort of assignment is not an issue for them (II.C.1).
The only condition that was communicated to them implied that they would have to be able to explain their judgment to the management of the company afterwards as if this were a real-life task. This did not mean that the management had to agree with it. It was also stressed that there was no choice between types of minibuses to be made. The subjects had 1.5 hours to complete the task, which turned out to be more than sufficient for everyone. Hence time pressure was not present (II.C.2).
A problem that came up was that, in order to understand the assignment, the subjects had to absorb a greater quantity of information than we would have liked. Our solution was to present the information in pieces. First, the subjects had to read a short text on the general purpose of the assignment. Then, some information on the company and the decision context followed. By now, the subjects knew that they had to make an importance assessment concerning minibuses, but they did not yet know about which attributes. Also, they were given two brochures about minibuses, so that they had a better idea of what a minibus actually was. They were given fifteen minutes to read through and process the information. Finally, they were given the attributes they had to make an importance assessment about. Hence, information control was a clear issue in this assignment (II.C.3).

### 3.3 Data collection (phase III)

In our research project we used audio recording of the sessions. Interaction with the experiment supervisor was kept at a minimum to avoid interference (III.1). In order to get acquainted with thinking aloud, the subjects performed three exercises before commencing the assignment, as suggested by Ericsson & Simon (1993).
Afterwards an exit interview was conducted with each subject in which issues like their confidence in the quality of their work were addressed.

### 3.4 More on validity issues for this attribute weighting research project

Let us reflect on some of the validity issues. Already in the research design phase it turned out that the validity issues are related in several ways. A clear example is issue I.2 (avoiding automated behavior by a proper combination of subject and assignment). It was implicitly addressed by taking students without relevant experience as subjects, while devising an assignment that was within the realm of comprehension of the students, and by avoiding time pressure.

The exit interview was also used implicitly for some validity checks on time pressure and information control. There were no complaints on time pressure whatsoever. Although, in the exit interview, some subjects expressed a desire for more information, there are no indications that information processing capabilities of the subjects influenced their performance to any significant extent.

The issue of commitment of the subjects to the assignment can also be judged in retro-perspective after the experiment. It seemed not to be a problem. Judging by the elaborate deliberations of all subjects but one, they really did their best. The exception was one subject who felt that the level of importance of both attributes was so obvious that no long deliberations were needed.
Stress from having to think aloud had to be avoided. This is, of course, a general issue with think-aloud experiments. Giving the subjects the opportunity to practice thinking aloud addressed this issue. Indeed, in the exit interviews conducted, some subjects stated it was difficult or unpleasant to think aloud, but nobody indicated this to have been a great problem.

This completes the description of the data-gathering phase of our study. The next step is the data analysis. For this, we developed a general framework, which is discussed in the next section.

## 4 A general model for preliminary protocol analysis concerning attribute-weighting processes (phase IV)

The development of our analytical framework for analyzing attribute-weighting processes started with Simon's general problem solving model, which will be discussed in section 4.1. This leads to a general model of the importance assessment process (section 4.2), operationalized in a preliminary coding scheme (section 4.3). The validity of this approach is addressed in section 4.4.

### 4.1 The starting point: Simons general problem-solving model

Having to give an importance judgment can be seen as a problem, the way to its solution being the importance assessment process. There are hundreds of problem-solving models, many of them prescriptive, some descriptive (see for an overview Hicks, 1995). The problem-solving model of Simon (1979) is very general and directed at information processing by individual problem-solvers, thus fitting in the context of our problem statement.

Simon describes problem solving as 'moving through a problem space' (Hunt, 1994; Newell & Simon, 1972; Simon, 1979). It is clear that this view accommodates a dynamic perspective from the outset. The problem space is the way in which the problem-owner (the person tasked with solving the problem) sees the task environment (the task at hand). The problem-owner is seen as an information-processing system, going from one node in

the problem space to another. In the case of a complex, unstructured problem, the problem space may change as the problem-solving process evolves. A problem space contains all kinds of solutions to a problem and all the results of steps made during the problem-solving process, insofar as they are results of 'legal moves'. A legal move is a transition from one step in the problem-solving process to the next that is in accordance with the problem definition. It might be that not all elements of the problem space are relevant for the solution of the problem, but they are all permitted as results of legal moves. Of course, legal moves capture the essence of the process dynamics, as well as the essence of "solving".

In our study, the importance assessment process is seen as a movement of an actor through his or her problem space. Therefore, it is essential to develop a way of representing the problem space of these actors.

## 4.2 A preliminary model of the importance assessment process

### 4.2.1 The problem space

Of course the problem space depends on the sort of problem considered: in our case it refers to elements associated with importance assessment processes. Essential elements will be attributes and weights. Their background in scientific theory will be discussed shortly. As for cognitive processes arguments are introduced as another element in the problem space. In the dynamic perspective of legal moves made during the problem solving process in importance assessment, we can now identify:

1: The attributes that the actor starts with;

2: All the modified attributes that could possibly be developed from the original attributes;

3: All the weights and weight ranges that could possibly be assigned to any attribute of the problem space;

4: All arguments that could possibly be used for justifying all possible attribute-weight combinations;

5. All possible attribute-weight combinations, plus their associated arguments. Note that not all arguments that could logically be associated with these attribute-weight combinations have to be actually associated with them. This is up to the actor. We are not concerned with whether the assessment process or importance judgment is 'logically' correct. The importance judgment is a subset of these attribute-weight-argument combinations;

6: All possible forms of the utility function used.

The above-mentioned elements form the basis for discovering regularities in the evolution of cognitive states that actors go through when performing an importance assessment. They do not yet provide a list of cognitive operations but merely indications of where to find them. In section 4.2.2, an initial classification of legal moves is given, which was further developed during the qualitative analysis of the protocols. Since the importance assessment process is highly personal, we have no outside norm for legality of a move. Hence in this sort of research on cognitive processes, if a person makes a move we accept it as a legal move. Hence, all moves are legal. Still, we retain the term 'legal moves', in order to follow Simons terminology.

To derive the basis for the problem space of the importance assessment process, we refer to the elements as recognized in the theoretical concept of a utility function. In this concept there is a set of attributes on which an alternative that a decision maker considers, is scored. The scores on attributes are aggregated into an overall utility of an alternative.
The relative importance of each of those attributes is called its weight. (Keeney & Raiffa, 1976) and the simplest utility function is the linear additive function, which can be expressed as:

$$U_i = \sum_{j=1}^{N} A_{ij} W_j$$

where Ui is the utility of alternative i, Aij is the score of alternative i on attribute j and Wj is the weight of attribute J (identical for all alternatives).

Concretely, the score on each attribute of an alternative is multiplied by the weight of that attribute and the results, called 'partial utilities' are added to get the total utility, or attractiveness, of an alternative. The alternative with the highest attractiveness should logically be chosen. An example as used in our study is given in Table 1. A decision-maker has to choose which of two types of minibuses he should purchase for his traveling company.

| | Score on safety (weight = 0.4) | Score on comfort (weight = 0.3) | Score on running costs (weight = 0.2) | Score on price (weight = 0.1) | Total utility |
|---|---|---|---|---|---|
| Type A | 3 | 5 | 1 | 2 | 3.1 |
| Type B | 4 | 4 | 2 | 5 | 3.7 |

*Table 1: The utility of two types of minibuses*

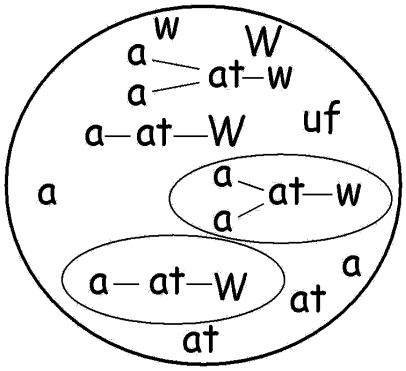
In this case, the type B bus should be chosen for it has the highest total utility.

When performing an importance assessment, it is obvious that arguments are given for the various cognitive operations that are performed, such as attributes that are considered to be relevant or weights that are assigned. Note that, as will be the case in the remainder

of this chapter, 'arguments' here do not only stand for single arguments but also for chains of interrelated arguments.

Weights and scores need not be numerical values, but may be expressed as fuzzy ranges or in a qualitative way in the problem space. For example, a weight may be .3, but may also be somewhere between .3 and .4, or may be given in a qualitative way like 'very important' or 'not important'. So, their measurement level can vary from nominal to ratio, (Blalock, 1981; Swanborn, 1973; Swanborn, 1987). Our framework will accommodate this generality.

Figure 1 shows our problem space for an importance assessment process and its associated weight judgments. In this Figure we can find all elements listed at the start of this section. The encircled weight-attribute-argument combinations represent the combinations eventually included in the importance judgment. The other elements are used at some stage in the decision making process but are not included in the resulting judgment.




uf = shape of the utility function, at = attribute, w = weight value, W = weight range, a = argument

*Figure 1: The importance assessment process*

It is clear that this problem space is infinitely large. There are an infinite number of weights, even if weights are set between 0 and 1. Also, the number of arguments in favor of or against incorporating a certain attribute and for any weight for any attribute that an actor may consider is infinite. No actor can oversee the entire problem space, but no actor needs to. The theoretical problem space, in contrast to the actual problem space, is not what the actor has in his or her mind, but rather the elements that the actor could pick in the quest for an importance judgment.

How does the above help us constructing a general model for preliminary analysis of the think-aloud protocols of importance assessment processes? Firstly, it shows possible elements that mental activities of the actors (legal moves) may be related to. This is further discussed in section 4.2.2. Secondly, the different combinations of elements may

point to phases in the assessment process that can be distinguished, for example: the definition (and elimination) of attributes versus their weighting.

Now that the problem space of the importance assessment process has been defined, it is time to look at some legal moves of this process. We will not discuss all legal moves we derived from the model, but confine ourselves to some examples that give insight in the way our research method works.

### 4.2.2 A classification of legal moves

The general categories of the legal moves we can a priori identify, are in line with Simon's (1979) two sub-processes for the solution of complex problems, i.e. 'an understanding process that generates a problem space from the text of the problem, and a solving process that explores the problem space to try to solve the problem' (p. 268). The first sub-process is often called 'structuring the problem'.

#### 4.2.2.1 Structuring the problem

Simon (1979) indicated that, whereas structured problems like the Tower of Hanoi problem are clearly defined in terms of starting situation, legal moves and solutions (and hence problem space), ill-structured (wicked) problems may need structuring before the search for a solution can start. The structuring process can pertain to:

- the starting situation, like further definition of the attributes to be weighted;
- the result, for example the sort of importance judgment desired (qualitative versus quantitative, the extent of exactness), which sort of arguments are likely to be considered sufficient or legitimate;
- the available means, like information and the organizational context;
- the legal moves to be employed.

The main legal moves of this phase; choosing the shape of the utility function and processing attributes, are discussed below. Other legal moves are not addressed because they are not unique to the importance assessment process.

***Choosing a general form of utility function in the problem space***

In no research that we are aware of, subjects have been asked to explicitly choose a general type of utility function (for example, an additive or multiplicative, linear or non-linear function) before, during, or after the weight assessment process. But it is possible that an actor, familiar with the phenomenon of utility functions, explicitly or implicitly chooses such a function before starting the weighing process.

This possibility will be taken into consideration.

***Processing attributes***

The notion of the processing of attributes comes from two sources; the obvious necessity of comparing attributes in order to determine their relative importance, that is to say, to express them in some sort of common denominator, and the notion of cognitive processes as a series of steps in which one step forms the input for the next one. A logical way to analyze the processing of attributes would be to divide it into a series of steps in which attributes are progressively modified until they reach the stage in which comparison is allowed. In Simon's model, the output of one legal move (a modification of an element of the problem space), in this case an attribute, forms the input for the next legal move. The notion of describing processes as a series of transformations where the output of one transformation is input for the next is also well known in systems theory (De Leeuw, 1997).

***Ways of processing attributes***

The following ways of processing attributes have been identified, both from attribute characteristics described in, for example, methodological and statistical literature (for example, Blalock, 1981; Cooper & Schindler, 2003; Swanborn, 1973; Verschuren, 1980) and from research on attribute judgments. We restrict ourselves to ways of processing that we expect to be reliably identifiable in the protocols.

1: Splitting attributes (see, for example, Borscherding, Schmeer & Weber, 1995; Póyhónen & Hámálánen, 1998). 'Safety', for example, can be split up in sub-attributes that make a minibus safe, like 'strength of the chassis' and 'the availability of seatbelts for passengers on rear seats'. It can also be split up in several effects on passengers like 'the number of accidents' and 'the number of fatalities per accident';

2: Integrating. This is the opposite of splitting. An actor may take several sub-attributes together in one attribute;

3: Concretization of attributes. This means lowering the level of abstraction. For example, safety may be defined as 'the chance that a minibus arrives at its destination without being involved in an accident'. The difference with splitting is that the whole attribute is thought to be covered by the newly formulated attribute;

4: Abstraction as the opposite of concretization;

5: Re-formulation. This implies giving the attribute a new name without effectively changing the measurement unit, or it implies changing the measurement unit without affecting the measurement level, the relationships with sub-attributes (1 and 2) or the level of abstraction (3 or 4). For example, an actor may rephrase 'availability of a stereo set' as 'does the minibus have a radio?' or 'is the braking distance (assumed to be in meters) better or worse than average' as 'how long is the braking distance?

Changes in the way subjects describe attributes and their use during the think-aloud sessions will be used as indicators for the ways in which those attributes are processed.

#### 4.2.2.2 Solving the problem

***Absolute versus relative weighing (pertaining to the weights and attributes in the problem space)***

Absolute weighing means that weights are given to each attribute in isolation, without comparing the importance of the various attributes to each other. Relative weighing means that the subject relates the weights of attributes to each other. The distinction between absolute and relative weighing can also be found in the various types of elicitation methods commonly used (see, for example, Harte & Koele, 1995; Jaccard, Brinberg & Ackerman, 1986; Saaty, 1980).
Timmermans (1993) uses the distinction in absolute and relative scoring of attributes. An example of absolute scoring is 'this minibus has a high level of safety'. Its relative equivalent would be: 'The Opel seems to be safer compared with the Volkswagen'.

***Holistic versus dimensional weighting (pertaining to the weights and attributes in the problem space)***

Timmermans (1993) distinguishes between holistic and dimensional judgment. A holistic judgment covers the alternative as a whole; i.e. 'this is a very attractive minibus'. A dimensional judgment covers only some attributes, i.e. 'this minibus has comfortable seats'. Likewise, in this chapter, and in the remainder of this thesis, a distinction is made between holistic weighting (making a holistic importance judgment), where an attribute as a whole gets a weight, and dimensional weighting, where any or all sub-attributes of an attribute get weights. In the experiment performed in this research, subjects were supposed to perform a holistic importance judgment, but as splitting up attributes is a way of problem structuring (see above), dimensional importance judgments are possible.

***Linking arguments to attribute-weight combinations (pertaining to attributes, weight and arguments in the problem space)***

Often, this legal move will be performed in combination with one or more of the previous ones. An actor may link an attribute to a weight value, or a weight range, and then provide arguments for this action. As with the preceding legal moves, the result may be provisional or final. The actor may consider (groups of) arguments in favor or against the importance of an attribute without linking them to a specific weight value.
Many types of arguments for weights can be identified. One type of arguments stands out in the literature, that is, handling risk (Kahnemann & Tversky, 2000; Keeney, 1992). Because it was impossible to overlook and hence was one of the perspectives from which the initial analysis of the protocols took place, we included this type of argument in our classification of legal moves.

Now that the main elements of the importance assessment process have been identified, a preliminary coding scheme can be developed as the starting-point for the analysis. In the next section, the design of the coding scheme is described.

## 4.3 The tool for preliminary protocol analysis

The primary aim of this tool is to serve as a preliminary coding scheme. It has to identify some of the elements that have been discussed in the previous sections, so as to get some structure in the 'pile of data' that the protocols represented. The coding scheme was further developed during the qualitative analysis, and hence provided a way to structure the results of this analysis. The word 'preliminary' needs some explanation. It is used because the coding scheme developed for and during the qualitative analysis was for the most part rather general and was used more to identify and structure phenomena than to provide precise or so-called 'hard' results. It was used more as a structuring tool than as a measuring instrument. It was 'preliminary' relative to the coding scheme used for the quantitative analysis.

The first part of the coding scheme dealt with identifying the (sub)attributes used by the subjects and the extent of (sub)attribute processing. We started with underlining every (sub)attribute used by a subject and then establishing whether each (sub)attribute was the result of decomposition, integration, specification, abstraction or re-formulation of a previously mentioned attribute. As the subjects seldom identified processing activities explicitly, they had to be inferred. So, a set of rules was developed to separate, for example, specification from decomposition.

An example of such a rule is: if attribute X1 is processed into only one new attribute X2 at a lower level of abstraction and it is clear that the subject sees X2 as covering X1 entirely, it is a case of concretization. If it is clear that the subject feels that X2 only partly covers X1 and hence that more sub-attributes are needed to completely cover X1 (regardless of whether the subject actually mentions these other sub-attributes) it is a case of decomposition. In this way, coding rules were developed for all processing activities. Appendix 1 shows the result of the complete coding scheme for one particular subject. Although this 'attribute-processing scheme' provides some quantitative data, such as the number of sub-attributes generated, it was meant to be largely descriptive and served as a basis for the qualitative analysis only.

With this scheme at hand, more information could be extracted from the protocols. For example, the instances at which weights were given to attributes could be made explicit. This made it possible to recognize not only absolute, relative, dimensional and holistic weighing, but also two kinds of weighting that would eventually emerge as phases in the model, i.e. homogeneous and heterogeneous weighting. The former means weighing two sub-attributes of the same main attribute against each other. The latter means: weighting sub-attributes of different main attributes. Examples are, respectively, the weighing of headroom and legroom (both comfort) and the weighing the quality of the braking system (safety and the amount of legroom or comfort).

Looking at the weights that were given might give an idea of the shape of the utility function, if any, that the subjects used. For example, all subjects used weights that were independent of the score of the attributes. As there were no indications that scaling techniques were used, it seems fair to conclude that the subjects either used no utility function at all, or used a linear additive function.
In sum; with the attribute-processing scheme, three elements of our model (attributes, weights and the shape of the utility function) have been addressed, plus the attribute processing activities and the various types of weighing. 'Addressing' does not mean 'analyzing in-depth'. Rather, depicting the attribute scheme for each subject is only a preliminary step to the start of the further qualitative analysis.

In the next section, the qualitative analysis, in which the second step of the development of the coding scheme took place, is described.

### 4.4 Validity issues concerning the general model and the coding scheme

The validity issues numbered IV.1-4 are relevant here. By taking Simon's general problem-solving model and combining it with literature on decision theory, we hope to have maximized construct validity (IV.1). The construct validity in this research probably was not lower compared with other exploratory studies. The lack of relevant literature was one of the reasons for using the Grounded Theory approach in the next part of the analysis (see the next section).

Defining variables or phenomena unambiguously (issue IV.2) proved difficult. In the end, by crosschecking the protocols (did a certain criterion for distinction that was suitable for protocol one also hold for protocol two), we believe to have achieved a satisfactorily result. But in case other groups of subjects would have been studied, for example, experts, the rules may have to be adjusted.
We took further measures to handle ambiguity. Firstly, in the coding scheme, we included the segments from the protocol that pointed at a certain phenomenon, and checked the coding after some time to see whether our insights had changed. Also, comparing citations from different subjects became relatively easy this way. This frequently led to adjustments in the coding. Secondly, we did not try to infer what a subject might have meant to say, but only what he or she actually said. This seems logical, but sometimes the temptation was strong to interpret before coding. Refraining from inferring may, of course, have led to either incorrect coding or dismissal as irrelevant of certain phenomena. The latter was countered by the fact that there was a lot of repetition in the protocols, a variable that has been missed on one occasion was almost sure to be spotted on another, even if it was formulated slightly differently.

On the whole, we feel that ambiguity was sufficiently low for the further qualitative analysis. Our procedures for safeguarding against too much ambiguity also were checks on consistency (IV.2). This holds especially for crosschecking the protocols and re-studying them over time. As the style of formulation of the subjects was so varied, our iterative way of working was bound to highlight at least some of the inconsistencies in

the coding.

Coder-independence (issue IV.3) was a thorny issue, for the coding required much experience. Solutions that were implemented were to have the more simple coding activities checked by an assistant. Another solution was to check for internal consistency. For example, if a subject indicated in the exit interview that he had not changed his rank order of importance for the attributes (safety and comfort), we would check this against the number of preference reversals the coder had identified. These solutions gave us sufficient confidence in the quality of the coding.

The method of analysis should not restrict our scope of attention beforehand (issue IV.4). This was the reason why we started with a very general model (Simon's model), paying the price of aggravating some of the previous validity issues.

## 5 The qualitative analysis (phase V)

The qualitative analysis was done according to the Grounded Theory Approach. This approach will be described in section 5.1, followed by the description of our analysis in section 5.2. Validity issues are discussed in 5.3.

### 5.1 The philosophy behind the qualitative analysis: Grounded Theory

As discussed before, there are as yet no models available that describe importance assessment processes. Therefore, a method was needed that imposed as few limitations in perspective as possible, so that the assessment process could be viewed with an open mind.

A well-known method that answers to these requirements is the Grounded Theory method (Strauss & Corbin, 1998). This method is meant to generate knowledge about concepts, and relationships between concepts, that are studied in qualitative research.

In this study, the sub-phases 1-5 as mentioned in section under phase V in Section 2 were followed largely, but not completely, sequentially.

### 5.2 The phases of the Grounded Theory approach in the qualitative analysis

#### 5.2.1 Labeling phenomena

Identifying all attributes and the ways in which they were processed, as depicted in Figure 1, formed the start of the analysis (as related in Section 4.3). The first step in using the preliminary coding scheme for further qualitative analysis was identifying:

- The types of arguments used by the subjects (including the special category 'handling uncertainty');

- Ways in which weights were linked to attributes and arguments, and the number of weight-attribute-argument combinations was progressively reduced.

As a preparatory step for further qualitative analysis in this phase, segments in the protocols that pointed at certain variables or activities were identified. These were subsequently inserted in the coding scheme. This is still a rather preliminary method, and rather subjective and explorative. But it shows a first structure hinting at phenomena. In order to give an impression, an example of a part of such an explorative qualitative coding scheme is given in Appendix 2. Essentially, this is the application of the coding scheme developed in phase III.

The next step was the identification of so-called 'working rules', i.e. general labels for the specific mental activities of the subjects. For example, subjects might say: "I know that good suspension is important because my uncle had a minibus with poor suspension and that was very uncomfortable". This would be labeled as 'use of personal experience'. Note that some mental activities may fall under more than one label, even though we tried to avoid this as much as possible. The Grounded Theory approach does not prescribe a set method for labeling phenomena, enabling a combination of creativity and systematic induction.

The key question here is, of course, under which conditions a subject's statement is considered for being re-formulated in more general terms in the first place. For some statements it is clear that labels are useful for describing the importance assessment process. But especially statements that do not strike the researcher as relevant, if only because he has not encountered similar statements made by other subjects, are in danger of being unjustifiably ignored. That is why the phases of the Grounded Theory approach were gone through in an iterative fashion. All protocols were studied at least fifteen times in order to extract meaningful statements from them.

### 5.2.2 Discovering categories

After the working rules had been identified, they were grouped into categories. This was done by means of a number of iterations. First, the rules were simply grouped on the basis of their apparent similarity. For example, all rules that indicated a change in importance of a (sub)attribute were grouped under the category 'weight change rules', and distinguished from rules that pertained to the elimination of (sub)attributes. Weight change rules and elimination rules pertain to an element of the problem space developed in Section 4, i.e. arguments for an attribute-weight combination.

Other categories that pertained to arguments could be identified as well, in particular the so-called reference source rules. These rules define the perspective a subject takes while using an argument, for example the perspective of clients or of the drivers of the

minibuses.

For some rules no suitable category could be found, so they were put into a rest category. An example is the wish to work as systematically as possible.

The next step was relating the categories to each other, sometimes aggregating categories. When the relationships were clear, this provided a system for naming the categories. This step will be dealt with in the next section.

### 5.2.3 Naming categories

Up till now, provisional names have been given to the categories. The next step was to name categories in terms of their place and function in the model representing the importance assessment process. First, a general structure of the model had to be devised. The idea was to try and place categories in some logical sequence. This did not imply that subjects actually followed - or should follow - this sequence, but that in their thought processes certain phases could be distinguished. This approach is in line with Simon's (1979) problem-solving model (see Section 4) and with many other descriptive and prescriptive problem-solving models (Evans, 1991; Hicks, 1995; Lipschitz & Bar-Ilan, 1996).

With the categories that had been generated in the previous steps in mind, the protocols were examined, not at the level of individual statements, but now at a more general level of groups of statements pertaining to one and the same element of the problem space (see section 4.2.1).
The pattern that was discovered was that four clusters of weighting activities were applied by the subjects when performing the assignment:

- Absolute weighting (see Section 4.2.2.2);
- Relative dimensional weighting of sub-attributes pertaining to the same main attribute;
- Relative dimensional weighting of sub-attributes pertaining to different main attributes;
- Holistic weighting the two main attributes safety and comfort against each other.

During the development of the model (see Section 4) we already identified these ways of weighting as legal moves. The qualitative analysis, however, showed that these hitherto isolated phenomena are essential in the importance assignment process in terms of clustered activities, phases in Simon's terminology. This is an essential preparatory step towards measuring the significance of these clusters in a quantitative way. So, of all the phenomena identified in the preliminary analysis tool, only a few were necessary to form the backbone of our model. This was a vital gain from the qualitative analysis. Once the

phases of the model were named, the categories discovered earlier were grouped under these phases whenever possible. Some categories that could not be linked to specific phases were classified as 'auxiliary activities' and analyzed separately. The resulting model is presented in Chapter 3.

### 5.2.4 Developing categories in terms of their properties and dimensions

While the analysis so far was focused on developing the model that describes the importance assessment process, the next two steps were aimed at operationalizing the model so that differences in thought processes between subjects within the framework of the model could be described, as a precursor to the quantitative analysis.

For the category 'processing of attributes', rather precise measurement was possible, as explained in section 4.3. Several aspects of intermediate and final importance judgments were also measured, like the number of preference reversals, the number of (sub)attributes eventually weighted and whether sub-attributes or main attributes were weighted in the final importance judgment.

### 5.2.5 Establishing relationships between the various categories and with other variables

The most basic relationships that have been established are expressing dominant rules in terms of effort devoted to the clusters of activities in the phases as discovered. A working rule was dominant for a subject if it was both observed in the think-aloud protocol and in the exit interview protocol. We assume that, if this occurred, a working rule was really important for a subject, Furthermore, relationships with variables known from other theories on characteristics of expertise, characteristics of process planning and influences of bounded rationality were established. For a discussion of such relationships we refer to Chapters 3, 4 and 5.

This completes the description of the way the qualitative analysis was performed. The result of this analysis was a general model of the importance assessment process, including some detailed descriptions of the way the individual subjects performed this process. Furthermore, it served as the basis for the quantitative analysis, which is discussed in the next section.

### 5.2.6 Validity of the qualitative analysis

For this phase, the previously discussed issues under phase IV are again relevant, but they were already discussed adequately in section 4.4 and we don't want to repeat ourselves. The only new issue is V.1-3. As for objectivity in establishing relationships; in this phase of our study, no definitive statements about relationships can be made; only qualitative

arguments for relationships will be given. On the whole, in terms of Cooper & Schindler (2003), the research in this phase is more descriptive than causal.
However, some objectivity was enhanced by discussing perceived relationships with qualified colleague researchers to avoid 'tunnel vision'. In a qualitative phase this is the best one can hope for.

## 6 The quantitative analysis (phase VI)

In section 6.1, the way the quantitative analysis that was performed is described. In section 6.2, some validity issues are discussed.

### 6.1 The steps in the quantitative analysis

The aim of the quantitative analysis was to answer the following questions:

1: To what extent could the phases (clusters of activities) developed by means of the Grounded Theory approach in the previous section indeed be observed among a significant percentage of subjects? In other words, besides the fact that the phases are logically defendable, are they empirically relevant in a significant way as well?

2: How much effort was devoted to each of the phases of the model of the importance assessment process?

Counting of the properties and dimensions described in section 5.3.4 was not part of the quantitative analysis. It was done during the course of the qualitative analysis.

The procedure for the quantitative analysis was relatively straightforward. Some salient aspects will now briefly be described.

1: The development of the coding scheme. It was established by the researchers how segments should be derived from the protocols and be allocated to the various coding categories. This turned out to be relatively simple. Segments are the parts in which the coding scheme is divided. They are the smallest entities in a protocol to have meaning by themselves; the smallest entities that can thus be coded independently of each other. A segment may correspond to a sentence, but it may also be part of a sentence, or even a single word (in our case, the naming of an attribute).

2: Development of the coding procedure. The procedure developed was inspired by Chi (1997), Ericsson & Simon (1993), and Van Someren et al. (1994). Two coders were asked to code all protocols. A focal point here was inter-coder reliability (Baarda & de Goede, 2001). Because the coding scheme comprised a large number of possible codes to assign, the probability of inconsistency seemed high. Furthermore, although the coding scheme seemed straightforward, the protocols were so diverse that it was to be expected that coders would sometimes have difficulties to distinguish between

adjacent phenomena. Therefore, it was decided to have, in the beginning of the coding processes, a few meetings during which the coders could, along with the researcher, discuss possible problems. These meetings were not meant to directly adjust the coding but to improve the coding scheme. Before each meeting, each coder made a list of problems that he or she had encountered. Then, he or she re-formulated each problem in a general way. For example, if a coder hesitated between 'evaluation' and 'attribute judgment' in case a subject said: "If I give safety a weight of 0.8 and comfort 0.2, then that is too extreme. Safety is worth more than that", then he or she would formulate as a general problem: "If a subject gives a judgment and then immediately modifies it, is the second judgment classified as a judgment or as an evaluation?" Then, during the meeting, it was decided that modified judgments would not count as evaluations but as judgments. Afterwards, both coders would make any corrections necessary in the protocols. While this violates the rule that coders should work completely independent of each other, the advantages of improving the coding scheme 'on the fly' were considered greater than the drawbacks.

3: The coding itself. First, the protocols were segmented, not by the researcher who had already done the qualitative analysis, but by independent coders. The numbers of segments associated with a model element of phase IV or a category of phenomena (phase as discussed in the previous section) are used for quantitative analysis. The results were stored in SPSS-files. SPSS, or Statistical Package for the Social Sciences is a widely used software program for the storage and analysis of quantitative data (Norusis, 1999).

4: The analysis. The first step was to make frequency distributions (of the segments) of all phases as found in the qualitative analysis. Also, some specific issues were addressed that had caught the attention of the coders, like differences between male and female subjects. Also, it was assessed to what extent the subjects jumped back and forth between the phases, using a modified version of the approach of Lipschitz & Bar-Ilan (1996).

Typical examples of the types of results our method of analysis can generate are as follows.
The percentage of effort devoted to absolute sub-attribute weighting, measured in terms of the number of segments, is 27 %. Or, 55 % of the subjects put some effort into heterogeneous sub-attribute weighting. For more details on this phase model we refer to Chapter 3. Figure 2 shows graphically, for one particular subject, the phase to which each sequential segment belongs. If the subject would have executed each phase sequentially in time (so, not jumping back and forth between phases) the figure would show six blocks of increasing height from left to right. This subject followed the phases not completely sequentially. The analysis will be the subject of a future publication. We just want to illustrate the sort of results that the quantitative analysis can lead to.

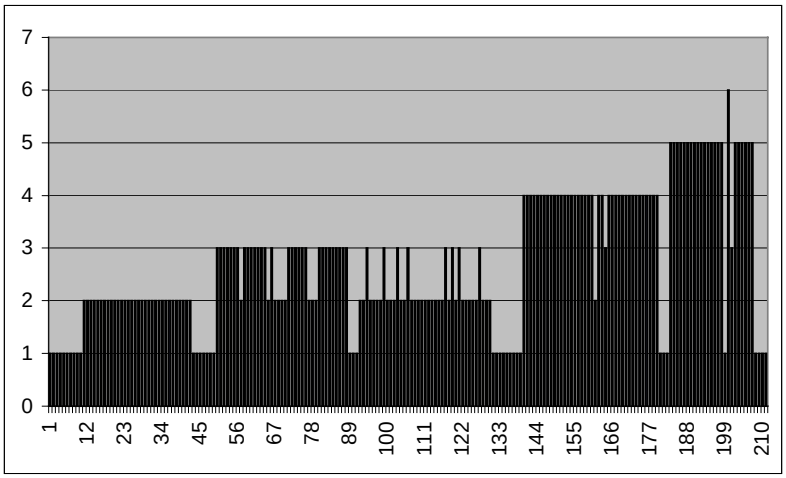



*Figure 2: Distribution of phases over sequential order of segments for one of the subjects*

## 6.2 Validity of the quantitative analysis

The issues VI.1-2 pertain to this phase of the study. VI.1 concerns the content validity of the coding categories. For example, we used the number of segments associated with a particular phase as an indicator of the effort devoted to that phase. This was the most practical solution, but there are other indicators of effort, such as time. The content validity issue was all the more relevant because little research existed on importance assessment processes. To determine the inter-rater reliability (issue VI.2) we used Cohen's Kappa. The value of .97 gave us confidence in the coding. We want to stress that the quantitative analysis was based on the results of the qualitative analysis, providing an extensive check on the logic of the operationalization of the categories introduced and the objectivity of assumed relationships.

This completes the description of the research design. The next section is devoted to the assessment of the external validity of the research resulting from the chosen design.

# 7 Assessment of the external validity of the results (phase VII)

Up till now, we discussed the validity of separate phases in our research. This provides, in our view, an adequate picture of the internal validity of the research, which we think is, on the whole, adequate. What remains, then, is the external validity. The research we present in this thesis was not aimed at developing a research design, but at studying a certain type of cognitive process. The research design was merely a means to and not an end of our research, and the generalization, induction if you want, of our research activities into the guidelines presented in this chapter are a fall-out of our research, an added bonus. Consequently, the general validity of the design has not been determined. We can merely present the requirements to assure validity the way we believe these

requirements were fulfilled on our research. We do not see this as a major drawback, however. Every generic design needs to be adapted to whatever research project it is used for, so the validity needs to be assessed in every specific case. We believe the validity requirements we define make this possible. All in all, we believe that our design is both generic enough and yet addresses the peculiarities of research on cognitive processes sufficiently to be used for a wide range of research on cognitive processes, as long as the validity requirements stated in Section 3.1 are evaluated at the beginning of each specific research project.
So far, we addressed the external validity of our generic design. But what about the external validity of our own results? The fact that we used a small sample in a laboratory context makes it impossible to generalize our results to real-world settings outright. We feel that our results are applicable to real-life settings similar to our laboratory setting: individual layman performing non-routine importance assessments uncontaminated by concurrent evaluation of alternatives, within an organizational context. This situation may not often occur in practice. Still, we believe that the mental activities play a role in real-life organizational importance assessment processes, even though we do not know to what extent and in what form. This can be covered in future research. Our work at least provides a solid basis for this. We now have a model of importance assessment processes that wasn't there before, and that can be further developed.

## 8 Discussion and recommendations for further research

Summarizing, the research approach described here is new in several respects:

- It uses the think-aloud method for studying problem-solving of unstructured problems for which no accepted heuristics exist and the solution of which is highly determined by personal values;

- The analysis of the think-aloud protocols has started with only a very broad model;

- It concerns cognitive processes that, to our knowledge, have not been studied before, so no ready-made experimental setups were available.

In the illustration concerning our attribute weighting research project we demonstrated that our design method enabled us to start without a clear theoretical frame of reference and yet arrive at quantitative data. Usually, it's one or the other. Therefore, we consider it to be a valuable addition to methods such as the Grounded Theory approach. The role of a tool for the preliminary analysis of was emphasized. This research method can be recommended for use in further research on importance assessment processes, for example for studying experts, but also for other cognitive processes of a highly unstructured nature. The method is labor-intensive, but no other major practical problems were encountered. We feel that the method is especially worth considering for areas where psychology meets other disciplines, such as management studies, where methods like case studies and surveys are often used, and which provide very limited insight in cognitive processes. An example would be the process of strategy formulation in a

company. In a case study, written material and interactions between actors can be studied, but what goes on in the minds of those actors can only partially be inferred from their actions. These cognitive processes may be highly personal and may not easily fit into models describing logical problem solving.

Yet, our research method could well be used in such a situation. Actors involved in strategy formulation processes could be given unstructured problems pertaining to various elements (phases) of the strategy formulation process and could be asked to solve them thinking aloud. No prior model of analysis would be required.
Most problems that we encountered in our research were specific for our research problem. The choice of subjects and the nature of the attributes to be weighed (not too straightforward, but not too difficult either), and the way of providing the subjects with enough information to conduct the assignment without overloading them required long deliberations. The only significant problems not specific to our research but typical for our research design scheme pertained to phase IV: the design of a general model for the preliminary data analysis. The first challenge was to assure that there was indeed no model specifically aimed at importance assessment processes. Where do you look, and when can you be sure to have looked in all likely places? The next challenge was to find the clues for the very first outlines of the general model. How can you be specific enough to make the model have added value without narrowing your view too early so that you ignore more promising paths?
The following experience-based tips may be useful for researchers who want to use our design scheme:

1: First write down your own ideas about what the general model could look like on the basis of the protocols, and only then start surveying the literature for more ideas. Starting to examine the literature too early may ‘force’ you in specific lines of thinking prematurely. You have to have a broad arsenal of possible approaches before you settle for a particular one. Remember that it was the lack of usefulness of the literature that drove you to our research design scheme in the first place;

2: Be creative, play ‘advocate of the devil’ don’t dismiss any idea out of hand and look in the reality of your own daily life for inspiration. One of the authors used writing music, fiction and articles for an aviation trade journal as inspiration for the model;

3: Re-read the protocols, not necessarily systematically, but use your intuition when deciding which (parts of) protocols you want to examine. Maximize serendipity by trying to translate everything you encounter, even, or especially, not related to your research, in terms of the problem that is so be modeled. The inspiration for the notion of the importance of absolute weighting came from a poem that went with a Christmas present for one of the authors;

4: Use as many approaches to your problem as possible, and only choose a definitive approach (general structure of your model) when you feel you have run out of new ideas;

5: Before you choose your definitive approach, think (and read) about possible ways of modeling. Miles & Haberman (1994) may provide inspiration for qualitative modeling of complex processes;

6: When having your own ideas more or less in place, talk about the subject with others. Don't be alarmed if they have different views on the model to be designed; in this stage, nothing is settled yet. Be prepared that people will tell you that your problem was solved a long time ago, let them explain to you their models, which will likely be of little use, thank them for their ideas and use the sensible elements in your own model.

These tips may not appeal to everyone, but when no theoretical framework exists for your research, you have to grab every possibility to get 'method in the madness'. The actions described in the tips were of great help to us. More tips may be found in the literature about creativity and problem solving.

Our conclusion is that our method provides a way to study individual cognitive processes while introducing an organizational setting. In this sense, the method has the potential of offering the best of both worlds. Moreover, this study can provide the basis for further research in three areas:

1: Applying the method of analysis on importance assessment processes in other contexts. The subjects in this study were laymen. It has yet to be proved that the concepts used can describe importance assessment processes as conducted by experts, or by actors working in a real organizational context. Actors involved in the acquisition process of, for example, a capital good, within an organization could be asked to participate in an experiment similar to the one we have conducted in our research. The assignment would of course have to be adjusted, that is to say, the company, the capital good and attributes to be weighted would have to be in line with the real-life acquisition process the actors are involved in;

2: Development of similar frameworks of for preliminary analysis for other types of cognitive processes concerning the solving of complex (unstructured) problems. An area that comes to mind is problems where both rational reasoning and values are involved (as is the case with importance assessments), like business decisions where ethics plays a role. For example, research has been done on the choice of methods for transporting and storing dangerous (radio-active) materials. See, for example, Keeney, 1992). While elicitation methods exist for assessing the perceived importance of attributes relevant for the choice, it would be interesting to study the way actors handle the ethical problem of weighting the importance of subjecting people to safety hazards versus, for example, financial attributes.

**Literature**


Baarda, D.B. & de Goede, M.P.M. (2001). *Basisboek methoden en technieken*. Groningen: Stenfert Kroese.

Blalock, H.M. (1981). *Social Statistics*. Revised 2nd edition. Auckland etc: McGraw-Hill.

Borcherding, K, Schmeer, S. & Weber, M. (1995). Biases in multi-attribute weight elicitation. *Contributions to decision making-1*, Caverni, J.P, Bar-Hillel, M. & Jungemann, H. (ed). Amsterdam [etc.]: Elsevier Science BV.

Chi. M.T.H. (1997), Quantifying qualitative analyses of verbal data: a practical guide, *The journal of the learning sciences*, Volume 6, 271-315.

Chi, M.T.H, Glaser, R. & Farr, M.J. (Ed.) (1988). *The nature of expertise.* Hillstate: Lawrence Erlbaum Associates.

Cooper, D.R. & Schindler, P.S. (2). *Business research methods, 8th edition.* Boston: Irwin/McGraw-Hill, 2002 cop. 2003.

Davison, G., Vogel, R. & Coffman, S.G. (1997) Think aloud approaches to cognitive assessment and the articulated thoughts in simulated situations paradigm, *Journal of consulting and clinical psychology*, Volume 65, 950-958.

De Leeuw, A.C.J. (1997). *Organisaties: management, analyse, ontwerp en verandering*. Assen: Van Gorcum.

Easterby-Smith, M, Thorpe, R. & Lowe, A. (1993). *Management research, an introduction*. London [etc]: Sage Publications.

Eisenhardt, K. (1989) Building Theories from case study research, *Academy of management review*, Volume 14, 532-550.

Erocsson, K.A. & Hastie, R. (1994). Contemporary approaches to the study of thinking and problem solving. In: Sternberg, R.J. (Ed) *Thinking and problem solving*. San Diego: Academic Press.

Ericsson, K.A. & Simon, H.A. (1993). *Protocol analysis: verbal reports as data* (2nd ed.). Cambridge, MA [etc.]: Bradford Books/MIT Press.

Ericsson, K.A. & Smith, J. (1991). *Towards a general theory of expertise*. Cambridge: Cambridge University Press.

Evans, J.R. (1991). *Creative thinking in the decision and management sciences*. Cincinnati: South-Western Publishing Co.

Harte, J.M & Koele, P. (1995). A comparison of different methods for the elicitation of attribute weights: structural modeling, process tracing and self-reports. *Organizational Behavior and Human Decision Processes, volume 64*, 49-64.

Hicks, M.J. (1999). *Problem solving in business and management; hard, soft and creative approaches.* London [etc.]: International Thomson Business Press.

Hunt, E. (1994). Problem solving. In: Sternberg, R.J. (Ed) *Thinking and problem solving*. San Diego [etc.]: Academic Press.

Jaccard, J, Brinberg, D & Ackerman, L.J (1986). Assessing attribute importance. *Journal of Cvonsumer Research, volume 12*, 463-467.

Kahnemann, D. & Tversky, A. (2000) (ed.). *Choices, values and frames*. Cambridge: Cambridge university Press.

Keeney, R.L (1992). *Value-focused thinking.* Cambridge MA [etc.]: Harvard University Press.

Keeney, R.L & Raiffa, H. (1976). *Decisions with multiple objectives, preferences and value tradeoffs.* New York [etc.]: John Wiley & Sons.

Lipschitz, R, & Bar-Ilan, O. (1996). How problems are solved; reconsidering the phase theorem. *Organizational Behavior and Human Decision Processes, volume 65*, 48-60.

Miles, M.B. & Haberman, A.M. (1994). *Qualitative data analysis; an expanded sourcebook*. London: Sage.

Newell, A. & Simon, H.A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ [etc]: Prentice-Hall.
Norusis, M.J. (1999). *SPSS 9.0; guide to data analysis*. London: Prentice Hall.
Póyhónen, M & Hámáláinen, R. (1998). Notes on the weighting biases in value trees. *Journal of Behavioral Decision Making, volume11*, 139-150.
Riemsdijk, M.J. van (1999). Dilemma's in de bedrijfskundige wetenschap; enkele achtergronden. In: Riemsdijk, M.J. van *Dilemma's in de bedrijfskundige wetenschap*. Assen: van Gorcum
Saaty, T.L. (1980). *The analytic hierachy process*. New York[etc]: McGraw-Hill.
Seguinot, C. (1996). Some thoughts about think-aloud protocols, *Target*, Volume 8, 75-95.
Strauss, A. & Corbin, J. (1998). *Basics of qualitative research; Grounded Theory, procedures and techniques.* Thousand Oaks CA [etc.]. Sage Publications.
Simon, H.A (1979). Information Processing Theory of Human Problem Solving. In: Estes, W (ed.) *Handbook of Learning and Cognitive Processes* (Vol. 5). Hillsdale: Lawrence Erlbaum Associates Inc.
Swanborn, P.G. (1973). *Variabelen en hun meting*. Meppel: Boom.
Swanborn, P.G. (1987). *Methoden van sociaal-wetenschappelijk onderzoek*. Meppel: Boom.
Timmermans, D. (1993) The impact of task complexity on information use in multi-attribute decision making, *Journal of behavioral decision making*, Volume 6, 95-111.
Van Someren, M.W, Barnard, Y.F, & Sandberg, J.A.C. (1994)*The think-aloud method; a practical guide to modelling cognitive processes*. London {etc}: Academic Press.
Verschuren, P. (1980). *Strukturele modellen met theoretische variabelen*. Nijmegen: University of Nijmegen (Ph.D. thesis).
Whitney, P. & Budd, D. (1996), Think-aloud protocols and the study of comprehension, *Discourse processes*, Volume 21, 341-351.

## Appendix 1: Example of an attribute-processing scheme

The schemes should be read as follows. Safety always gets the number 1 and comfort the number 2. Decomposed attributes at the first level get the numbers 1.1, 1.2, 2.1, 2.2 etc. At the second level, the numbers consist of three digits and can be, for example, 1.1.1, 1.1.2 etc. A letter placed after a certain attribute number means that the attribute is a re-formulation. If an attribute is the abstraction of another attribute, this is noted between brackets. An integration is always the result of two of more attributes being processed, and is also indicated between brackets.
The sub-attributes of the first level are listed as much as possible in the order in which the subjects mentioned them.

### The processing of 'safety' by subject 4

1: Safety
1a: If an accident happens, you want to get out in one piece, preferably unhurt (specification)
1b: I want to get out in one piece or with very minor injuries, but not so that I can sit in a wheeled chair for the rest of my life
1c: Accidents
1.1: Number of deaths per year with a certain brand
1.1a: Number of accidents with which it has occurred
1.1b: Number of deaths per year with accidents (from the context it is clear that it is meant per type)
1.1c: Accident numbers (from the context it is clear that it concerns deaths per year)
1.1d: Maximum so many deaths per year
1.1e: How many deaths per year with accidents and with how many accidents does this happen?
1.1f: Number of deaths per year
1.2: Number of serious injuries
1.2a: Number of serious injuries per year per accident
1.2b: How often does it occur (serious injuries)
1.2c: Figures about serious injuries
1.2d: Number of serious injuries per year
1.2.1: Paralyzed (downwards) from a certain body part or really loose a body part
1.2.1.1: Paralyzed
1.2.1.2: Body part coming off
1.3: Seatbelts
1.3a: Are seatbelts in the car?
1.4: Seat broke loose
1.5: Anti-skid system
1.6: Are there headrests?
1.6a: Headrests
1.7: Can headrests be adapted?
1.7a: Are headrests adaptable?
1.7b: Are they adjustable in height (no specification because this is what he meant with 1.7 and 1.7a)
1.7c: Are headrests adjustable?
1.7d: Adjustable headrests
1.8: Safety for driver
1.9: Safety for assistant-driver
1.10: Safety for passengers

1.11: Airbag
1.11.1: Airbags on the side
1.12: How does a bus fare if you smash into it from the front, the rear, the side and from above?
1.12a: With crash tests what was the result (abstraction)
1.12b: Result with type of accident
1.12c: Result with crash tests
1.12.1: If an airplane crashes on your car
1.12.2: If such traffic pole like you have in Enschede comes crashing into your car from underneath
1.12.3: From the side they come
1.12.3a: If someone comes from the side
1.13: To what extent does a baby sit safely in the car?
1.13.1: Has it got baby seats?
1.13.2: Does the possibility exist to install them (baby seats)
1/13.1/1.13.2(a): Baby seats are they there, can they be installed (integration)
1.13.3: Do baby seats have to be with the face forward or with the face rearward?
1.13.3a: Which baby seats are dangerous, which are not dangerous?

**Appendix 2: Example of the qualitative coding scheme**

- remarks of the coders are made in italics or indicated by the word 'note':

- the numbers in the right column refer to 'working rules' that were identified during the qualitative analysis (see Section 5.3.1).

The distinction between decision rules, weighting rules, implicit and explicit rules could not be unambiguously maintained and was not used during the analysis.

| | **Subject 11** |
|---|---|
| **Number of reformulations safety.** | 0 |
| **Number of reformulations comfort.** | 4 |
| **Reformulation safety taken into account, unity plus level of measurement** | Original formulation (safety) taken. Level of measurement at least ordinal. |
| Scale? | No, so no common scale. |
| **Reformulation comfort taken into account, unity plus level of measurement.** | Original formulation (safety) taken. Level of measurement at least ordinal. |
| **Scale?** | No, so no common scale. |
| **Number of decompositions/specifications safety plus levels.** | First level: 4, second level: 1 specification. |
| **Number of decompositions/specifications comfort plus levels.** | First level: 7, second level: 9, third level: 1 specifiction, fourth level: 2. |
| **Number of sub-attributes safety taken into account plus level of measurement.** | 0 |
| Scale? | Not applicable. |
| **Number of reformulations taken into account sub-attribute safety.** | Not applicable. |
| **Number of sub-attributes comfort taken into account, plus level of measurement.** | 0 |
| **Scale?** | Not applicable. |
| **Number of reformulations taken into account sub-** | Not applicable. |

| | |
|---|---|
| **attribute comfort.** | |
| **Number of integrations/abstractions sub-attributes safety.** | 0 |
| **Number of integrated/abstracted sub-attributes safety taken into account, plus level of measurement.** | Not applicable |
| **Scale?** | Not applicable |
| **Number of integrations/abstractions sub-attributes comfort.** | 'you are picked up at your door and dropped at Schiphol' abstracted to 'comfort of the hole journey' en 'you will be waited for if you have a delay of less than 1,5 hour' abstracted to ' you don't have to adjust to train hours, but Plane Drive will adjust to you'. |
| **Number of integrated/abstracted sub-attributes comfort taken into account, plus level of measurement.** | 0 |
| **Scale?** | Not applicable. |
| **Number of (sub)-attributes that explicitly fall off after being taken into account.** | 0 |
| **Weights of level of measurement.** | Interval. |
| **Safety weight given for every score (sub)attribute?** | One weight for all scores. |
| **Safety weight given for every score (sub)attribute comfort?** | Weight independent of comfort score. |
| **Comfort weight given for every score (sub)attribute?** | One weight for all scores |
| **Comfort weight given for every score (sub)attribute safety?** | Weight independent of safety score. |
| Level of conjunction limit given<br>Note: Even when sub-attributes aren't weighted, they can be the conjunction limit indicator. | No conjunction limit. |
| **Common denominator?** | No. |
| **Preference reversal** | ❑ Number of abstractions weighted: 2<br>❑ Gross: 2<br>❑ Net: 0 |

| | |
|---|---|
| **Decision rules explicit**<br>❑ System in decomposition etc.<br>❑ Conjunction/rules for exclusion/elimination<br>❑ Rules for scale construction<br>❑ Rules for different scores<br>❑ Patterns information gathering<br>❑ Global level of measurement<br>❑ Etc.<br>Note: Indicate if it is about alternatives, attributes or weights. | ❑ Decomposition in comfort for passengers and comfort for driver 2.2.<br>❑ Decomposition in comfort for the bus and comfort for the hole journey 2.2.<br>❑ You have to give a weight factor to sub-attributes. Other test persons are indeed doing this, but aren't addressing it explicitly as rule/goal and*: she only compares sub-attributes of the same main attribute.*<br>❑ With some sub-attributes: pairs wise comparison. *She indicates that this is the method to get weights of sub-attributes, though she doesn't mention AHP, but she doesn't do the method* 3.0.6.2.<br>❑ Safety and comfort are close together*: is this about weight or about empirical relation?*<br>❑ Company is telling things in brochure as they think they matter. Note: implication: don't just acknowledge the folder of Planedrive 3.0.10.4. |
| **Decision rules implicit**<br>For categories see decision rules explicit. | ❑ Level of measurement sub-attributes as known almost always ordinal or yes/no, 1 time nominal, 1 time ratio.<br>❑ First invent sub-attributes yourself, then brochure 2.3.3.<br>❑ 'Weight' is interest. 5.1.<br>❑ Weights normalized to 1. 3.0.9. |
| **Weighting rules explicit**<br>❑ External reference frame<br>❑ Thought experiments, procedure as well as evaluation of weights, for example by hypothetical scores<br>❑ Restricted number of values of weights?<br>❑ Systematic pairs wise comparison<br>❑ Etc.<br>NB: Indicate if it is about alternatives, attributes or weights. | ❑ Clients perspective: comfortable sitting is important for the passenger (weight) 3.0.5.3.2.<br>❑ But with that one over there I see those benches, I wouldn't sit on them on the way to Schiphol 3.0.5.1.1.<br>❑ For the chauffeurs comfort devices like cruise control are important. Note: I am taking that she means the chauffeurs opinion (she also said 'from the chauffeurs point of view'), although it can be that only she thinks it is important. 3.0.5.4.2.<br>❑ The company itself emphasizes comfort. 3.0.5.5.3.<br>❑ I would think seatbelts are more important than an airbag. An airbag only goes for the driver and the passenger sitting next to him 3.0.4.10.<br>❑ Car frame is important for the whole car, so that has to stand above seatbelts. The interview shows that she means, that the whole car is protected. 3.0.4.7 and 3.0.4.10.<br>❑ … for passengers comfort and drivers comfort, because that are two different things according to me. And because driving people is your business, customer's comfort is most important, and that's what you have got to look at mostly. 3.0.4.10.<br>❑ Yes, of course comfort is not only characteristics of the car, (…) the whole journey is comfortable, then comfort also contains being picked up at your door and dropped at Schiphol, (…) yes I, m taking here it's about those cars. Note: The interview shows that she leaves attributes, that aren't considering vans but the whole journey, out of |

| | |
|---|---|
| | consideration. 3.1.1.1.<br>❑ Comfort at least equals train, now it seems more to me, because you are picked up at your door. 3.0.5.9.1.<br>❑ Safety is really important, but I think comfort is even more important, because you already assume some sort of safety at all vans. Later she says: there aren't unsafe cars anymore. 3.0.4.1.<br>❑ Comfort is the ding thing the passenger noticed most. Note: Not very clear. 3.0.5.3.1. |
| **Weighting rules implicit** For categories see weighting rules explicit. | ❑ Weights in last instance determined by test person and customer.<br>❑ 3.0.5.10.1 and 3.0.5.10.2. |
| Handling uncertainty<br>❑ Chance on event as attribute<br>❑ Other weight<br>❑ Extra attribute<br>❑ Multiple weights<br>❑ Discount on score | ❑ If something happens one time, safety is really important. Note: I assume because of (big) consequences. 4.1.<br>❑ The chance for an accident is of course really small, but if it happens, you have to get a van that is prepared. 4.3. |
| **Rules mentioned during interview** Note: with analysis: don't consider rules that are mentioned during interview, but aren't traceable to think aloud session. Rules mentioned after probing by the researcher aren't considered: chance of rationalization afterwards is to big. | ❑ Yes, if you read through the assignment, just what is coming in your head first about safety and comfort. (…). And then I checked safety and comfort at both cars and if there where any aspects I didn't mention. Note: Firs make something up yourself, then brochure during decomposition. 2.3.3.<br>❑ Yes and with comfort I've mostly let myself lead by passengers comfort: Not clear what she means: looking at protocol, I don't assume she is for example leaving chauffeurs comfort out of consideration. Later she also says: I also mentioned the chauffeur's comfort. 3.0.5.3.3.<br>❑ Well at the VW-van, I wouldn't want to sit in there at such a bench the whole journey to Schiphol. Note: Own perspective, score. 3.0.5.1.1.<br>❑ Then I thought comfort can be seen as don't having to go to the station to take the train and to adjust to certain hours, but as comfort of having a company picking you up at your door. (…) Especially with that aspect of comfort, that you can see much wider than just the car. But because you have to weigh between those vans, it appears to me that you still have to look at the car and not at the rest. 3.1.1.1.<br>❑ And because the car frame is for the whole car and for all passengers, I thought that was more important than just airbags that are only for the people in the front seats. Or you have to have airbags everywhere, but still then I would find the car frame for the whole car most important, because it is the outside, and if you have an accident, that still is what gets damaged where it gets hit by forces. 3.0.4.10 and 3.0.4.7.<br>❑ But that (driver's comfort) I would find less important than comfort of passengers you have to accommodate, because they are your customers. 3.0.4.10. |

| | |
|---|---|
| | ❑ So, that way I've made that assessment and that (driver's comfort) seen as less important and didn't mention it the driver's comfort anymore. 3.0.4.10 and 3.1.1.4.<br>❑ Comfort is that which the passenger will seen and feel, and statistically speaking chances of an accident are rather slim. 4.3<br>❑ And because I think safety is a bit more important (because if you need it you've got to have it) I settled for a bit across the middle. But not much, because I think that passenger comfort is quite important. 4.1 and 3.0.7.2.<br>❑ Weights in last instance determined by subjects and clients: 3.0.5.10.1 and 3.0.5.10.2. |
| **What kind of information should you have had (afterwards)?** | None. |
| **What will be done differently next time?** | ❑ Entangle even more sub-attributes with comfort.<br>❑ Reading material a little bit better (to be able to entangle more sub-attributes).<br>❑ Reading the assignment better, so it's clear that it's about the van's comfort and not about the comfort of the whole journey.<br>She would rank sub-attributes and assign rates to main attributes the same way next time. |
| **Remarks** | ❑ During interview she mentioned, not knowing if it was about comfort of the van or comfort of the whole journey, as moment of trouble. But it appeared from protocol this didn't affect the method. |

# Chapter 3: On importance assessment and expertise in non-routine decisions

## An exploratory study on the cognition of weighing processes of capital goods' attributes[3]


### Abstract

How do actors involved in the acquisition of a capital good assess the importance of its attributes? And what is the role of expertise? Numerous instruments exist for measuring the importance attached to attributes, but little is known about the importance assessment process that precedes these importance judgments.

Expectations concerning the behavior of actors facing non-routine importance assessment problems are tested, yielding some interesting results. Firstly, the behavior of these actors is consistent with a newly developed phase model. Even with a non-routine problem, structuring the assessment problem takes less effort than the actual weighting. Surprisingly, weighting attributes in isolation gets much more emphasis than weighting them against each other, despite the latter being the essential part of importance judgments. Despite the subjects being laymen, they showed high confidence in their work.

Finally, propositions concerning the behavior of experts are formulated, based on Van der Heijden's (1998) dimensions of expertise.


### 1 Introduction

Let us first introduce the problem area with a brief case description. Consider a firm that transports passengers to and from an airport some 150 km away. Its clients are, for various reasons, unable or unwilling to travel by car or by public transport. Years ago, a previous manager bought a fleet of Volkswagen minibuses that have to be replaced in the coming year. The minibuses that are available on the market differ from each other in many respects, such as ease of maintenance, running costs, safety features, passenger capacity, passenger comfort and so on. These different minibuses' characteristics can be labeled with the term 'attributes.'

As the present, recently employed, manager wants to standardize on one type of minibus, he has to decide which type to acquire. In order to do this, he needs to decide how important the various characteristics of the minibuses are to him. Is ease of maintenance more important than passenger comfort? If this is the case, and if he has to choose between minibus A that is not so comfortable but easy to maintain and bus B that is more

[3] Heerkens, H. and Van der Heijden, B. (2002). 'On importance assessment and expertise in non-routine decisions: an exploratory study on the cognition of the weighting process of capital goods' attributes', *Int.J. Management and Decision Making*, Vol. 3, Nos. 3/4., pp. 370-398.

comfortable but also more difficult to maintain, he'll probably choose bus A. But what if bus A is somewhat easier to maintain, but bus B is a lot more comfortable? Then the manager has to decide how much more important ease of maintenance is to him compared with comfort.

This example has several characteristics that managers are often confronted with. Firstly, his challenge is to make a choice in a situation where it is unlikely that one alternative is the most favorable with regard to all characteristics. It cannot be expected that one minibus is best in all respects compared to all other minibuses. Secondly, the task at hand is not a routine decision. It will occur no more often than once every so many years. Besides, on each occasion the attributes of the available minibuses will probably have changed, as may be the case regarding the importance of these attributes. Minibuses may have become more comfortable, but clients may rate comfort increasingly important. The manager can be considered an expert when running a fleet of minibuses is concerned. But to what extent does this specific expertise help him?

A rational manager who faces a non-routine decision like the one in the case description basically has two tasks:

1. He has to decide on the relevant attributes and on the importance of each of these attributes, in terms of its influence on the final choice. The thinking process about attribute importance is called the process of *importance assessment* and results in an *importance judgment*: the weight given to an attribute;

2. He has to assess the *values of the attributes* of each alternative to be taken into consideration. Given the importance judgment on each attribute, the most attractive alternative can then be chosen.

In this chapter, we will focus on the first problem: how do actors in non-routine decision-making arrive at an importance judgment?

> *The goal of this study is firstly to develop a generic model of the importance assessment process and secondly to identify how the way laymen (actors with a low level of expertise) go about importance assessment processes can be characterized.*

From a managerial perspective, the importance assessment problem is of interest, as it neither implies a completely logic-driven problem nor a completely value-driven problem. The concept of 'importance' implies that the actor concerned has personal goals, norms and values, or at least preferences with an affective component. But no actor, at least when functioning within an organization, can afford to be too much led by personal preferences. The manager in our case would use financial, strategic or operational arguments to justify his decision in order to, for example, convince the bank to lend him money for the purchase.
Thus, an importance assessment calls upon the ability to solve logical problems, yet it always includes dealing with affective elements as well. The amount of expertise the

assessor has does not matter in this respect. Laymen as well as experts are subject to this personal element in the decision making process.
We are especially interested in decisions where one can expect that the degree of expertise influences the way in which the importance assessment is made. On the other hand, the decisions concerned should not be routine decisions, whether the actor is a layman or an expert. In that situation the decision-maker would be able to fall back on previous importance judgments. Going through the importance assessment process would perhaps not be necessary.
This goal sets two important conditions for our study. Firstly, the decision to be studied should be, as stated before, a non-routine decision. We chose the decision described in the introduction: the acquisition of minibuses. Secondly, we had to control the degree of expertise. We did not compare experts with laymen, for reasons explained later. Instead, we studied actors who were not experts in the field of the particular decision situation (in our case: buying minibuses) but who were qualified as decision-makers in the sense that they possessed:

- a general awareness of the field the decision was about;
- general reasoning skills;
- general problem-solving skills;
- experience with self-evaluation (reflection on performance).

Even if actors possess these qualifications, it is of course not certain whether they will use them in a non-routine decision setting and, if so, in what way.

In this thesis, actors who possess the above-mentioned qualities are called ‘laymen’. We define laymen as actors with little or no experience in the task to be fulfilled. This pertains both to the content of the task, (in our example the task pertained to minibuses) as well as to the procedure to be executed (making an importance assessment). An expert, on the other hand, is a person who has a certain amount of experience with the content of the task (the manager in our example is considered an expert because he is familiar with operating minibuses). The expert may also have experience with the procedure to be executed, but this is not necessary, and in this thesis we focus on actors for whom the task is not routine. So, whatever experience they have with making importance assessments in the case of minibuses should be limited. This definition corresponds with the characteristics of experts common in the literature (see, for example, Chi, Glaser & Farr, 1988), with the exception that most literature we find no combination of a high level of content expertise with a non-routine problem. In other words, in most studies of expertise, experts are assumed to have both the content (domain) knowledge and are familiar with the problem at hand.

Some elaboration may be necessary in order to clarify the distinction between ‘level of expertise’ and ‘level of routineness’. Being an expert does not mean that a decision is routine. An actor can be an expert on cars without ever having had to buy one. Thus, an

expert may know which cars are available, how well they score on certain attributes, how attributes are causally related to each other (for example, how the quality of the suspension of a car may influence both its safety and comfort), and he may have a preference for a certain type of car. But weighting attributes may still be a non-routine task for him, if only because he has never consciously done it. For someone knowledgeable about and skilled in weighting but ignorant about cars, the weighting obviously is a non-routine problem also. Only for experts with extensive knowledge about and skills in weighting and assessing cars could the weighting of car attributes resemble a routine decision. But even this does not need to be the case, for even if an expert has all content and weighting knowledge available, he need not have processed this knowledge and skills into a method readily available for dealing with a particular weighting problem. An extreme example of experts facing a non-routine problem was the first trip to the Moon in 1969. Each of the three astronauts making the trip was as much an expert as was humanly possible at the time. The astronauts had trained for years, and had even made simulated landings on the Moon with an aircraft that emulated the satellite's low gravity field. Yet, nobody would call the first moon shot a routine task. On the other hand, it is difficult to see how a problem can be routine for a layman in the field.

In sum, the terms 'routine' and 'non-routine' apply to decisions and the terms 'layman' and 'expert' apply to actors involved in making decisions. A non-routine decision is a new decision for which no standard methods or procedures are available to the actor concerned. A routine problem is one that has been has been dealt with by the actor before and certain methods and procedures are available to handle it. An expert is an actor who has knowledge and skills available to (elements of) a decision, like content knowledge about the decision context or about weighting of attributes in general, or both. A layman lacks this knowledge and skills. For a layman in the field concerned, a problem is always non-routine.

An important limitation of our research is that we confine ourselves to situations where weighting of attributes is performed. In some choice strategies, weighting is not necessary (Bettman, Luce & Payne, 1998, see Section 4.2 of Chapter 1). Also, we specifically look at situations where the weighting of attributes is done before alternatives are evaluated. This need not always be the case (Bettman, Luce & Payne, 1998). When important decisions have to be made within an organizational context, like the acquisition of a capital good, the conditions on which our research is focused are likely to be in order.

In the next section we will start with a brief theoretical overview on the process of importance assessments and judgments. Subsequently, we will go into the theory on expertise. Expectations are then formulated with regards to certain aspects of expertise (Section 3), followed by the research methodology (Section 4) and the resulting model of the weight assessment model (Section 5). Expectations are tested in Section 6. Finally, we will go into some limitations of the study and we will suggest an agenda for further research (Section 7).

# 2 Theoretical framework

## 2.1 Scientific relevance

Our study is new in several aspects.

Existing literature provides a general indication of phases in problem solving (see, for example, Simon, 1979 and Chapter 1). However, no specific model for the phases in importance assessment exists. Various normative models for eliciting weights (importance judgments) are available (Keeney & Raiffa, 1976), as is a vast body of knowledge on the behavior of experts (for an overview, see Van der Heijden, 1998). But the literature concerns the result of importance assessment (the weights elicited from and used by decision-makers) and behavior of experts solving logical, analytical problems. The way importance assessments are made remains a 'black box'. Understanding the importance assessment process is vital if we want to give decision-makers instruments to help them with these assessments.

Most research on expertise has concentrated on cognitive issues like the mental representation of problems, the role of domain-specific knowledge, the use of solution algorithms and heuristics and the role of logical reasoning (Bereiter & Scardamalia, 1993; Chi, Glaser & Farr, 1988; Ericsson, 1996; Ericsson & Smith, 1991). Insofar as importance assessments have been addressed, they were either seen as 'black boxes' or analyzed in the context of logical reasoning or external factors. Examples of research questions that have been addressed are: 'Which arguments do experts and laymen have for their importance judgments?' and 'To what extent does domain knowledge influence the judgment?' (Camerer & Johnson, 1991; Johnson, 1988).

## 2.2 Views on expertise: the linking of expertise and importance assessment

In research on decision-making there are two views about experts. The first view, which emerges from behavioral research on decision-making, is sceptic about experts. Data suggest that experts from a wide range of expertise domains are not much better predictors than less experienced people. Furthermore, this view suggests a simple technology for replacing experts, i.e. a linear regression model (for example in the domain of medical diagnosis, using illness symptoms and medical judgments as inputs) (Patel & Groen, 1991). The regression does not mimic the thought process of an expert, but it probably makes *more* accurate predictions than the expert does. In some cases a linear regression model containing only three or four variables may accurately predict experts' behavior (Van Dam, 1993) despite experts often believing that they take into account a large number of factors in their decisions.
The second view, stemming from research in cognitive science, suggests that expertise is a rare skill that is developed only after much instruction, practice, and experience. The cognition of experts is more sophisticated than that of laymen; this sophistication is presumed to produce better predictions. This view suggests a model that strives to mimic

the decision policies of experts, i.e. an 'expert (or knowledge-based) system containing lists of rules experts use. Such an expert system tries to match, not exceed, the performance of the expert it represents.
Whereas behavioral decision theory emphasized the performance of experts, cognitive science usually emphasizes differences in experts' thinking processes (Johnson, 1988). In this contribution we use the cognitive approach. The main justification for our approach lies in the fact that expertise and expert behavior can only be understood by studying the phenomena as they are, in their full complexity, and not by 'fragmentations' into 'simplified' research questions. Observing the behavior by means of think-aloud protocols enables us to uninhibitedly take stock of different (chains of) phases respondents go through while making an importance assessment. Also, we can analyze the approach of each subject in terms of expertise-related characteristics. Thus in our approach process and expertise characteristics are related. As far as we know, identification of these relationships has not been achieved before.

## 2.3 Research on dimensions of expertise

Much research has been done on the way experts such as doctors, investment brokers, judges, typists and pilots perform their tasks, often focusing on differences between laymen and experts (see for example Bereiter & Scardamalia, 1993; Chi, Glaser & Farr, 1988; Ericsson, 1996; Ericsson & Smith, 1991; Van der Heijden, 2000). Lots of efforts have been made to develop a general theory of expertise. Sternberg (1995) has tried to place various European and American contributions on expertise in a more general framework that encompasses all of the various conceptions or aspects of expertise. Glaser (1987) was the first who tried to summarize his thoughts on expertise in a set of propositions. He succeeded in encapsulating his thoughts in a set of more or less general statements. Expertise is the result of a continuous development process. It is very specific and development of expertise is influenced by task demands in the course of experience, and other factors.

Although the picture of expertise is biased on account of the highly structured domains in which it has been studied, compared with the study of expertise in complex professional domains like management wherein periodically importance judgments are asked for, it has proved to be very useful as a starting-point for the understanding of the phenomenon and to discover its compiling ingredients (Van der Heijden, 1998; Van der Heijden, 2000).
Theoretical and empirical controversy abounds with regard to the understanding and potentials of the concept of expertise. The divergent meanings attached to the concept of expertise create great confusion mainly owing to the domain-specific character of expert behavior (c.f. Curtis, 1986; Logan, 1985; McLagan, 1997). Because of the fact that most of the researchers restrict themselves to one domain of expertise and try to examine outstanding behavior in that particular domain (Chi, Glaser & Farr, 1988; Ericsson, 1996; Ericsson & Smith, 1991), the understanding of the concept in general is greatly in need of elaboration. A compilation of expertise research and subsistent ingredients that altogether compose professional expertise was made by Van der Heijden (1998, 2000).

Subsequently, the concept of professional expertise has been operationalized in a multi-dimensional way.
Results of studies on expertise seem to show unanimously and quite clearly that expertise comprises

1: a *knowledge* component;

2: a *skills* component;

3: a *meta-cognitive knowledge* component.

In these components, the qualification requirements for decision-makers can be recognized. General field awareness falls under the knowledge component. Reasoning and problem-solving skills are part of the skills component. Experience with self-evaluation falls under the meta-cognitive component.

Nonetheless, the study by Van der Heijden (1998, 2000) among professionals from different job fields, made it clear that these three components are not sufficient to cover the phenomenon of professional expertise in its entirety. Measurement of cognitive abilities and skills is not enough to fully cover the construct (see also Ericsson & Lehmann, 1996; Trost, 1993). Motivational aspects and self-insight, as well as social skills, social recognition and growth capacities are important interactors and moderators. That is to say, there is a compelling reason for the proposal of a broader type of measurement, in which cognitive abilities and overt skills play an explicit, but partial role (Van der Heijden, 1998; Van der Heijden, 2000). For sake of clarification the different dimensions will be explained briefly.

The *knowledge* dimension comprises the different types of knowledge that are inherent to a certain professional field. These different types of knowledge are declarative knowledge (‘knowing that’), procedural knowledge (‘knowing how’) and conditional knowledge (‘knowing when and where or under what conditions’) (Alezamder, Schallert & Hare, 1991).

The second (skills) dimension has to do with the skills an employee needs to perform the required professional tasks. Once the activities and responsibilities have been defined, it is clear which skills are necessary to perform a given job. A person can only be referred to as an expert if his or her overt behavior demonstrates the capacity to perform qualitatively well in a particular domain in terms of benchmarking with colleagues. This component of professional expertise is called the dimension of professional *skill*s. These skills can pertain to the content of the area of expertise, like being able to work with certain software, and process skills, like communication (writing) and project management skills.
This third dimension is the so-called *meta-cognitive knowledge* dimension (‘knowing about knowing’ or ‘knowing that one knows’). This dimension, that has to do with self-insight or self-consciousness, is known by a wide variety of names: meta-knowledge, executive control knowledge, self-knowledge, regulative knowledge and meta-cognitive

strategic knowledge, to mention but a few.
Yet, notwithstanding the importance of knowledge and skills professional expertise can only exist by virtue of being respected by knowledgeable people in the work setting. Even where people are to a large extent endowed with different types of knowledge, with capability for self-regulation and with the capacity to perform quickly and competently, they may not be labeled as specialists. It seems reasonable to presume that other mechanisms play a part in the assessment of people as to whether they are talented or not. It is very important to be able to develop impression management skills, social intelligence and communication skills. That is, the fourth dimension of expertise: *acquirement of social recognition* is of high importance in professional settings. As will be shown, in this study this dimension boils down to *presentation skills*.
A fifth dimension that has been identified by Van der Heijden (1998) is the dimension of *growth and flexibility*. People who are capable of acquiring more than one area of expertise within adjacent or radically different fields, or who are capable of acquiring a strategy to master a new area of expertise or expert performance in another territory can be termed 'flexperts' (Van der Heijden, 1996). These are people who are both flexible and in possession of expertise. In terms of decision theory, growth can be labeled as *learning.*

The central characteristics of professional competence or expertise in current working life are change-orientation and multi-dimensionality (Tikkanen, 1998). The adaptability and flexibility requirements of an individual employee are based on the continuous changes and transformations taking place at different levels (the individual worker, the job organization and the labor market policy) and on their interaction as well as on global developments, such as that of technology and economy.

To prevent unnecessary elaborations, a short outline has been given here. For the interested reader, we refer to an elaborated outline in the thesis by Van der Heijden (1998) and in the article on the psychometric studies that indicate satisfying reliability and validity of the five-scale instrument (Van der Heijden, 2000). In this contribution we aim to judge the benefit of activities in the assessment process by placing them into the multi-dimensional expertise framework by Van der Heijden, with emphasis on the aspects 1, 2 3 and 4.

## 3 Expectations concerning the weight assessment process from the perspective of expertise

### 3.1 Introduction

Based on the theoretical insights discussed above, some expectations are formulated concerning importance assessment in the context of non-routine problems. We speak of expectations and not of theories because we were not able to assure statistical validity, due to the small number of subjects.
This study concerns individual weight assessment processes at a single point in time, for reasons discussed later. So the first three of Van der Heijden's (1998) dimensions were

the most important. No expectations were formulated concerning social recognition in general. But actors may be expected to consider how the results of their work are received in an organizational environment, even if we do not study the functioning of the actor within organizations. This means that the dimension of social recognition boils down to the preparation of the *presentation* of the results in a way that maximizes recognition of the quality of the importance assessment performed by an actor. The fifth dimension, learning, was not an issue in our study. Apart from the unintentional learning by fulfilling the assignment, we assume learning did not take place.

Before we address expectations concerning Van der Heijden's dimensions of expertise, which pertain to specific elements of the importance assessment process, we formulate an expectation about the structure of the importance assessment process as a whole.

### 3.2 The general structure of the assessment process

As stated earlier, our study concerns laymen who are qualified as decision makers because they possess, amongst other things, general problem-solving skills. It seems logical that actors possessing these skills will use them to devise a solution method for a problem that they have no previous experience with. After all, these skills are all they have readily available to solve the problem. When operating in an organizational context, not explicitly or implicitly devising some sort of structured way to approach the problem does not seem to be an option, for the results are likely to have to be communicated to - and judged by – other actors in the organization. So, problem-solving skills can play an important part in structuring importance assessments. We expect to be able to detect a use of problem-solving skills by identifying a general layout – to be formalized in a model - of the way actors perform importance assessments. Once we succeed in constructing a model of the importance assessment process, we expect phases can be recognized analogous to those found in problem-solving models. In the literature, the problem-solving approach is an established way of describing human mental activities (Ericsson & Simon, 1993; Evans, 1991; Nowell & Simon, 1972; Simon, 1976; Simon, 1979; Sternberg, 1995). It will become clear that the model we have developed in order to test expectation 0 will allow us to set the analytical framework for dealing with subsequent expectations.

Therefore, we formulate the following expectation:

*Expectation 0: Subjects use general problem-solving strategies.*

Indicators for this expectation are:

- the use of a problem-solving approach in which the phases of a general problem-solving model can be recognized. It will become clear that our model of the importance assessment process largely follows these phases;
- the decomposition of the problem in sub-problems. An obvious manifestation would be the decomposition of attributes in sub-attributes.

## 3.3 Knowledge

Van der Heijden (1998) describes as indicators of the knowledge dimension relevant for our study: the ability to apply knowledge to new, unfamiliar problems, having technical knowledge (what we earlier called field awareness) and having an arsenal of problem-solving strategies. The actors we study score low on these indicators. Their reasoning skills may help them solving new problems, but they have only rudimentary field awareness. There is no reason to suppose that actors facing non-routine problems have an arsenal of problem-solving strategies available for it. The actors may have a feeling about the importance of attributes, and may be able to generate arguments so support their judgments, but they don't have a dedicated structure to fit their cognitive activities in.
This lack of a dedicate structure means that the actors are likely not to have a clear idea about the attributes to be weighted. There are several ways the actors can clarify the nature and meaning of the attributes: by formally defining them or by splitting, or decomposing, them into sub-attributes (as already mentioned above). We had no idea whether or how actors would use definitions, but there is literature available on the decomposing of attributes in importance assessment tasks (Weber, Eisenführ & Von Winterfeldt, 1988). Because actors are not expected to have a clear idea of the meaning of the attributes to be weighted, we expect them, if they decomposed attributes, to generate large numbers, and a great variety, of sub-attributes, so as to avoid having to attach specific, narrow meanings to attributes. Now what is 'large'? Research indicates that decisions of actors can to a great extent be predicted accurately using a linear additive model containing only three to four attributes (Van Dam, 1993), although actors generally think that they take many attributes into consideration. By analogy, one would expect that three to four sub-attributes would be sufficient to capture the meaning of an attribute accurately enough to serve as a basis for decisions. If we assume actors use more sub-attributes than necessary, then we feel that in our test we can settle for a three- to fourfold redundancy in sub-attributes as an operationalization of "more than necessary", so we take the round number of ten sub-attributes per attribute.
Therefore, we formulated the following expectations.

*Expectation 1: Subjects take into account at least ten sub-attributes for each of the two attributes to be weighted*

The indicator for testing this expectation is the number of aspects (sub-attributes) that safety and comfort are decomposed into.
Facing non-routine problems, the actors that are the subjects of our study lack the insight in the problem area needed to analyze causal relationships between (sub-)attributes. After all, they have only general field awareness. Establishing causal relationships can be relevant for weighting. It can lead to the use of a common denominator, like money. The relative effects of safety and comfort on a variable related to money, like revenues of the minibus company, could be used as an indicator of the attributes' importance. Establishing causal relations can also be used for eliminating redundant attributes. For example; if braking distance is a sub-attribute of safety and weight determines braking distance, it is useless, even dysfunctional, to consider both weight and braking distance in assessing the safety of a minibus. Finally, insight in causal relationships can be used as a

form of framing the problem, getting a 'mental model' of what 'safety' and 'comfort' actually mean.

*Expectation 2: Subjects will not concern themselves with explicit causal relationships between (sub-)attributes.*

Several indicators of the concern for causal relationships are possible, like the number of segments in the think-aloud protocols pertaining to causal relationships and the explicit use of methods for establishing such relationships, like cognitive mapping. We don't use the first indicator because it says nothing about the number and importance of causal relationships eventually established, but we use the second one. We also use the extent of integration (taking together) of sub-attributes. If sub-attributes A and B are taken together as attribute C, we assume that subjects establish a causal relationship between A and C and B and C, respectively. In the example of the minibus decision: if 'quality of brakes' and 'acceleration' are integrated as 'active safety', then we assume that the quality of the brakes and the acceleration that the minibus can attain determine active safety.

As stated, for expectation 1, we used as an indicator the number of sub-attributes in which an attribute is decomposed. Whereas there is an obvious relationship between an attribute and its sub-attributes, this need not be a causal relationship but may merely be an implication that there is some sort of a relationship, or a specification relationship (the sub-attribute is a specific form of the main attribute). So, we don't use the number of decompositions as an indicator for expectation 2.

Actors possessing only rudimentary field awareness would not have sufficient knowledge (let alone skills) for applying methods and techniques specifically aimed at making importance assessments. Such methods and techniques include: the many methods developed to elicit weights that people attach to attributes (sometimes without themselves knowing it). It would, for example, be possible for actors to imagine for themselves choices between a number of alternatives with varying values on a number of attributes, and then, based on these choices, infer the weights that they use. We do not expect actors dealing with non-routine problems to use such sophisticated methods.

*Expectation 3: Subjects do not use methods and techniques specifically aimed at making importance assessments.*

As indicators we used firstly the explicit mentioning and subsequent use of specific methods, or the use without explicit mentioning, but in that case the use had to have some level of consistency. For example, some weight elicitation methods involve pairwise comparison of (sub-)attributes.
Now if expectation 1 can be accepted and subjects indeed generate many sub-attributes, there are bound to be some incidental pairwise comparisons. But to qualify as a method, there should be consistent pairwise comparison, with a clear conclusion in the end concerning the weight values.

Utility theory states that the attractiveness of an alternative is expressed in the summation of scores on the various attributes, multiplied by their weights (Keeney & Raiffa, 1976). The product of the score and weight of an attribute is called the partial utility of that attribute. The sum of the partial utilities is the total utility (attractiveness) of the alternative. So, utility theory assumes that a person judging an alternative has some concept of (partial) utility that makes various attributes comparable to each other. When utility theory is applied properly, the scores of all attributes are expressed in the same dimension, or on a common dimensionless scale, like a five-point scale. So, all possible scores of each attribute are translated to a score of, for example, 1 to 5. This would mean that weights are also expressed on a common scale. After all, the products of scores and weights all have the same dimension: (partial) utility. In that case, the importance of attributes need not be established by comparing them and deciding whether one attribute is more important than the other. One could just attach an importance value to each attribute individually, just like you can see from the speedometer how fast a car is going without having to compare its speed to that of other cars on the road.

For actors dealing with non-routine problems, absolute weighting has the advantage that no difficult direct comparisons with other attributes are required. The comparison of attribute importance follows from the weights themselves. For example, if safety gets absolute weight 0.4 and comfort 0.6, then comfort is more important than safety even if they have not been directly compared. Of course, comparison could lead to adjustment of the weights, but that is optional. Also, questions could be raised about the validity of weights given without comparison to other attributes, but that need not make the method less attractive for actors facing importance assessment problems that may be difficult for them to solve anyway.

So, if we assume that actors reason according to utility theory, we could formulate the following expectation:

*Expectation 4: Subjects express the importance of each attribute on an absolute scale.*

As an indicator we used the average amount of effort subjects spent on weighting attributes in isolation (‘safety is important’) versus comparing the importance of attributes (‘safety is more important than comfort’). But this is not enough. Actors can weigh attributes in isolation and yet use different scales for each attribute (whether this is sensible or not). They may even use identical or similar scale values like ‘quite important’, ‘really important’ and the like without meaning the same in each case. So, as a second indicator we look at the presence of a common scale, either by a consistent use of scale values (a certain value indeed consistently means the same level of importance) or by the presence of an explicit algorithm according to which absolute weights are converted into relative ones. Such an algorithm is not possible if there isn’t some sort of common scale, or scales convertible into each other.

### 3.4 Skills

Van der Heijden (1998) formulated several indicators of skills that are appropriate for this study. They are: the ability to use specific methods and techniques and to work independently, and the general skill level as perceived by the actor.

One type of skills: general problem-solving skills, has already been covered in expectation 0. If this expectation is accepted, another expectation is logical. If actors adopt a phased problem-solving approach, it would be interesting to study the amount of effort spent on structuring the problem versus the effort spent on solving it. It would be reasonable to assume that actors having limited field awareness and problem-solving skills confronted with a non-routine problem would spend relatively much effort on structuring the problem. Kotovsky & Fallside (1989) indeed provide proof for this. It is difficult to quantify the effort spent on the two activities. Kotovsky & Fallside studied a totally different type of problem (strictly logical problems) and provide no usable quantitative clues. We settled for a simple 50% rule: actors spend at least as much effort on structuring as on solving the problem.

Therefore we expected the following:

> *Expectation 5: Subjects spend at least as much effort on structuring the non-routine assessment problem as they do on solving the problem.*

The indicator is the average percentages of statements pertaining to the structuring phases of our model compared to the solving phases.

Not only at the beginning of the importance assessment process, but also at the end, the characteristics of the actors we studied lead to an expectation. We can expect that actors are well aware of their lack of proficiency in solving a non-routine assessment problem. Therefore, it seems reasonable that they spend quite a lot of effort on evaluating their work. They will likely not readily accept their way of working, or the result of their work, as adequate, but they will try to assess whether they actually got it right. We expect them to spend a significant amount of effort on evaluation. As to what is ‘significant’, we decided on a threshold of the average effort per phase, the exact number to be decided upon after the model of the importance assessment had been constructed.

> *Expectation 5’: Subjects spend an above-average amount of effort on evaluating the quality of their work.*

The indicator we used is the average number of segments in the think-aloud protocols pertaining to evaluation (as we will see, evaluation was identified as a phase of importance assessment).

### 3.5 Meta-cognition and presentation

Van der Heijden's (1998) indicators of metacognition relevant for this study are: capability of judging the level of one's own knowledge and skills in a particular problem situation, capability to judge one's own ability of separating main issues from details, to weigh pros and cons of working methods and techniques, to keep oversight over complex situations and the ability to judge which skills one is missing when confronted with a new problem. We did not measure these indicators but assessed the confidence the subjects had in their own importance judgment. The actors we studied have, as we have seen, little, if any, experience with the task given to them, but they possessed self-evaluation experience. So, the actors can be expected to have adequate cognitive abilities to assess their own performance. This leads us to expectation 6:

> *Expectation 6: The subjects have little confidence in their own judgments (since they are very aware of their lack of knowledge and skills to solve the non-routine problem presented to them).*

But an alternative line of reasoning is possible. Actors involved in acquisition decisions may possess a considerable amount of training in project management, communication and presentation of their work. These characteristics were not assumed beforehand, but they may be there nonetheless. It is possible that they, while realizing their lack of content knowledge and specific importance assessment skills, they are quite confident in their project management skills required to handle the assignment properly, and in their presentation skills, needed to convince others of the quality of their work. This, combined with the not unreasonable notion that importance judgments are to a certain extent a matter of personal opinion anyway, gives rise to the assumption they may well have considerable confidence in the quality of their work. We enter the realm of van der Heijden's fourth dimension: social recognition, which we have narrowed to presentation. So, an alternative to expectation 6 is:

> *Expectation 6': Subjects are confident in their weight judgments (based on the project management, communication and presentation skills they possess).*

The indicators for these two expectations are the confidence in their judgments subjects expressed in the interview after having concluded the assignment, the frequency of preference reversal, the acceptance of expectation 4 as an indicator of the use of project management skills and reflections by the subjects on their performance made during the interview afterwards. Each of our expectations will be investigated in Section 6.

## 4 Research methodology

Only a short outline of our research methodology is given here. For an extensive description we refer to Chapter 2.

**Sample and assignment**

18 undergraduate students of the University of Twente were given an individual assignment based on a fictional case similar to that used in the introduction to this chapter. University students might be assumed to have enough analytical abilities to perform the assignment satisfactorily. At the university, all students receive basic training in general problem-solving techniques. They are specifically indoctrinated in the importance of having a proper problem definition (splitting the problem if necessary), in explicitly choosing a problem-solving method and evaluating the solution. Yet, they don't have enough knowledge and skills in this area that would enable them to rely on previous experience of importance assessments. Hence, the danger that they give weights based on previously obtained knowledge is minimized. This is what we wanted, given that our study is about non-routine decisions. The students had extensive experience with projects concerning the solving of management problems, and insight in the theory of management. So they can be assumed to have some feeling for operating within a simulated organizational context.
Their assignment consisted of supporting the acquisition process of new minibuses by a local company. The subjects were asked to establish the importance of two characteristics of the to-be-acquired minibuses *vis-à-vis* each other, and were asked to imagine that they would be advising the management team during the acquisition process. The attributes, *safety* and *passenger comfort*, were chosen to prevent comparability by some readily available algorithm or heuristic or easy expression in a common denominator such as money. The information that was supplied included a brochure of the company, a leaflet explaining the decision context and two brochures on minibuses; one on a Volkswagen and one on an Opel. The latter enabled the subjects to get familiar with the specific capital good to be acquired. It was made clear that these examples of minibuses did not mean that the subjects had to make a choice between them. A pre-set structure of (sub-)attributes was avoided as much as possible.

The minibus problem was chosen because subjects were assumed to have superficial familiarity with them, so that they could give an importance judgment at all. With, for example, airplanes this would likely have been much more difficult.

**Procedure**

The respondents were asked to think aloud during the assessment process. The general guidelines for think-aloud studies given by Ericsson & Simon (1993) were followed, including a practice session to familiarize the subjects with the think-aloud strategy. All verbal information given by the respondent was recorded and typed out verbatim. After completion of the assignment, a short interview was conducted. In total, each session lasted for a maximum of two hours, for which the subjects were paid 20 Euro. Two pilot sessions were conducted, which led to some minor adjustments of the assignment.
Two kinds of analyses have been performed using the typed out protocols:

1: A largely qualitative analysis according to the general rules of the 'Grounded Theory' approach (Strauss & Corbin, 1998). This was aimed at discovering regularities in the

protocols by inferring general working rules from the think-aloud protocols and in some cases determining the percentage of subjects applying certain rules.
Using this analysis, several phases in the importance assessment process could be distinguished. They will be discussed in the next section;

2: A more quantitative analysis based on a formal coding scheme that was designed on the basis of the qualitative analysis. As mentioned above, certain phases were discovered during the qualitative analysis. In the quantitative analysis the effort put into each phase was assessed, expressed in the number of verbal protocol segments devoted to each phase. Two coders performed the coding activities. Although they worked independently of each other, during the coding of the first six protocols weekly meetings were held to discuss general coding issues in order to enhance its reliability. The coders retrospectively applied the refinements to the coding scheme independently. The overall Cohen's Kappa (Baarda & De Goede, 2001) for inter-rater consistence was .97 over a total number of verbal segments of 7253.

The expectations are largely analyzed in quantitative terms. Mostly, we used average numbers across all subjects in testing the expectations. For expectations pertaining to individual behavior as observed, we use the following acceptance criterion. We accept an expectation if 80% of the subjects (15, rounded off upwards) behaves according to the expectation. For negatively formulated expectations, we reject them if 15 subjects act contrary to the expectation. This is a somewhat arbitrary threshold, but with the small number of subjects involved, this seems a reasonable level of significance. Sometimes some additional criteria are taken into account; where appropriate this will be mentioned. In cases where it was not a case of an indicator either present or not present with individual subjects, but where the value of an indicator could be measured quantitatively, average values were used as criteria.
The threshold of 15 could in theory lead to situations in which an expectation is not accepted but its complement also has to be rejected. This is the case where fewer than 15 but more than 3 (18-15) subjects behave according to the expectation. We had a means of dealing with this, but it will become clear that this situation did not occur.
In the next sections we will go into the results of the think-aloud study

**Limitations**

As outlined in Chapters 1 and 2, this research was conducted with a small group of subjects. These subjects were students with no prior experience with either the acquisition of minibuses or formalized organizational importance assessment processes. The research was conducted in a laboratory context. This means that the results have relatively low statistical and external validity. The first is, in our view, not a great problem. The trends in the results seem to be quite clear and multiple indicators were used for many variables, thereby increasing internal validity. The second point, however, means that no definite conclusions can be drawn for other groups than the group we studied: laymen performing importance assessment away from a real-life context. We can, and do, make propositions about how actors in real-life situations may behave, based on our results and on the literature. But these propositions have yet to be tested in future

research. The basic regularities in importance processes that we describe will, we expect, be present in some form in real-life situations. After all, it is not uncommon for individuals (albeit often with a certain degree of expertise) to make importance assessment processes under circumstances similar to those in our research. So our research provides a basis from which to look at real-life situations. But, again, only further research can establish to what extent and in which form our results describe real-world importance assessment processes.
As explained in Chapter 1, our research pertains only to decision contexts where there is explicit weighting and where the importance assessment process is separated from the evaluation of alternatives.

# 5 Results: the Weight Assessment Model (WAM)

## 5.1 General structure of the WAM

Firstly, the phase model that has been derived from the data is given. In the next section *the way attributes are processed* and *the confidence in the end result* will be discussed.
Our Weight Assessment Model (WAM) consists of 7 main phases and 6 auxiliary activities. The 7 main phases are presented in a sequential way in table 1. In reality, subjects often go back and forth between phases and most often address phases more than once. The auxiliary activities pertain to areas like information search and planning. We do not deal with them in the context of this chapter.

The WAM follows the general format of phased problem-solving models (Ericsson & Simon, 1993; Evans; 1991; Hicks, 1995; Nowell & Simon, 1972; Siomon, 1976; Simon, 1979; Sternberg, 1995): problem structuring, often called problem identification (phase 1), problem analysis (phase 2), problem solution (phases 3 to 6) and evaluation (phase 7). In many models an implementation phase is also found, but as our problem had no implementation component this phase is not relevant in this study.

From the table, it is clear that most phases could be observed with every subject. Exceptions were phases 4 and 5. In each phase subjects could be found who had either 5% or more of their total number of statements, including auxiliary activities, devoted to that phase, or at least 30 statements. Thus we conclude that it is justified to distinguish these phases and that they are not just isolated coding incidents.

Let us now discuss briefly the different phases of the model as shown in Table 1.

| Phase | Phase name | **% of segments devoted to the phase** | **% of subjects with which elements of phase were observed** |
|---|---|---|---|
| 1 | Problem identification | 6.74 % | 100 % |
| 2 | (Sub-) attribute processing | 30.33 % | 100 % |
| 3 | Absolute sub-attribute weighting | 27.22 % | 100 % |
| 4 | Homogeneous sub-attribute weighting | 4.53 % | 66.7 % |
| 5 | Heterogeneous sub-attribute weighting | 1.50 % | 55.6 % |
| 6 | Attribute weighting | 12.54 % | 100% |
| 7 | Evaluation | 17.14 % | 100% |

*Table 1. The phases of the WAM*

## 5.2 Phases of the WAM

- ***Phase 1: Problem identification***

  This phase consists of activities like reading the assignment and, if desired, re-formulating it in one's own words. When re-formulating aspects of the assignment, subjects may also mention things they are *not* to do, for example weight attributes concerned with costs. Essentially, this phase concerns defining, so to speak, the task lying ahead.

- ***Phase 2: (Sub-) attribute processing***

  We find aspects of problem analysis in phase 2. Not only the elements constituting the problem are defined (elaboration on the character on the character of the attributes to be weighted), but also the possible, not necessarily causal, relationships between these elements.

  If one wants to weight attributes, one should first know what one is weighting. Attribute processing concerns giving the attributes a more precise meaning. This can be seen as a case of framing (Akin, 1994).

  Attribute properties like measuring level, measuring unit, level of abstractness and precision can change as a result of processing. The following forms of processing were identified:

  a: Decomposing. An attribute can be split up in several sub-attributes. For example: 'safety' can be split up in 'braking distance', 'acceleration', 'strength of the structure', 'presence of seatbelts' and so forth. Each of these sub-attributes can, of course, decomposed further;

  b: Re-formulating. When an actor gives an attribute or sub-attribute a different name while meaning the same attribute with a similar, not necessarily identical, measurement unit, the attribute is re-formulated. For example, 'comfort' may be

re-formulated as ‘passenger comfort’, ’comfort of the passengers’ or ‘traveling comfort’;

c: Concretizing an attribute or sub-attribute. For example ‘the minibus should be roomy enough’ might be made concrete, as ‘people should not have to sit with their suitcases on their laps’;

d: Integrating (sub-)attributes into a new ‘sub-attribute. For example: after having named sub-attributes of safety like ‘braking distance’, ‘acceleration’, ‘strength of the structure’ and ‘presence of seatbelts’, the first two could be taken together (integrated) into ‘active safety’ and the last two could be integrated into ‘passive safety’;

e: Making an attribute more abstract. This is the complement of concretization.

The next phases concern the actual weighting process.

- ***Phase 3: Absolute (sub)-attribute weighting***

With ‘absolute’ weighting (Timmermans, 1993), we mean that a statement about the importance of a (sub)-attribute is made without reference to the importance of other (sub)-attributes. For example, if an actor says: “safety is important”, or “having seatbelts in the minibus is absolutely essential”, it is not clear whether safety is more important than comfort, or whether having seatbelts is more essential than having an anti-skid system.

In phase three, as well as in the subsequent phases, subjects not only assign weights, but they also give the reason for giving these weights. For example, a subject may say: “I think a good anti-skid system is important (weight assignment), because (argument) it reduces the chances of getting involved in an accident”. It is possible to decide that some sub-attributes defined earlier, are to be excluded from the weighting activities in subsequent phases. So, this phase, like phase two, can serve to give the subject a clearer view of the (sub)-attributes to be weighted.

- ***Phase 4: Homogeneous sub-attribute weighting***

This phase is the first in which ‘true’ weighting takes place; i.e. the balancing of the weight of one sub-attribute against that of another. We call this ‘relative weighting’ (Timmermans, 1993). In this phase, two or more sub-attributes of the same main attribute are weighted against each other, and arguments for the weighting are given. For example, the presence of seatbelts, the quality of an anti-skid system and the strength of the structure (all sub-attributes of safety) may be weighted against each other.

- ***Phase 5: Heterogeneous sub-attribute weighting***

  This phase differs in only one respect from the previous one, i.e. the sub-attributes that are weighted belong to different main attributes. For example, the quality of the anti-skid system (a sub-attribute of safety) might be weighted against the quality of the seats (a sub-attribute of comfort).

- ***Phase 6: Attribute weighting***

  This phase concerned the integral weighting of the (two) main attributes. This was the original assignment. Only 13 of the 18 subjects reached this phase. 5 stopped after having completed phase 5. Some of these subjects consistently judged sub-attributes of safety to be more important than sub-attributes of comfort, or the other way around. In that case, a judgment of the main attributes can be readily inferred. But this does not lead to the conclusion that subjects actually completed phase 6, for they may not have made the inference we made. Some of the subjects who stopped after phase 5 made some statements pertaining to phase 6. For example: at the beginning of the assignment they would remark that, at first sight, safety seemed more important to them than comfort. Hence phase 6 was considered as being addressed (even with 100%), though sometimes with no explicit weight judgment as the conclusion of the assignment.

- ***Phase 7: Evaluation***

  This phase comprises the reflections by subjects on their activities and the results. Several types of evaluation were identified, like the extent to which the assignment had been fulfilled, evaluations of weights judgments (was the subject, on hindsight, happy with assigned weights) and evaluations of arguments (how good were the reasons for particular weight judgments).

## 5.3 Use of general problem solving strategies.

Looking at Table 1,

*Expectation 0: Subjects use general problem-solving strategies.*

can be easily accepted.

All subjects went through the phases of problem structuring, analysis, solution and evaluation, even though some solution phases (in particular homogeneous and heterogeneous sub-attribute weighting) were sometimes rarely used. All subjects split the problem into sub-problem by decomposing the main attributes, as stated above.

Furthermore, phase 2 could clearly be identified as the phase that subjects enter immediately after they have read the assignment (which is part of phase 1). One can

easily detect the logic behind this; before one can attach a weight to a (sub)-attribute (phases 3 to 6), one has to identify it and process the different sub-attributes. Usually, phase 3 precedes phases 4 to 6. If subjects assign weights to attributes on an absolute importance scale (expectation 4), then phases 4 and 5 could well be unnecessary, for the subjects could immediately proceed to phase 6. But phases 4 and 5 could be used as checks and refinements on the absolute weights given in phase 3. In that case, the sequence of phase 3, 4 and 5 is also logical. Moreover, the assignment was to compare the importance of safety and comfort. If sub-attributes of safety can be weighted against sub-attributes of comfort (phase 5) then it does not add much to the fulfillment of the assignment to weigh those sub-attributes against other sub-attributes of the same attribute (phase 4). Likewise, if safety and comfort are already weighed (phase 6), the assignment is fulfilled and there is no need for weighting sub-attributes (phases 4 and 5). Phase 6 concerned the desired result of the assignment (weighting safety against comfort), so it is logical that it is the last weighting phase, followed only by evaluation.

This concludes the description of the different phases of the decision making process. In the next section, the other expectations we formulated earlier are tested.

## 6 Results: The relevance of the WAM in the light of expert behavior: testing of the expectations

*Expectation 1: Subjects take into account at least ten sub-attributes for each of the two attributes to be weighted*

This expectation is accepted. As mentioned earlier, attributes could be processed in one of the following ways: decomposing, concretizing, re-formulating, abstracting and integrating. The way of processing of attributes has a profound effect on subsequent phases of the weight assessment process. The result of the processing phase can be seen as an intermediate product, i.e. the attributes are formulated in such a way that they can, ideally, be given weights easily, unambiguously and with a clear motivation.

Table 2 shows the number of sub-attributes for ‘safety’ at various levels of decomposition. A level is defined as the number of splits that resulted in a certain sub-attribute. So, if ‘comfort’ is split in a number of sub-attributes, amongst which ‘quality of the seats’, which in turn is split in ‘width of the seats’ and ‘height of the armrests’, then there are two levels of decomposition. For ‘comfort’, the numbers were roughly similar to those of ‘safety’. It can be seen that the attribute is decomposed in a large number of sub-attributes. The average number of sub-attributes per subject for ‘safety’ was 19,6. For comfort the number was 24,4. 15 subjects generated 10 or more sub-attributes for safety. All subjects generated 10 or more sub-attributes for comfort. So, we accept expectation 1.

| Number of attributes as a result of decomposition | Number (%) of subjects, first level | Number (%) of subjects, second level | Number (%) of subjects, third level |
|---|---|---|---|
| 0 | | 4 (22) | 15 (83) |
| 1-5 | 2 (11) | 7 (39) | 1 (6) |
| 6-10 | 5 (28) | 4 (22) | 1 (6) |
| 11-15 | 5 (28) | 2 (11) | 1 (6) |
| 16-20 | 5 (28) | 1 (6) | |
| 21-25 | | | |
| 26-30 | | | |
| 30-35 | 1 (6) | | |

*Table 2. The decomposition of 'safety'*

What did the subjects do with all these sub-attributes? One could think that the subjects wanted to use the sub-attributes in order to discover common denominators, for example by identifying sub-attributes that safety and comfort have in common and that hence might be eliminated or given equal weights. Yet, no evidence supports this notion.

*Expectation 2: Subjects will not concern themselves with explicit causal relationships between (sub-) attributes.*

This expectation is accepted. No systematic methods for finding causal relationships were used, like cognitive mapping. Not a single subject tried to define or frame the two attributes to be weighted (safety and comfort) in order to express them into a common denominator, for example money or another explicitly mentioned common scale on which the attributes could be scored. Instead, they decomposed the attributes in a large number of sub-attributes, as already discussed.

When a systematic procedure could be inferred for weighting some sub-attributes (for example, the distinction between active and passive safety), this was occasional and not maintained with respect to all relevant sub-attributes. What is more, there was hardly any integration of sub-attributes, something that one expects when weights are to be given not to sub-attributes but to main attributes. This is in line with expectation 2, in the sense that for integration the (causal) relationships between (sub-) attributes need to be known. Indeed, relationships between attributes were addressed only incidentally. If they were discussed, they seldom had consequences for the weighting. Integration was a rare occasion, as can be seen from Table 3. This table shows the number of (sub-) attributes integrated. In the acceptance of expectation 2 we classify a subject as a non-incidental user of integration if integration occurs in at least 4 cases over safety and comfort together. As we see that fewer than 15 subjects integrate at all, expectation 2 is rejected as far as this indicator is concerned.

| Number of sub-attributes being integrated | Number (%) of subjects integrating sub-attributes of safety | Number (%) of subjects integrating sub-attributes of comfort |
|---|---|---|
| 0 | 9 (50) | 13 (72) |
| 1-5 | 6 (33) | 5 (28) |
| 6-10 | 3 (17) | |

*Table3: Integration of safety and comfort*

Another way to look at the significance of integration is to observe how many (sub-) attributes are the result of integration. The maximum number of attributes that were the results of integration was 4 (1 subject). 2 integrated attributes were found with only 3 subjects. Only in two instances was a sub-attribute resulting from integration given a weight during the final weight assignment. Integration always resulted in a new sub-attribute, not in the main attributes to be weighted according to the assignment (safety and comfort). The logic of the integration was often implicit and nearly always purely qualitative. No indexing or other quantitative methods were used. In sum, integration was by and large irrelevant.
It can be seen that all subjects ended phase 2 with a large number of sub-attributes, with no system to guarantee that the main attributes have been adequately covered, or that assessing the weights of the sub-attributes would in any way be easier than assessing the weights of the main attributes. This is also in line with expectation 2. The relationships between the sub-attributes and main attributes were unclear, so that weights assigned to sub-attributes did not seem to say anything about the weights of the main attributes (see Chapter 3). It could be assumed that subjects used the processing of attributes mainly for framing purposes, i.e. to find out what ‘safety’ and ‘comfort’ actually mean. But since all subjects devoted considerable effort to phase 3 (absolute weighting) the processing of sub-attributes appeared to mean more than just a framing function. The subjects obviously were not only interested in obtaining concepts of ‘safety’ and ‘comfort’. The sub-attributes were important in their own right.

*Expectation 3: Subjects do not use methods and techniques specifically aimed at making importance assessments.*

Expectation 3 is accepted: no more than three subjects violate this negatively formulated expectation. Only a small minority (3 subjects) used some form of pairwise comparison with some consistency. 8 subjects used a method to progressively refine the range of weights taken into consideration, either starting with extreme weights and working towards the middle or starting with equal weights and working towards extremes. This could be classified as a method well suitable for weighting, but since it can equally well be used for, for example, estimation of attribute scores or other properties, we did not qualify it as a specific weighting method.

*Expectation 4: Subjects express the importance of each attribute on an absolute scale.*

Although the results point strongly at the possibility of subjects using absolute scales of importance, they do not allow us to confirm this expectation. Phases 4 and 5 (relative sub-attribute weighting are insignificant compared to phase 3 (absolute weighting). This gives credence to the notion that phases 4 and 5 are merely used to check the results of phase 3. If phases 4 and 5 would be steps in establishing weights of sub-attributes instead of a mere check, more effort would surely be devoted to these phases. On the other hand, if subjects would use phases 4 and 5 as checks for phase 3, we would expect them to frequently hop between phases 3 and 4 or 3 and 5, probably with evaluation (phase 7) in between. Although the protocols were not analyzed in this respect, a cursory look at the graphical representations of phase changes provided no indications that this was the case to any significant extent.

Another observation casting doubt on the use of an absolute scale is that phase 6 takes 12.5% of the effort. While this is less than phase 3 (27%), it represents much more than an incidental activity. The effort per attribute reinforces this notion: only two attributes were weighted in phase 6, while twenty or more attributes were frequently weighted in phase 3. The insignificance of phases 4 and 5 relative to phase 6 may be explained by the fact that the assignment stated quite clearly that safety and comfort had to be weighted. So, weighing sub-attributes was not enough.

Finally, there are no reliable indications that subjects used absolute scales to any extent. 13 of the 18 subjects gave ordinal weights as the end result of the assignment. Even after the experimenters asked the subjects to give quantitative (interval or ratio) weights, 6 subjects persisted in giving ordinal weights. Examples of ordinal weights often used are: ‘quite important’, ‘really important’, ‘not very important’ and the like. It is impossible to say whether these expressions were part of a scale in the sense that their value relative top each other had any consistency. Some subjects used explicit scales, but they were the exception rather than the rule. Only three subjects used scales that could be assumed to be absolute if used consistently (which was impossible to determine) to any significant extent. And even then, these scales were usually used at the sub-attribute level and only in one case did the subject (when asked during the interview afterwards) link the weights of the sub-attributes directly to the (ordinal) weights of safety and comfort.
The notion of subjects constructing an absolute scale of importance is attractive, but cannot be accepted, nor clearly rejected, by our results. This is an area for further research. Questions present themselves like ‘is there one scale for all attributes or are there several scales, for example one for each set of sub-attributes pertaining to the same main attribute’? The endowment effect (Kahnemann, 1994; Kahnemann, Deutsch & Thaler, 1990) seems to point at the possibility of more than one scale. The price a person is willing to pay for a good is sometimes lower than the sum he wants to receive before giving it up. Both scores and weights could play a role here, but the example shows that the question whether people have an absolute importance scale available is far from answered. But the significance of phase 3 is obvious in our results. The relationship between the phases of the WAM is further discussed in Chapter 4.

There is another reason for the comparative rarity of relative weighting that should be considered: the conscious or unconscious desire to reduce complexity. With absolute

weighting of n attributes, there ate n weights to establish. With weighting of all possible pairs of attributes against each other (the most efficient form of relative weighting) the number of weights becomes n(n-1)/2. So, the weighting becomes much more complex, and even more so when inconsistencies in the weighting have to be addressed (A is more important than B, B is more important than C, C is more important than A). As the complexity of the importance scales do not differ in principle between absolute and relative weighting, absolute weighting yields a reduction in complexity. This explanation cannot be proven, but subjects certainly employed various other means of complexity reduction (see Chapter 5).

*Expectation 5:* *Subjects spend at least as much effort on structuring the non-routine assessment problem as they do on solving the problem.*

Table 1 shows that, contrary to our expectation, 37.03% of the average effort was devoted to the structuring phases (1 and 2) and 45.79% to the weighting phases. Only 2 of the subjects devoted more time to the structuring phases than to the weighting phases. We have no clear explanation for this. It is possible that experts would devote even less time to the structuring phase. If this were the case, our threshold for accepting expectation 5 may be too high. But for the moment it is rejected.

*Expectation 5':* *Subjects spend an above-average amount of effort on evaluating the quality of their work.*

This expectation is confirmed (see Table 1). The model encompassed seven phases, and 15% is one seventh of the segments, rounded off upwards. So, the threshold for accepting the expectation is 15%. The averaged effort devoted to evaluation is more than this. It is impossible to say whether the reason is indeed, as we suspect, the lack of expertise of the subjects in weighting the attributes under consideration, but the notion seems credible.

**Confidence in the end result**

*Expectation 6:* *The subjects have little confidence in their own judgments (since they are very aware of their lack of knowledge and skills to solve the non-routine problem presented to them).*

This expectation is rejected. Although some subjects remarked that assigning weights is to a certain extent an arbitrary process, all subjects showed signs of having confidence in the weights eventually assigned. Some empirical evidence of this is summarized below:

- *Robustness*: After completing the assignment, we asked the subjects why they assigned the particular weights they did and not slightly different weights (for example 0.6 for safety instead of 0.7). All subjects remained with their original weights, although four acknowledged that slightly different weights would also have been an option. Of course it is possible that the subjects wanted to stick to their

original weights because they had invested considerable effort into performing the assignment, but it was pointed out to them explicitly several times that there were no 'right' or 'wrong' answers; By this, we hoped to at least partly negate the risk of 'social desirable' answers by removing the impression that it was proper to seem confident in one's own weights. We could not provide an incentive for accuracy, since there were no objectively identifiable 'right' solutions (weight judgments).

- *Consistent preference evaluation*: Only one subject of the 13 who in the end weighted safety and comfort explicitly as overall concepts (and not only their sub-attributes) showed preference reversal between the first time they made a statement about the relative importance of safety and comfort and the final weighting. So, most subjects who stated at the start of the assignment that safety was more important than comfort would likely stick to this for the remainder of the assignment, although the exact weights might change. This is all the more remarkable because 4 of these subjects changed preferences during the assignment, but in the end came back to their original order of preference.

So, the subjects seemed to have considerable confidence in their own weight assignments. . This is not a sign of lightheartedness as most of them showed expressions indicating that at certain times they experienced difficulty in coping with the assignment. Also, except for one subject, there are no indications that any of the subjects failed to do their utmost to perform the assignment as good as possible. The one exception had such clear-cut ideas about the importance of safety and comfort that he felt he did not have to elaborate on the assignment.

All in all, while the subjects were not experts in the field of minibuses nor did they have experience in performing importance assessments like this, they had confidence in their own work. This is contradictory to expectation 6, based on the expertise framework by Van der Heijden (1998, 2000). Looking at the dimensions distinguished by Van der Heijden, having self-confidence with regard to a certain task is an important aspect falling in the category of meta-cognitive knowledge. Having insight into one's strengths and weaknesses enables a person to evaluate his or her own decision process. And our subjects could rely only to a limited extent to content knowledge (a laymen's knowledge on minibuses) and even less on meta-cognitive knowledge.
So, these subjects may not have been capable to assess the validity of their own judgments, i.e. in this case they are subject to the leniency effect (Carcio, 1991). They are inclined to give a more rosy image, i.e. more faithful judgment of their own decision making process. Besides, our subjects did not need to be concerned with the approval of bosses or colleagues, so their self-confidence was not hindered by fear for a lack of social recognition. In real-life professional settings, which the model by Van der Heijden relates to, gaining respect and approval by others is important.
So, the confidence of the subjects might very likely have been lower in real professional settings.

The above explanation of the results is based on the negative relationship between the capabilities for metacognition and confidence assumed in expectation assumed in

expectation 6. But it is possible that confidence in the end result is high despite a high level of metacognition. Such an alternative explanation is given in expectation 6':

> *Expectation 6: Subjects are confident in their weight judgments (based on the project management, communication and presentation skills they possess).*

This expectation is confirmed in the sense that the indicators discussed above were present. However, alternative explanations cannot be ruled out. We already saw that a general problem-solving strategy could be recognized in the subjects' work. In Chapter 5 we argue that complexity-reduction strategies (which can be seen as a type of project management strategies) were employed, but in the interviews afterwards subjects did not indicate that these strategies had increased or assured their confidence in the end result. They could have felt this way, but it cannot be proven. There were indications that subjects realized the communicative aspects of the assignments: 3 stated that their own personal opinion didn't matter but that they had to look at the problem from the perspective of the company. All subjects took perspectives of interested parties other than themselves into account (clients, the management of the company, the drivers of the minibuses). Some of them indicated explicitly that they felt that they could justify their importance judgment to the management of the minibus company. But with respect to this expectation, the relationships with project management, communication and presentation skills remain somewhat speculative.

# 7 Conclusions and recommendations for further research

## 7.1 Behavior of laymen

Based on Simons (1979) general problem-solving model, we developed a phase model for representing the way individual actors assess the importance of attributes of capital goods in non-routine situations. The model is consistent with the problem space for weight assessment problems devised on the basis of Simons model and on the elements of a generic linear additive utility function. In this chapter, we addressed some salient observations done during the think-aloud study.

The results of our think-aloud study show, not surprisingly, that apart from the WAM phase structure, the subjects did not assess weights in a systematic way. They did not in any explicit way try to relate the various attributes to each other, either by developing common scales to score the attributes or by seeking a common denominator, like money. They hardly integrated the many sub-attributes they generated. Neither did subjects use a system in decomposing 'safety' and 'comfort' to uncover incompleteness and redundancies that also might have served as a basis for subsequent integration. Furthermore, they paid very little attention to weighting the sub-attributes against each other (phases 4 and 5 of the Weight Assessment Model). They did explore the problem broadly (expectation 1 accepted), did not concern themselves with causal relationships

(expectation 2 accepted) and used elements of general but not clearly of specific problem-solving methods (expectations 0 and 3 accepted). They emphasized problem structuring less than expected (expectation 5 rejected). Altogether, this leaves us with the impression that non-routine decision problems are tackled using an intuitive approach. All the same, they had confidence in the end result (expectation 6 rejected) but this could not be attributed unambiguously to their project management, communication and presentation skills (even though expectation 6' is confirmed). Considerable effort was devoted to evaluation (expectation 5' accepted).

Our results are quite unambiguous in terms of internal validity. The value of Cohen's Kappa is high. Furthermore, the assignment was, while difficult, not beyond the subjects' ability and all subjects seem to have done their best to perform the assignment as well as they could. The reason for this assessment is that no subject took the easy way out in case of lack of confidence by stating that safety and comfort were equally important. One subject arrived at equal weights but he had a good reason; he combined the weights that he would assign personally with the weight he thought that clients would give.

However, a further discussion of the external validity and the scope of generalization is necessary. The question is to which degree the setting of the study and the subjects were truly representative of real life decision situations.

## 7.2 External validity

Real-life decision processes are much more complex than those studied in our study. For example, it is to be expected that actors will make weight assessments not only before but also during the choice of alternatives, if they use weights at all for making a choice. Also, it is likely that there are more than two attributes, and contextual aspects like the preferences of the management have to be taken into account by the actors involved. A major shortcoming in many expertise studies is the fact that it is debatable whether the standardized laboratory tasks that are mainly used in expertise research capture the real-life problems that people encounter in work settings. This means that there are some validity problems inherent to the choice of the performance tasks that have been studied. Although we have tried to counter the problem of previous experience, which would give some subjects an advantage compared with others by taking non-experienced students who did not have the expert's domain-specific knowledge, the situation lacks 'real-life character'.

In order to make a start on addressing these problems, a study should be attempted of the phenomenon of professional expertise in its entirety and in different real-life settings. One could consider studying the process of importance judgments among more or less experienced managers in working organizations.

Another limitation is that students were used as subjects. They had no experience in making weight assessments for the acquisition of capital goods, except perhaps goods for their own personal use. In organizations, major decisions are often taken by groups of

actors who communicate their importance judgments and their preferences for certain alternatives to each other, and who hopefully in the end arrive at a decision that everyone can live with. But actors bring their individual judgments and preferences to the table, perhaps having entertained them long before the start of the formal decision process. It is on these individual actors that we focus. This is partly a matter of prudence. Our priority is studying the reasoning processes of actors. If we would study groups of actors, it would be practically impossible to observe individual actors closely enough to assess their reasoning processes. Furthermore, it would be hard to distinguish individual reasoning processes from group dynamic processes. Our research method is well suited for studying reasoning processes of individual actors in detail, but unsuited for studying group processes. Once a model for representing individual reasoning processes is available, group decision processes may be studied using it.

The last limitation is that we do not deal with boundary conditions that the organizational context may impose on the (perceived) freedom of an actor in his or her importance assessment process. For example, the opinion of one's superiors may have a profound influence on one's importance judgment. In this study, the actors faced no pre-set limits on their freedom of reasoning. This may be an issue for future research.

The limitations that are mentioned previously were deliberately accepted and can also be turned into strengths. What we have tried to do is to establish a baseline case: how do actors make weight assessments when they are not influenced by knowledge about the available alternatives, by an organizational context or by past weight assessments which might serve as a basis for undue rationalization? The students that formed the sample had the potential to, and perhaps will one day, make strategic acquisition decisions for organizations. They were not very familiar with minibuses, but had seen them, sometimes driven in them, and generally were familiar enough with them to make a weight assessment concerning two attributes that they should be able to comprehend to a certain extent. Based on the interviews afterwards, we have reasons to believe that we succeeded in this respect.
It is by no means certain that the constraints of this study impair the applicability of our findings to real-life decision processes. Having thought about the weights and hence having a frame of reference to work with, actors may adjust their judgments in later stages of the decision process, for example when they actually have to choose between alternatives or when they have to communicate with other actors. This seems especially likely with non-routine decisions, the kind of decisions this study is about. So, the importance assessments made during the preliminary stage of decision processes may well be similar to those observed in our study. Also, in some elicitation methods, actors are asked to give weights at an ordinal level (is attribute A more or less important than attribute B). The weight judgments the subjects made in our study would likely be valid enough to be used in these elicitation methods. And we should remember that, contrary to the solving of logical, well-structured problems, there are no 'right' or 'wrong' weight judgments. However, how actors behave in practice remains speculation until real-life importance assessment processes are studied.

Now that we know in broad terms how people make importance assessments when they are influenced as little as possible by factors like prior experience and the organizational context, in future research these factors can be progressively introduced in order to study their influence on the weight assessment process. We conclude this chapter by stating some propositions concerning the behavior of experts performing importance assessments and by suggesting some issues for future research.

## 7.3 Behavior of experts

The results of our study can be interpreted in the light of research on differences on the behavior of layman and experts. As noted earlier, no conclusions about the behavior of experts can be drawn, as we did not compare laymen with experts. But relating our findings to the literature on expertise, some propositions can be made that can be tested in future research.
Important characteristics of experts are that they have a large amount of content knowledge and that they are capable of representing a problem in the abstract terms needed for the application of general solution algorithms or heuristics (Chi, Glaser & Farr, 1988). Taking these characteristics into account, in our research we found that our subjects displayed the behavior of laymen in the following respects. They split attributes in large numbers of sub-attributes (having no clear idea of the meaning of the attributes; expectation 1). They pay no attention to causal relationships between (sub-)attributes (expectation 2) and devote considerable effort to evaluation (expectation 5')..We therefore propose experts, even in non-routine decision-making:

- to use ready-made or self-developed formal definitions of attributes, aimed at finding characteristics for which a weighting algorithm or heuristic can be applied. This may include an absolute importance scale;

- when decomposing attributes and intending to assign weights to the resulting sub-attributes, to pay attention to the system of decomposition so as to make it complete but not redundant (regardless of how successful the expert is in this respect);

- to integrate sub-attributes to the maximum extent so as to reduce the number of (sub-)attributes to be weighted and to find common denominators for as many (sub-)attributes as possible;

- to devote only a limited amount of effort to evaluation.

If future research proves these propositions correct, the performance of laymen may be improved by stimulating expert-like behavior. The first two skills mentioned above can certainly be developed. Concepts like mental models or cognitive maps seem suitable for this. So, although these instruments have yet to be adapted for developing importance assessment skills, our research so far shows that the importance assessment process can be analyzed, understood and possibly improved (by the use of tools like cognitive mapping).

## 7.4 Recommendations for further research

We would like to suggest the following areas of research. Firstly, research on importance assessment by experts. Are they more rational, systematic or consistent than laymen? Do they go through the same phases in much the same way, but only using more content knowledge, or is their approach fundamentally different? For example, do they attempt to find a common denominator for comparing attributes? Another interesting question is whether experts are better in making importance judgments outside their own expertise area. Some authors have suggested that expertise is confined to a certain area; experts are not better than laymen in solving problems outside their own area of expertise. By comparing the way experts perform weight assessments both within and outside their area of expertise we can find out whether this limitation on expertise also holds for a higher-order problem-solving skill like performing weight assessments. This research should concern a range of weighting problems (minibuses, organizational strategies, hiring of personnel), with various numbers of attributes. Using the coding scheme and analytical tools that we developed and which are available to others on demand could do part of this research. It could provide a basis for statistically valid quantitative analysis, which was not possible with the limited number of subjects in our study. It might even be possible to set up a database with think-aloud protocols for this purpose.

Secondly, one could conduct research as a basis for developing and testing instruments aimed at helping actors who have to make weight assessments. This could be based on research on expert behavior concerning the ideas formulated in Section 7.3. For example by stimulating them to decompose attributes systematically or take (causal) relationships between attributes into consideration. An important pitfall here is that the effectiveness of these instruments is difficult to measure. A decision-maker may not know for years whether he made the right decision. For lack of comparison, he may never know. Indicators of the effectiveness of weight assessments may be the perceived quality of the assessment by a panel of experts, or the converging of experts' weight assessments after they have been taught to use the instruments, compared to the situation before the use of the instruments was taught.

Finally, research into other areas of management should be conducted. Management science is a multidisciplinary field. Heuristics have been developed for integrating information from several disciplines into a decision on, for example, the strategy of a company (see for an overview Johnson & Scholes, 1999). But what goes on in the head of managers? Do they really follow these normative heuristics? Insofar as integrating information from various disciplines means assessing the weight of factors to be taken into consideration, our model may be used for analyzing the reasoning of managers in different fields.

**Literature**


Akin, Ó. (1994). Creativity in design. *Performance Improvement Quarterly, volume 7*, 9-23.

Alexander, P.A., Schallert, D.L., & Hare, V.C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research, 61*(3), 315-344.

Baarda, D.B. & de Goede, M.P.M. (2001). *Basisboek methoden en technieken*. Groningen: Stenfert Kroese.

Bereiter, C., & Scardamalia, M. (1993). *Surpassing ourselves. An inquiry into the nature and implications of expertise*. Chicago: Open Court.

Bettman, J.R, Luce, M.F. & Payne, J.M. (1998). Constructive consumer choice processes. *Journal of consumer research*. Volume 25, 187-217

Camerer, C.F., & Johnson, E.J. (1991). The process-performance paradox in expert judgment. How can experts know so much and predict so badly? In: Ericsson, K.A., & J. Smith (Eds.), *Toward a general theory of expertise. Prospects and limits.* Cambridge: Cambridge University Press.

Cascio, W.F. (1991). *Applied psychology in personnel management.* London: Prentice-Hall International.

Chi, M.T.H., Glaser, R., & Farr, M.J. (Eds.) (1988). *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Curtis, B. (1986). By the way, did anyone study any real programmers? In: E. Soloway & S. Iyengar (Eds.), *Empirical studies of programmers.* Norwood: New Jersey: Ablex.

Dam, K. van (1993). Beoordelingsprocessen: structuur- versus procesgericht onderzoek. In: Koele, P. & van der Pligt, J. (Ed) *Beslissen en beoordelen*. Meppel: Boom.

Ericsson, K.A. (1996). *The road to excellence. The acquisition of expert performance in the arts and sciences, sports, and games*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Ericsson, K.A., & Lehmann, A.C. (1996). Expert and exceptional performance: Evidence on maximal adaptations on task constraints. *Annual Review of Psychology, 47,* 273-305.

Ericsson, K.A., & Simon, H.A. (1993). *Protocol analysis: verbal reports as data* (2nd ed.). Cambridge, MA [etc.]: Bradford Books/MIT Press.

Ericsson, K.A., & Smith, J. (1991). *Toward a general theory of expertise. Prospects and limits*. Cambridge: Cambridge University Press.

Evans, J.R. (1991). *Creative thinking in the decision and management sciences.* Cincinnati: South-Western Publishing Co.

Glaser, R. (1987). Thoughts on expertise. In: C. Schooler & W. Schaie (Eds.), *Cognitive functioning and social structure over the life course* (pp. 81-94). Norwood, NJ: Ablex.

Hicks, M.J. (1995). *Problem solving in business and management; hard, soft and creative approaches.* London [etc.]: International Thomson Business Press.

Johnson, E.J. (1988). Expertise in admissions judgment. Unpublished doctoral dissertation. Carnegie-Mellon University.

Johnson, G, & Scholes, K. (1999). *Exploring corporate strategy, fifth edition*. London [etc]: Prentice Hall.

Kahneman, D, Knetch, J.L. & Thaler, R,H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of political economy*, volume 96, 1325-1348.

Kahneman, D. (1994) New challenges to rationality assumption, *Journal of institutional and theoretical economics*, volume 150, 18-36.

Keeney, R.L., & Raiffa, H. (1976). *Decisions with multiple objectives, preferences and value tradeoffs.* New York [etc.]: John Wiley & Sons.

Kotovsky, K. & Fallside, D. (1989). Representation and transfer in problem solving. In: Klaad, D. & Kotovsky, K. (ed). *Complex information processing; the impact of Herbert A. Simon*. Hillsdald, NJ, 69-108.

Logan, G.D. (1985). Skill, and automaticity: relations, implications and future directions. *Canadian Journal of Psychology, 39*, 367-386.

McLagan, P.A. (1997). Competencies: the next generation. *Training & Development, 51*(5), 40-47.

Newell, A., & Simon, H.A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Patel, V.L., & Groen, G.J. (1991). The general and specific nature of medical expertise: a critical look. In: Ericsson, K.A., & J. Smith (Eds.), *Toward a general theory of expertise. Prospects and limits.* Cambridge: Cambridge University Press.

Simon, H.A (1979). Information Processing Theory of Human Problem Solving. In: Estes, W (ed.) *Handbook of Learning and Cognitive Processes* (Vol. 5). Hillsdale: Lawrence Erlbaum Associates Inc.

Simon, H.A. (1976). *Administrative behaviour.* New York: The Free Press, London: Collier McMillan Publishing.

Sternberg, R.J. (1995) Expertise in complex problem solving. A comparison of alternative conceptions. In: P.A. Frensch & J. Funke (ed.) *Complex problem solving. The European perspective* (pp. 295-321). Hillsdale NJ: Lawrence Erlbaum Associates, Publishers.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research; Grounded Theory, procedures and techniques.* Thousand Oaks CA [etc.]: Sage Publications.

Tikkanen, T. (1998). *Learning and education of older workers: Lifelong learning at the margin.* Jyväskylä Studies in Education, Psychology and Social Research 137. University of Jyväskylä.

Timmermans, D. (1993) The impact of task complexity on information use in multi-attribute decision-making, *Journal of behavioural decision making, 6*, 95-111.

Trost, G. (1993). Prediction of excellence in school, university and work. In: K. Heller, F.J. Mönks & A.H. Passow (Eds.), *International Handbook of Research and Development of Giftedness and Talent* (pp. 325-336). Oxford: Pergamon Press.

Van der Heijden, B.I.J.M. (1996). *Life-long expertise development: goal of the nineties.* Book of Proceedings of the Fifth Conference on International Human Resource Management. Hyatt Islandia San Diego, California, USA, 24-28 June 1996.

Van der Heijden, B.I.J.M. (1998). *The measurement and development of professional expertise throughout the career. A retrospective study among higher level Dutch professionals.* Ph.D.-thesis. University of Twente, Enschede, the Netherlands. Enschede: PrintPartners Ipskamp.

Van der Heijden, B.I.J.M. (2000). The development and psychometric evaluation of a multi-dimensional measurement instrument of professional expertise. *High Ability Studies. The Journal of the European Council for High Ability, 11*(1), 9-39.

Weber, M, Eisenfúhr, F. & Winterfeldt, D. von (1988). The effect of splitting attributes on weights in multi-attribute utility measurement. *Management Science, volume 34*, 431-445.

# Chapter 4: Thinking about attribute importance

## Structuring and phase relationships in attribute weighting processes[4]


### Abstract

The process of weighting attributes in a non-routine decision problem can be described in a phase model (Chapter 3). The phases of the model follow a logical sequence, but do actors weighting attributes actually relate the phases to each other? That is: do they, in a particular phase, expand on the work of earlier phases? Our results indicate that actors follow an incremental approach:

- actors do not define the criteria for the end result of the weighting process, such as the required precision and measurement level of the weights;
- actors do not define attributes with a view on making them as suitable for weighting as possible, for example, by trying to find a common denominator;
- there are no identifiable quantitative or qualitative relationships between the phases of absolute, relative, homogeneous and heterogeneous weighting;
- given the attention devoted to processing of attributes and to absulute weighting, these phases must have some relevance for the actors, perhaps for framing the weighting problem.

It turns out that thinking ahead in the sense of the logic of the phase sequence seems to play a minimal role in non-routine decision problems.


## 1 Introduction

### Importance assessment and decision theory

When a choice has to be made between alternatives (goods, courses of action), the attractiveness of each alternative can be expressed in terms of the scores of each alternative on a number of attributes and the weight (importance) of each of these attributes (Keeney & Raiffa, 1976). For example, if you buy fruit at the grocery shop, you may want to choose between apples and oranges, judging them on taste, price, healthiness and the like. Each attribute has a certain importance; some people think taste important, some think healthiness is less important than price, and so on.
Many methods exist for measuring the importance that decision-makers attach to the various attributes. They have considerable predictive power (Harte & Koele, 1995;

[4] This chapter is based on an article with the same title, which will be submitted for publication to Journal of Behavioral Decision Making.

Jaccard, Brinberg & Ackerman, 1986). But they show only the *result* of the cognitive processes of the decision-maker: the weight eventually assigned to each attribute. The research reported in this chapter focuses on the way decision-makers arrive at attribute weights. Setting the final weights by the decision-maker (or one or more other actors involved) is called the importance judgment or weight judgment. The cognitive process of arriving at the judgment is called the importance assessment or weight assessment process. Although many aspects of the decision process have been studied, for example individual choice processes (Kahnemann & Tversky, 2000; Keeney & Raiffa; 1976, Keeney, 1992), rationality (Simon, 1960, 1976; Shafir & Leboeuf, 2002; Shafir, Simonson, & Tversky, 1993), group decision processes (Janis, 1971, 1972; Janis & Mann, 1977), organizational decision making (Koopman, 1980), and the measuring of preferences (Jaccard, Brinberg & Ackermann, 1986; Harte & Koele, 1995), the way importance assessments come about has never been studied in depth.

Why is it necessary to study importance assessment processes? Firstly, we believe that if we understand how decision-makers construct importance judgments, we may be able to develop instruments to help them to do it better, or increase their confidence in their work. Secondly, if we can give decision-makers insight in their own thinking processes, they may be able to communicate their thoughts better to others. This may well enhance the efficiency and perhaps the effectiveness of decision-making processes. Thirdly, if we know how importance assessment works, we may be able to find new and better explanations for (often irrational) behavior observed in research cited in Section 1.1.
And finally, knowledge about importance assessment processes can contribute to filling an important gap in management science. The multidisciplinary character of management science is often seen as making it stand out from other areas of research (Easterby Smith *et al*, 1993; Van Riemsdijk (ed.), 1999). But the way in which various disciplines are integrated in management decisions is largely a ‘black box’. ‘For example, how do managers integrate economic, social and technical information when they have to decide whether to enter a new market or to increase their share of an existing market? Weighting, as is clear from the example at the beginning of this chapter, is integrating pieces of incompatible information. This is exactly what managers do all the time and what research on management is all about. So, if we understand importance assessment processes, we may understand the integration of multidisciplinary information better, and hence management processes themselves.

## 1.1 The problem statement

In Chapters 2 and 3 we presented a phased model of the importance assessment process analogous to the many examples of general problem solving models available. (Simon, 1979; Evans, 1991; Hicks, 1995). This model is used as a starting point for this chapter and is briefly described in the next section. We will use the model as a basis for formulating expectations concerning the way actors go through the various phases of the importance assessment process. One of the first questions one could logically ask after having developed a phased model is whether the phases are recognizable in the behavior of actors in practice. We gave an affirmative answer to this question in Chapter 3 and a

summary of the results is shown in Table 1. The next logical question is whether the phases as observed in the model and in the behavior of actors in practice are not only logically but also empirically related to one another. The phases in our model seem to have a logical sequence; they implicitly assume that each phase builds on the work done in previous phases. This will be further explored in Sections 2 and 3. But does this logic show in the behavior of actors in practice?

That is the core question in this chapter. So, our problem statement is:

> *To what extent and in which way are the phases that actors performing importance assessments go through systematically related to each other, in the sense that the output of each phase is used as input for subsequent phases?*

We look at the problem from two perspectives: the rational and the incremental perspective (see Section 2.2). If actors behave rationally, we expect clear relationships between phases, and goal-oriented behavior. If actors behave incrementally, we expect diffuse relationships between phases and trial-and-error behavior. The two perspectives make it possible to formulate specific expectations concerning importance assessment behavior.

Our study serves to characterize the importance assessment process as one of the two perspectives mentioned above, which are well known to researchers on problem solving. (see Section 2.2). This provides a more detailed insight in the importance assessment process and links it to a theoretical concept used for the studying of many types of behavior, thus making mutual comparison possible. Furthermore, it provides insight in the way rational or incremental behavior manifests itself in practice, thus creating a new link between these perspectives and the practical situations they aim to describe.

In this thesis, as mentioned before, we consider the importance assessment process for specific actors and problems:

- laymen (i.e. not experts). Laymen have no possibility to reproduce knowledge or skills from earlier importance judgments when asked to perform an importance assessment. Therefore, importance assessment processes can be more clearly observed with laymen;

- non-routine problèmes. With non-routine problems, there is a definite need to assess the importance of attributes. With routine problems, importance judgments may be readily at hand, so the relevance of importance assessment is much less;

- organizational context. Private decisions do not concern us. Our focus is on business decisions, where formal decision procedures are often used in which instruments for improving the importance assessment process would fit in;

- individual actors. Our aim was to study the importance assessment process without the danger of contamination by group processes. Furthermore, the think-aloud method we used in our research (see Section 2) is only applicable to individual actors;

- Decision processes without time pressure. We limit ourselves to decision processes where there is ample – not necessarily unlimited – time for assessing the importance of attributes. This may be the case when expensive goods that are crucial to the future of a company have to be acquired. Weeks or months may be devoted to careful deliberation on all aspects of such a momentous decision;

- Situations where weighting of attributes is performed. In some choice strategies, weighting is not necessary (Bettman, Luce & Payne, 1998, see Section 4.2 of Chapter 1);

- Situations where the weighting of attributes is done before alternatives are evaluated. This need not always be the case (Bettman, Luce & Payne, 1998).

When important decisions have to be made within an organizational context, like the acquisition of a capital good, the last two conditions mentioned above are likely to be in order.

In section 2, the Weight Assessment Model that served as a structure for studying the behavior of our subjects is summarized. Also, two generic types of the way the model may manifest itself in practice are described: the rational and the incremental approach. Then the research method is addressed in section 3. In section 4, the results are discussed. This is done on the basis of expectations we formulated about the behavior of the subjects, based on our model. A brief discussion follows in section 5.

## 2 Theoretical background

### 2.1 The Weight Assessment Model (WAM)

Our Weight Assessment Model (WAM) consists of 7 main phases and 6 auxiliary activities. The 7 main phases are presented in a sequential way in Table 1. In reality, actors may go back and forth between phases and most often address phases more than once. The auxiliary activities pertain to areas like information search and planning. We do not deal with them in this chapter. Based on Simon's (1960) distinction of the problem-solving process in a structuring and a solving phase, we divide the WAM in a structuring cluster (phases 1 and 2) and a weighting (solving) cluster (phases 3 to 6). Phase 7, the evaluation phase, will not be covered in this chapter. We confine ourselves to the weighting itself, not to its evaluation. In the structuring cluster, the problem (in this case the assignment to weigh attributes) is formulated (phase 1) and the attributes are processed so that they can be readily weighted (phase 2). Then, the weighting takes place in the weighting cluster.

| Phase | Phase name | % of segments devoted to the phase |
|---|---|---|
| 1 Structuring cluster | Problem identification | 6.74 % |
| 2 Structuring cluster | (Sub-) attribute processing | 30.33 % |
| 3 Weighting cluster | Absolute sub-attribute weighting | 27.22 % |
| 4 Weighting cluster | Homogeneous sub-attribute weighting | 4.53 % |
| 5 Weighting cluster | Heterogeneous sub-attribute weighting | 1.50 % |
| 6 Weighting cluster | Attribute weighting | 12.54 % |
| 7 Evaluation cluster | Evaluation | 17.14 % |

*Table : The phases of the WAM*

**Phases of the WAM**

- **Phase 1: Problem identification**
  This phase consists of activities like reading the assignment and, if desired, re-formulating it in one's own words. Essentially, this phase concerns defining, so to speak, the task lying ahead.

- **Phase 2: (Sub-) attribute processing**
  If one wants to weigh attributes, one should first know what one is weighting. Attribute processing concerns giving the attributes a more precise meaning. This can be seen as a case of framing (Akin,1994).

  Attribute properties like measuring level, measuring unit, level of abstractness and precision can change as a result of processing. The following forms of processing were identified:

  a: Decomposing. An attribute can be split up in several sub-attributes;

  b: Re-formulating. When an actor gives an attribute or sub-attribute a different name while meaning the same attribute with a similar, not necessarily identical, measurement unit, the attribute is re-formulated;

  c: Concretising a (sub-)attribute concrete;

  d: Integrating (sub-)attributes into a new 'sub-attribute;

  e: Making an attribute more abstract. This is the complement of concretising.

  Definition is not included in the types of processing. The result of processing may be a description of an attribute that is so exact and formal that it can be called a definition.

The next phases concern the actual weighting process.

- **Phase 3: Absolute (sub)-attribute weighting**
  With 'absolute' weighting (based on Timmermans, 1993), we mean that a statement about the importance of a (sub)-attribute is made without reference to the importance of other (sub)-attributes.

- **Phase 4: Homogeneous sub-attribute weighting**
  This phase is the first in which 'true' weighting takes place: the balancing of the weight of one sub-attribute against that of another. We call this 'relative weighting' (based on Timmermans, 1993). In this phase, two or more sub-attributes of the same main attribute are weighted against each other, and arguments for the weighting are given.

- **Phase 5: Heterogeneous sub-attribute weighting**
  This phase differs in only one respect from the previous one: the sub-attributes that are weighted belong to different main attributes.

- **Phase 6: Attribute weighting**
  This phase concerned the integral weighting of the (two) main attributes.
  This was the original assignment.

- **Phase 7: Evaluation**
  This phase comprises the reflections by subjects on their activities and the results. Several types of evaluation can be identified, like the extent to which the assignment has been fulfilled, evaluations of weights judgments (is the actor, on hindsight, satisfied with assigned weights) and evaluations of arguments (how good are the reasons for particular weight judgments).

## 2.2 Two ways of problem solving

When we want to know how the phases of the WAM are related to each other in practice, and whether there is a relationship at all, we can formulate expectations from two different perspectives, both originating in the literature on structuring reality, problem solving and decision-making. These perspectives and the related methods go by many names, such as the 'hard' and 'soft' approach (Hicks, 1995), the 'rational' versus the 'incremental' approach (Lindblom, 1959; Simon, 1976), systems engineering versus soft systems methodology (Checkland & Scholes, 1990), the 'rational' versus the 'contingent' approach (Visscher, 2001), the management science perspectives of positivism versus constructivism (Easterby-Smith et al, 1993), the two cognitive systems of Teglasi & Epstein (1998) and others. The distinction differs in detail per author. We'll use the terms 'rational' and 'incremental' for the two perspectives, in line with two early publications in which the perspectives were clearly defined (Lindblom, 1959; Simon, 1976) and by which the perspectives are most widely known. It should be noted that the term 'rational' is used in many ways in the literature. We do not use it here in the specific context of, for

example, taking all alternatives and their consequences into consideration (see, for example, Rubinstein, 1998) but as a broad term describing a generic type of behavior that has many possible indicators. This will become apparent when looking at Table 2, in which the differences between the two perspectives relevant for our study are shown, based on the authors mentioned above. The essence is that in the rational perspective, people solving a problem, such as assessing the importance of attributes, proceed in a structured, goal-oriented, well-organized way. They follow a pre-defined series of steps, such as the phases in the WAM, not starting with a new phase until the previous one is completed, for the output of a phase is needed as the input of subsequent phases. The incremental perspective, on the other hand, assumes that problem-solvers do not follow a pre-structured path but move ahead in small steps, frequently going back to earlier steps as experience grows, settling for small steps at a time and generally proceeding in a 'trial and error' fashion.

| **Rational perspective** | **Incremental perspective** |
|---|---|
| Activities are executed sequentially. | Iteration between activities. |
| Goal-oriented. | Trial and error. |
| Striving for definitive solution. | Problem solved in little steps. |
| Planning of solving process in advance. | Planning, if any, 'on the fly'. |
| Relationships between various activities can be made explicit. | Relationships between activities are fuzzy. |
| Comparing all possible (or at least many) alternatives. | Compare only a few alternatives at a time (per iteration). |
| Process has clear beginning and end. | Problem definition and solutions may evolve continuously even after original problem has been solved or is no longer relevant. |
| Objective representation of reality in abstract terms. | Subjective, concrete, narrative representation of reality. |

*Table 2: The rational versus the incremental perspective: main characteristics: the major differences, based upon the publications mentioned above.*

The two perspectives reflect different visions on the problems to be solved and on the context in which they need to be solved. These visions are characterized by Rittel & Webber (1973) in their distinction of 'tame' versus 'wicked' problems. Some characteristics of wicked problems relevant for our study are (Rittel & Webber, 1973):

- there is no definitive formulation of a wicked problem (so, many problem definitions are possible; Hans Heerkens);
- wicked problems have no stopping rule (no defined moment when the problem has been solved; Hans Heerkens);

- solutions are not true or false, but good or bad (so there are no objective judging criteria or measurement; Hans Heerkens);
- there is no immediate or ultimate test of a solution of a wicked problem;
- there is no enumerable (or exhaustively describable) set of potential solutions;
- every wicked problem is essentially unique.

The concepts of 'wicked' and 'tame' problems are suitable to predict whether actors will behave according to the rational or the incremental perspective. Tame or 'structured' problems, with their unambiguous problem definition, clear judgment criteria and identifiable solutions, are, according to various writers, suited for a rational problem solving strategy (Hicks, 1995; Simon, 1979; Visscher, 2001). With wicked problems an incremental approach is in order, as Visscher (2001) observed while interviewing management consultants about their problem-solving strategies. If we compare the characteristics of importance assessment processes for non-routine decision problems to the characteristics of wicked problems, it is clear that an incremental problem-solving strategy is to be expected. While an importance assessment can be unambiguously formulated as 'give weights (importance judgments) for attributes X and Y' and so on, this says nothing about the definitions of the attributes to be weighted, about the meaning of 'weight' (importance), about how precise the weights have to be and so etcetera. So, there is no clear problem definition indeed. It is also not clear when the importance judgment is final (it is final when the actor stops weighting) or which weight is 'right'. While there may be a test of the solution of an importance assessment problem in the sense of the acceptance of the weights by relevant actors or the quality of the resulting decision the validity of such tests is far from a foregone conclusion. So, there is no pre-fixed set of solutions. In our study, we look at non-routine problems solved by laymen. For our research population, the problem at hand is thus unique.

Many normative models in decision theory tend to paint a logical and internally consistent picture of the importance assessment process, but in research on other aspects of decision-making is has become clear that these normative models need not accurately describe behavior in practice (see, for example, Kahnemann & Tversky, 2000 Shafir & Leboeuf, 2002; Shafir, Simonson, & Tversky, 1993, Weber & Borcherding, 1993; Weber, Eisenfúhr & Von Wintefeldt, 1988). So, we indeed have good reasons to expect incremental behavior.

It is clear that, based on the kind of problem this study deals with, and even excluding the social context from consideration, we can expect an incremental problem–solving approach. The expectations we generate reflect this. Although the WAM may at first sight appear to be a model for rational problem-solving, it is not. First of all it is not a prescriptive but a descriptive model. Furthermore, actors need not follow the phases or even address all phases. The phases in problem-solving models in general need not be followed for these models to be recognizable in the behavior of actors (see for further elaboration Lipschitz & Bar-Ilan, 1996). The phases of the WAM are so generally

defined (clusters of activities of a certain type) that there seems to be ample room for both rational and incremental behavior, as is reflected in our expectations.

## 3 The research method

Only the broad outline of the research method is given, as the method is extensively discussed in Chapter 2.

In order to gain insight in cognitive processes of our subjects, we used a think-aloud method. This is a good method for analyzing cognitive processes (Ericson & Simon, 1993). Methods such as choice experiments and process tracing show the results of cognitive processes, but not the processes themselves, while retrospective reporting methods, such as interviews and diaries, leave too much room for interpretation of the cognitive processes by the subjects themselves and are vulnerable to lapses of memory (Ericson & Simon, 1993). Besides, because there was no model of the importance assessment process, the formulation of interview questions would have been very difficult.

Obviously, our research should be seen as theory building, not theory testing. Therefore, we needed a research method with which we could investigate many variables, mostly ill-defined, without knowing which variables would turn out to be relevant for our problem statement, and what the relationships between the variables might be. Methods like experiments and process tracing obviously didn't qualify in this respect, but the think-aloud method proved suitable for the task. This becomes clear when we look at the assignment.

### Sample and assignment

18 undergraduate students of the University of Twente were given an individual assignment based on a fictional case. University students might be assumed to have enough analytical abilities to perform the assignment satisfactorily, without having enough knowledge and skills in this area that would enable them to rely on previous experience of importance assessments. Hence, the danger that they give weights based on previously obtained knowledge is minimized.

Their assignment consisted of supporting the acquisition process of new minibuses by a local company. The subjects were asked to establish the importance of two characteristics of the to-be-acquired minibuses *vis-à-vis* each other, and were asked to imagine that they would be advising the management team during the acquisition process. The attributes, *safety* and *passenger comfort*, were chosen to prevent comparability by some readily available algorithm or heuristic or easy expression in a common denominator such as money. The information that was supplied included a brochure of the company, a leaflet explaining the decision context and two brochures on minibuses; one on a Volkswagen and one on an Opel.

The latter enabled the subjects to get familiar with the specific capital good to be acquired. It was made clear that these examples of minibuses did not mean that the subjects had to make a choice between them.

**Procedure**

The respondents were asked to think aloud during the assessment process. The general guidelines for think-aloud studies given by Ericsson & Simon (1993) were followed, including a practice session to familiarize the subjects with the think-aloud strategy. All verbal information given by the respondent was recorded and typed out literally. After completion of the assignment, a short interview was conducted. In total, each session lasted for a maximum of two hours, for which the subjects were paid 20 Euro. Two pilot sessions were conducted, which led to some minor adjustments of the assignment.

Two kinds of analyses have been performed using the literally typed out protocols:

1: A largely qualitative analysis according to the general rules of the 'Grounded Theory' approach (Strauss & Corbin, 1998);

2: A more quantitative analysis, based on a formal coding scheme that was designed on the basis of the qualitative analysis. Two coders performed the coding activities. Although they worked independently of each other, during the coding of the first six protocols weekly meetings were held to discuss general coding issues in order to enhance the reliability of it. The coders retrospectively applied the refinements to the coding scheme independently. The overall Cohen's Kappa (Baarda & de Goede, 2001) for inter-rater consistence was.0 97 over a total number of verbal segments of 7253.

**Testing of the expectations**

The expectations formulated in this chapter are partly analysed in quantitative terms. Note that we talk of expectations and not of hypotheses, since the number of subjects was too small to test with any convincing statistical validity. For expectations pertaining to the frequency of behaviour to be observed, we use two thresholds. The most far-reaching threshold concerns the behaviour of the entire research population. This assumes that all subjects are more or less identical and behave in the same way, just as one would test expectations about, for example, the fuel consumption of a sample of cars of the same type. In this 'extreme case', we accept an expectation if 80% of the subjects (15, rounded off upwards) behaves according to the expectation. This is a somewhat arbitrary threshold, but with the small number of subjects involved, 80% seems an ambitious but realistic threshold.

It is obvious that generally we cannot expect all subjects to always behave in identical ways. Some will behave rationally, some incrementally. Still, in this case we want to know what the general trend is. Does the majority behave rationally or incrementally? We accept a majority expectation if 12 or more subjects behave according to the expectation (a margin of three subjects above the lower limit of 9 for a draw and also

three below the acceptance limit of 15 for the so-called 'extreme case'. This procedure avoids a nasty problem. In the 'extreme case', an expectation E is accepted if 15 to 18 subjects behave according to it. But this reasoning means that the expectation complement (EC) is also only accepted if 15 to 18 subjects behave in accordance with the converse of the property underlying E. Thus if more than 4 but fewer than 15 subjects behave according to E or EC, neither expectation can be accepted. Out of 19 possible results (0 to 18 subjects can behave according to an expectation), 10 values do not lead to clear conclusions. In the majority case, this number is only four (if 8 to 11 subjects display a certain behaviour).

For most expectations we use several indicators. While all indicators taken together determine whether an expectation is accepted or rejected, the 'weakest' indicator determines the qualification (extreme or majority case).

**Limitations**

The limitations of this research were discussed in Chapters 1 and 2 and summarized in Section 4 of Chapter 3. We will not repeat them here. Suffices it to say that our data do not readily support our conclusions being extended to real-life contexts. Our aim was to establish the structure of the importance assessment process uncontaminated by factors like group dynamics and previous experience of our subjects. The results can be used as a basis for future research introducing factors that manifest themselves in real-life situations. We expect our results to be mirrored in real-life situations to some extent. After all, it is not uncommon for individuals (albeit often with a certain degree of expertise) to make importance assessment processes under circumstances similar to those in our research. But we did not research factors such as 'contamination' by previous importance judgments, which may alter the importance assessment process. So, even if we expect our results to be applicable to real-life situations, we cannot say to what extent and in what form.
Our research pertains only to decision contexts where there is explicit weighting and where the importance assessment process is separated from the evaluation of alternatives. Whenever we make propositions about importance assessment processes in real-world contexts, these propositions can only be validated by further research.

## 4 How do actors go through the structuring and weighting clusters of the WAM and in which way were the phases related?

### 4.1 Introduction

As noted, we expect the subjects to use an incremental approach, given the wicked problem they are confronted with. And because the subjects were not experts and the problem, wicked or not, was not a routine problem for them, the assumption of incremental behavior is made all the more credible. So, the obvious thing to do is: formulate expectations from the incremental perspective. But there are several pitfalls:

- while the rational approach can be described relatively clearly with a limited set of indicators which manifest themselves in a limited number of ways, the incremental approach can manifest itself in many more ways. There are many more ways of being not goal-oriented than there are of being goal-oriented. If a relationship between actions (means) and goals is observed, we can assume that the subject is, consciously or not, goal-oriented. If we do not observe such a relationship, we assume the subject is not goal-oriented, so not rational, and thus incremental. In what way the subject is incremental we don't know, because we would have to test many expectations in order to adequately identify the subject's behavior. So if we would formulate expectations about specific kinds of incremental behavior, and we would not observe the behavior pertaining to these expectations, we would not be able to conclude that the subject is not behaving incrementally – we may not have looked at the right kind of incremental behavior;

- in line with the previous point: analyzing think-aloud protocols means to a large extent: looking for regularities, for patterns. (they may be patterns of inconsistency or irrationality). It is likely, although not certain, that discovering patterns of behavior of individual subjects pertaining to rational behavior are easier to discover than patterns pointing to incremental behavior. This is one of the drawbacks of our positivist research approach (Easterby-Smith *et al*, 1993); a structured, formal analysis of the think-aloud protocols. When we observe patterns, they are thus more likely to point at rational behavior than at incremental behavior. Observing a pattern in itself points to a structure in behavior, that is easily associated with rationality. There possibly is, then, a bias in our research method towards the observation of rational behavior;

- in a way, rational behavior has already been observed during the development of the WAM. The structuring cluster, which can be seen as a step in making wicked problems as tame as possible, and thus eligible for the rational approach, was given much attention and effort in the solving of importance assessment problems (see Table 1), The logic behind the WAM (see Chapter 3) also suggests that the importance assessment process is not completely incremental but has rational elements in it, at least at the level of the process as a whole;

- if we could perform our study from the rational perspective, this would have the advantage of being in line with normative decision theory. Theories like Multi-Attribute Utility Theory (Keeney & Raiffa, 1976), game theory (Sryrms, 1990) assume rationality, as do many decision experiments (Van der Pligt *et al*, 2000). Starting from a rational perspective will, in all likelihood, make it easier to formulate expectations and interpret the results.

To counter these pitfalls, we formulate expectations from the rational perspective. From these expectations, the complementary assumptions assuming incremental behavior will be derived. For the sake of efficiency, there will be no separate expectations for the extreme' and 'majority' cases (see Section 3). In the discussion of the results it will be noted whether the respective expectations are accepted for the extreme case, the majority case or rejected.

Section 3.2 deals with the structuring cluster (phases 1 and 2), and section 3.3 addresses the weighting cluster (phases 3 to 6).

## 4.2 The structuring cluster

### 4.2.1 Expectations

The structuring cluster starts with the formulation of the problem. As noted earlier, a weighting problem is a wicked problem. Given the nature of such problems, this means that: the attributes to be weighted have to be clear at the end of the structuring phase and that the end result in terms of, for example, the character of the weights, has to be made explicit. This can be seen as framing (Akin, 1994). A clear definition of the end result would help greatly in assessing the tasks that have to be fulfilled in order to complete the assignment. For example, if the weights need not be very precise, perhaps exact definitions of attributes are not necessary. If actors behave rationally, they can be expected to have, at the end of the structuring phase, exact definitions of the attributes to be weighted and a clear idea of what the weights should look like, so that the actual weighting is as straightforward as possible. If actors behave incrementally, they are likely to have fuzzy definitions of attributes and no clear concept of weights.

In order to achieve all this and do it at the high level of quality required for non-routine, far-reaching decisions like the acquisition of capital goods, we would expect that if actors behave rationally, they try to work in a more or less systematic way, successfully or otherwise. Many prescriptive problem-solving models basically aim to make the problem-solving process more systematic. Fine examples are the model by Kepner & Tregoe (1981) and morphological forced connection models (Hicks, 1995), but it is also the case for other problem-solving models (see for an overview Hicks, 1995). In Chapter 3 we addressed systematic working at the level of the importance assessment process as a whole and presently we will go into systematic working in specific phases of the process. In this chapter, we do not focus on the use of established methods, but on any signs of a systematic approach. We also do not look at methods specifically designed for weighting, but at systematic reasoning pertaining to specific activities within the phases, which need not be unique to weighting. If actors go through the assignment in an incremental way, we expect them not to use these methods. The relationships between phases and clusters are expected to be fuzzy, so there is no need for systematic working to secure good input for subsequent phases.

In order to assign weight values, one has to have a notion of what ‘weight’ means. The definition of ‘weight’ is far from a foregone conclusion (Alpert, 1971; Goldstein & Mitzel, 1992; Jaccard, Brinberg & Ackerman, 1986; Wilkie & Pessemer, 1973). A rational actor would tackle this problem before the actual weighting starts, if only to guarantee that all attributes are weighted in the same manner. An incremental actor would, if concerned with meaning of ‘weight’ at all, let the concept evolve during the weighting.

The same logic applies to the definition of the attributes themselves. In the rational approach, the attributes are explicitly defined before they are weighted. In the incremental approach, with its trial and error character, definitions may not be made explicit or there may be a constant iteration between defining and weighting, until in the end the actor is satisfied with the weight for a (possibly undefined) attribute. Since the rational approach is characterized by the representation of reality in objective and abstract terms, defining attributes (making their meaning more objective and usually entailing abstraction) could be expected. An incremental actor would probably use concrete examples of the attribute, which would likely vary during the weighting process.

Systematic attribute processing (phase 2) is a logical way to arrive at adequate definitions of (sub-)attributes. As is clear from table 1, subjects spent much effort on attribute processing. So we will look for systematic processing as a form of rational behavior. If actors behave incrementally, they are unlikely to process attributes systematically but instead process them haphazardly, each new processing step being determined by, for example, the outcome of previous steps, coincidence, the information that happens to be at hand and so on.

If actors have the general shape of a utility function (Keeney & Raiffa, 1976) in mind, which could be the case if they behave rationally, one would expect them to try and find common denominators, such as money, for the attributes. Then the weights could be derived from the relative impact of each attribute on the common denominator. This would be an eminent foundation for the weighting cluster. For example, if one unit of safety would have twice as much impact on revenues as one unit of comfort, safety could be considered twice as important as comfort. In Chapter 3 we reported that subjects were not concerned with finding a common denominator. So we will not devote attention to this issue here.

All in all, the following expectations can be formulated:

- For phase 1 (problem identification):
  - *Expectation 1A:*
    Subjects use the rational approach and translate the assignment in requirements for the end result, in order to define and structure the tasks to be fulfilled in later phases to complete the assignment;
  - *Expectation 1B:*
    Subjects use the incremental approach and do not translate the assignment in requirements for the end result, as they are not concerned with defining the tasks ahead;
  - *Expectation 2A:*
    Subjects use the rational approach and try to) work systematically in order to secure good input for the weighting cluster;

- *Expectation 2B:*
  Subjects use the incremental approach and do not try to work systematically, as they are not concerned with defining input for future phases in advance;

- *Expectation 3A:*
  Subjects use the rational approach and give a clear definition of 'weight' in order to establish possible weight values in later phases and to decide how exact definitions of attributes have to be;

- *Expectation 3B:*
  Subjects use the incremental approach and give no clear definition of 'weight' as these definitions, if required, will evolve during the importance assessment process;

- For phase 2 ((sub-)attribute processing):

  - *Expectations 4A:*
    Subjects use the rational approach and try to process attributes in a systematic way, in order to start the weighting clusters with clear definitions or concepts of the attributes;

  - *Expectations 4B:*
    Subjects use the incremental approach and do not try to process attribute in a systematic way, using implicit or evolving definitions or concepts of attributes in the weighting cluster;

  - *Expectation 5A:*
    Subjects using the rational approach explicitly define the attributes to be weighted in order to determine exactly what has to be weighted in the weighting cluster;

  - *Expectation 5B:*
    Subjects using the incremental approach explicitly do not explicitly define the attributes to be weighted as definitions, if used at all, will evolve during the *weighting process*.

### 4.2.2 Results: phase 1 (problem identification)

In Table 3, the expectations that are accepted are shown, along with the qualification of acceptance (extreme case or majority case). The terms 'extreme case' (15 or more subjects behave according to the expectation) and 'majority case' (12 or more subjects behave according to the expectation) were explained in Section 3.

| Expectation | Acceptance qualification |
|---|---|
| Expectation 1B: Subjects use the incremental approach and do not translate the assignment in requirements for the end result, as they are not concerned with defining the tasks ahead. | Majority case |
| Expectation 2B: Subjects use the incremental approach and do not try to work systematically, as they are not concerned with defining input for future phases in advance. | Extreme case |
| Expectation 3B: Subjects use the incremental approach and give no clear definition of 'weight' as these definitions, if required, will evolve during the importance assessment process. | Extreme case |

*Table 3: Results (phase 1)*

As for expectation 1A and 1B indicators were:

- statements that indicated a re-formulation of the assignment in the subject's own words that explicitly or implicitly contained requirements for the end result. Reading aloud or repeating the assignment does not qualify;

- a stated concern about aspects of the end result over and above what was already in the assignment (for example: the required precision of precision of the weights).

As for expectation 2A and 2B the indicators comprised: statements that indicated (a desire for) systematic reasoning. This systematic reasoning had to be more than coincidental. So, just the mentioning of the desire to work systematically without giving it any follow-up, or stating some systematic way of working that could be followed without putting it into practice, will not lead to acceptance of the expectation.

Expectations *1B and 2B are accepted*. No subject translated the assignment into requirements for the end result or, as far as could be identified, worked systematically during phase 1. None of the subjects wondered how precise the weighting had to be. In the assignment, only general qualifications like 'importance' were used, so there was no requirement as to the precision of the weights. If the subjects had wanted to reduce the complexity of the assignment as much as possible, they could have said whether safety was more important than comfort and left it at that. Indeed, 7 of the 13 subjects who weighted safety and comfort (and not their sub-attributes) and who did not initially set conjunction thresholds gave this kind of weighting before asked by the experimenter to be more precise. But the other 11 subjects either set conjunction thresholds or gave numerical weights, going beyond what the assignment explicitly specified and more towards what it implicitly demanded.

As for expectation 3A and 3B the indicator is whether subjects explicitly address the definition of 'weight' more than incidentally.

Expectation *3B can be accepted* and thus expectation 3A is rejected. 1 subject (18) explicitly concerned himself with the meaning of the concept of 'importance'. That is an insignificant minority, and it is obvious that subjects either had a clear idea of what 'importance' meant, were not sure what it meant but didn't realise it, or were not sure what it meant but did not consider that to be a problem in the sense that it would prevent them from completing the assignment. In the interviews after completing the assignment, the meaning of 'importance' was not mentioned as a problem.

### 4.2.3 Results: phase 2: ((sub)-attribute processing)

In Table 4, the expectations that are accepted are shown, along with the qualification of acceptance (extreme case or majority case).

| **Expectation** | **Acceptance qualification** |
|---|---|
| Expectations 4B: Subjects use the incremental approach and do not try to process attribute in a systematic way, using implicit or evolving definitions or concepts of attributes in the weighting cluster. | Extreme case |
| Expectation 5B: Subjects using the incremental approach explicitly do not explicitly define the attributes to be weighted as definitions, if used at all, will evolve during the weighting process. | Extreme case |

*Table 4: Results (phase 2)*

As for expectation 4A and 4B indicators were:

- a consistent system that could be observed in the processing of attributes, like the consistent decomposition of attributes in logical categories. An example is: categorizing all safety attributes of a minibus under either active or passive safety. Again, there should be a minimal level of consistence. Just mentioning active and passive safety as attributes and not using them to categorize other attributes doesn't count;

- statements indicating the use of some explicit and more than incidental systematic routine, for example linking attributes to theoretical concepts with a certain consistency. For example, using the physical relationship between mass, acceleration and stopping distance to identify the quality of the braking system of a minibus.

In this chapter we do not look at so-called stopping criteria: explicit criteria for deciding when decomposition had yielded enough sub-attributes and hence could be stopped. This aspect is covered in detail in Chapter 5.

It should be stressed that merely reading through the written material provided does not count as working systematically. The system in which the information was presented in the material did not provide a suitable system (implicit guidelines) for conducting the assignment as far as we know. Relevant information was scattered through the material.

Expectation *4B can be accepted*, so expectation 4A has to be rejected, as will be shown now.

Safety and comfort were decomposed in a large number of sub-attributes (see Chapter 3). The averages for the two attributes 'safety' and 'comfort' were 19.6 and 24.4, respectively. For isolated groups of sub-attributes, decomposition was at times systematic, but generally it seems to have been a case of just mentioning sub-attributes on the bases of the available information or the subjects' own knowledge until the subjects could think of no more relevant sub-attributes or until they had the idea that they had formulated enough sub-attributes. When a system was used, it was seldom, if ever, maintained. For example, one subject divided 'safety' in active safety' and 'passive safety' but never brought other safety sub-attributes under these two categories, which would have been the justification for using the system of active and passive safety in the first place. Like with re-formulation and specification, the main effect, if not a goal, seems to have been that the subjects could get a clearer image of 'safety' and 'comfort'. For the five subjects who did not weigh the main attributes (safety and comfort), the processing phase provided them with (loose concepts of) the sub-attributes they eventually weighted.

It is clear that fewer than 15 subjects conducted phase 2 in a systematic way, so expectation 4A is rejected and *4B is accepted* for the extreme case.

As for expectation 5A and 5B we use two indicators:

- the (explicit or implicit) measurement level of the attributes, possibly defined in a scale. The reasoning is that an interval or ratio measurement level implies a scale with defined values (thus the meaning of the attribute should be quite clear), whereas a dichotomous or ordinal scale roes not require this. Of course, this does not mean that attributes with ordinal scales are by default not defined clearly, so we also look at:

- explicit use of definitions on a more than incidental basis.

Expectation 5A has to be rejected and *5B can be accepted* for the extreme case. Only in isolated cases did subjects try to define (sub-)attributes formally. For example, one subject defined 'safety' as 'number of deaths per year'. But he then proceeded to use this definition as a sub-attribute alongside other sub-attributes. He hence used the same way as all other subjects to clarify the attributes to be weighted: splitting them into a number of sub-attributes. So, formal definitions were of no relevance whatsoever.
None of the subjects explicitly defined a consistent scale across attributes, for example a five-point scale, as required by multi-attribute utility theory (Keeney & Raiffa, 1976). One subject (17) consistently defined crude scales for all sub-attributes like 'yes/no' for

the presence of certain sub-attributes, 'fully loaded/not loaded' for the amount of baggage, and 'opinion of colleagues' for driving quality. It was clear that it helped him to get the meaning of sub-attributes clear. Sometimes subjects defined extremes of the possible scores of safety and comfort. But these extremes were not meant to define practical range, but as thought experiments in order to get a grip on the concept of comfort or safety, to indicate which alternatives they excluded or did not expect to have to choose from, or to make in which cases either safety or comfort would be important in any case. One of the subjects, for example, said: "Well, then I could describe comfort for myself as the worst case that you take a trip in a Ford Fiesta and a big limousine as the best case".

Of the 13 subjects that weighted safety and comfort, as was the assignment, all used an implicit ordinal scale for the two attributes. No one ordered the scale values (defined specific levels of safety and comfort) which is a demand for a proper scale, even on the ordinal level (Cooper & Shindler, 1998). Subjects merely used qualifications like 'very safe', 'safer than', less comfortable than'. A binomial scale, although not consistently used, was the most common scale used during the weighting of sub-attributes

The goal of phase 2 was never mentioned explicitly and was not investigated during the interviews, but 7 of the 18 subjects (39 %) explicitly stated that they wanted to define 'safety' and 'comfort' by saying things like: "What do I think 'safety' means?" and: "Well, yes, I think I'll start by defining the attributes". This should not be so much considered as a goal (the reason for giving a definition and the contribution to fulfilling the assignment) but as the description of an activity. And 'wanted' cannot be taken to be the same as 'tried', but at least it was an indication that the subjects were aware of the possibility or desirability of defining attributes. Therefore, we do not consider this indicator strong enough for contradicting acceptation of expectation 5B.

### 4.2.4 The structuring cluster: conclusions

We have shown that our subjects take an incremental approach to structuring the weight assessment problem presented to them. We expected this to be the case, but the facts that all expectations predicting incremental behavior are accepted in the extreme case makes the results all the more convincing.

The dominance of the incremental approach is in a way remarkable since the assignment stated that the subjects should be able to explain the result of the importance assessment process (the importance judgment) to others, for example the management of the minibus company. It is easier to justify an importance judgment if it can be shown that it has been reached in a systematic fashion. Yet, when queried afterwards about their importance judgments, when justifications were given, they consisted mainly of arguments why certain (sub-)attributes were important or not. The rationality of the process did not enter the conversation. Subjects would say, for example, that safety was important because accidents would scare prospective clients off. Arguments like this don't pertain to working methods.

All in all, the structuring cluster did not provide a solid basis for the weighting cluster: a great number of ill-defined sub-attributes with unclear relationships and no idea of the required precision of weighting.
This makes it all the more interesting to see how the weighting cluster is performed.

## 4.3 The weighting cluster

### 4.3.1 Expectations

In this section, we will again take the distinction between the rational and incremental approach as the basis for formulating expectations. Taking the rational perspective, as noted, the output of one phase should be the input for the next. Now this is a very strict demand. We would not expect actors, even if they are rational, to formulate their thinking process so diligently that an observer could ascertain a deterministic relationship between the output of one phase and the input of the next. So we loosen the requirement in the sense that the output of one phase should be recognizable in the input of the next. So the two need not be identical bus merely similar. This similarity can occur in two ways. There can be a quantitative relationship between the weights established in one phase and the weights established in the next. For example: the weight of an attribute in phase 6 (main attribute weighting) could be the sum or average of the weights of phase 3 (absolute sub-attribute weighting). Note that the term 'quantitative relationship' extends also to ordinal or binominal weights in this case. So an actor could decide that if one sub-attribute of safety, like the quality of the brakes is 'very important', but that the presence of seatbelts is 'not important' than the weight of safety could be 'important' (between 'very important' and 'not important'). It is also possible that the weight of an attribute is related to the number of its sub-attributes, assuming that all sub-attributes have equal weights. This is in parallel with the equal weights choice strategy (Bettman, Luce & Payne, 1998, see Section 4.2 of Chapter 1). There can also be a qualitative relationship between phases. The output of one phase can determine how the next phase is performed. For example: only those (sub-)attributes that get a certain minimum weight in phase 3 are considered for subsequent phases (analogous to the Image Theory (Beach, 1993; Beach, Puto & Heckler *et al*, 1996) for judging alternatives instead of weights).
In Table 5, examples of quantitative and qualitative relationships between phases are given.

| Types of weighting | Quantitative relationship | Example | Qualitative relationship | Example |
|---|---|---|---|---|
| From absolute to relative weighting | Absolute weights indicate relative weights | Absolute weights identical or similar to weight ratios of relative weights | Absolute weights indicate way relative weighting will be performed | Only attributes that have weights above a certain threshold are included in relative weighting |
| From homogeneous to heterogeneous weighting | Homogeneous weights indicate heterogeneous weights | If all homogeneous weights and one heterogeneous weight pair are known, some or all heterogeneous weights can be inferred | Homogeneous weights indicate how heterogeneous weighting will be performed | The most important sub-attribute of attribute A is weighted against the most important sub-attribute of attribute B, and so on with the second most important, third most important sub-attributes etc. |
| From sub-attribute to attribute weighting | Sub-attribute weights indicate attribute weights | Attribute weight is or approaches sum or average of sub-attribute weights | Sub-attribute weights indicate how attribute weighting will be performed | The definition of the attributes are modified by eliminating elements represented by unimportant sub-attributes |

*Table 5: Quantitative and qualitative relationships between phases.*

Since so little research has been done on importance assessment processes (see Section 1.1) we could not be sure to identify every possible relationship between phases before conducting our experiment. So we can only identify relationships if the subjects in our study mention them explicitly (which means that the subjects are somehow aware of them) or if we recognize them from the think-aloud protocols. In sum; there is no guarantee that if we do not identify a relationship, there really is none. However: we think we can isolate the types of protocol segments in which possible relationships could show up. They are:

- segments concerning activity planning, in which subjects may explain how weighting is to be performed;

- segments concerning weights mentioned by the subjects. We can compare weights in order to see if there are relationships;

- segments concerning arguments for giving certain weights. These arguments may pertain to weights given earlier;

- segments concerning evaluation of weights or arguments. In these, possible relationships may be explained. For example, subjects may check of an attribute weight is indeed the sum of the constituting sub-attribute weights and, if not, proceed to adjust the weight.

Obviously, the manifestation of any of the above relationships points at actors following a rational approach. From our analysis of the behavior of our subjects in the structuring cluster, however, we have every reason to expect an incremental approach.. So, relationships between phases will be either absent or diffuse, partly because phases will probably not have clear goals (see Section 2.2). Of course we expected incremental behavior from the beginning, given the 'wicked' nature of important assessments, and our initial expectations were borne out in the structuring cluster. So, for efficiency reasons, we will not formulate expectations for both the rational and the incremental approach, but only for the incremental approach. If any of these expectations is rejected, we will then formulate and test an alternative expectation concerning the rational approach.

We suspect, although we could not prove it, that actors do not use an absolute scale for importance while weighting (see Chapter 3). Having no absolute scale makes it difficult to compare weights, so this is an extra reason for expecting an incremental approach. In that chapter we also looked at the relationship between sub-attributes and main attributes in the sense whether the meaning of the main attributes as eventually weighted could be linked to the meanings given to the sub-attributes. We were not able to establish clear relationships. This makes us expect that weights of sub-attributes and attributes are also not related. After all, if there were definitional relationships this would be a good basis for thinking about weight relationships.

We formulate the following expectation:

- *Expectation 6:*
  The way the subjects proceed through the weighting cluster reflects an incremental approach and therefore there are no quantitative and/or qualitative relationships between absolute and relative weighting (phase 3 versus phases 4 to 6);

- *Expectation 7:*
  The way the subjects proceed through the weighting cluster reflects an incremental approach and therefore there are no quantitative and/or qualitative relationships between homogeneous and heterogeneous weighting (phase 4 versus phases 5 and 6);

- *Expectation 8:*
  The way the subjects proceed through the weighting cluster reflects an incremental approach and therefore there are no quantitative and/or qualitative relationships between sub-attribute and attribute weighting (phases 3 to 5 versus phase 6).

### 4.3.2 Results: the relationships between phases of the weighting cluster

Table 6 shows that expectations 7 to 9 can be accepted.

| Expectation | Acceptance qualification |
|---|---|
| Expectation 6: The way the subjects proceed through the weighting cluster reflects an incremental approach an therefore there are no quantitative and/or qualitative relationship between absolute and relative weighting (phase 3 versus phases 4 to 6). | Extreme case |
| Expectation 7: The way the subjects proceed through the weighting cluster reflects an incremental approach and therefore there are no quantitative and/or qualitative relationship between homogeneous and heterogeneous weighting (phase 4 versus phases 5 and 6). | Extreme case |
| Expectation 8: The way the subjects proceed through the weighting cluster reflects an incremental approach and therefore three are no quantitative and/or qualitative relationship between sub-attribute and attribute weighting (phases 3 to 5 versus phase 6). | Extreme case |

*Table 6: Results (weighting cluster)*

As for expectation 6, 7 and 8, indicators were:

- more than incidental statements about quantitative and/or qualitative relationships between phases made by the subjects as depicted in Table 5. We exclude statements about sequential relationships like 'Now I think I have weighted all sub-attributes, and it's time to weigh safety against comfort'. Such a statement may indicate a relationship between phases, but does not shed any light as to the nature of that relationship;

- observed quantitative or qualitative relationships between weights of (sub-)attributes or weights and weighting procedures in subsequent phases as depicted in Table 5. For example, if 'safety' was split in two sub-attributes, one with a weight of 0.1 and another with a weight of 0.2, and in the end 'safety' gets a weight of 0.3, then we assume that the weight of 'safety' is the added weights of the two sub-attributes.

We looked for these indicators both by reading through the entire protocols and later by focusing on particular types of segments as mentioned above.
The relationships between weights derived from absolute weighting and those that were the result of relative weighting were completely unclear, except in one case, and *expectation 6 is accepted* for the extreme case. The exception was subject 16 who sometimes defined common denominators for sub-attributes pertaining to the main attribute, like effectiveness' and common function'. In the first case, the

effectiveness/cost ratios could sometimes be recognized in the rank order of attributes in phase 4. Also, when attributes had a common function, only one of them should prevail (get the weight of the attribute denoting the function) in any decision, whereas the other two would get weights of zero. But these were the only examples of quantitative relationships between phases 3 and 4 to 6.
It is useless to speculate on possible relationships, but it seems an important area for future research. After all, so much effort is devoted to phase 3 that we have no choice but to assume that important groundwork is laid for later phases, at least for phase 6 (main attribute weighting), which was, after all, the essence of the assignment. Yet, not a single statement was found that pointed unambiguously at subjects using phases 4 to 6 as checks for weights assigned in phase 3. Sometimes, subject started with a general statement about the relative importance of safety or comfort before the actual weighting started, and later, when the assignment was fulfilled, observed that they had stuck to their original rank order of importance. But these were observations of facts, not tests. Sometimes, subjects went back to individual sub-attributes after establishing relative weights, but they usually went back to only one of the constituting sub-attributes and seemed to keep thinking in relative terms (for example, weight ratios, with statements like ‘safety thus has a weight of ¼’.

In all cases, some of the sub-attributes weighted in phase 3 did not return in later phases, but only 2 subjects explicitly attributed this in some cases to the sub-attributes not having attained a minimum level of importance. In all other cases, the sub-attributes simply were not mentioned any more (perhaps they were forgotten) or specific arguments were given why they received a weight of zero. While giving a sub-attribute a weight of zero excludes it from further weighting, this does not imply a relationship with the next weighting phases. The weighting of that sub-attribute is simply finished. In the case if sub-attributes being eliminated because their weights are low, this shapes subsequent phases: the weighting of those sub-attributes is not finished and they could logically well be weighted against other sub-attributes, but subjects choose not to do so.

Some subjects escaped relative weighting by proposing conjunction thresholds, especially for safety or its sub-attributes.

Concerning expectation 7 only in one case could a relationship between homogeneous and heterogeneous sub-attributes be established, so *expectation 7 is accepted* for the extreme case. One subject (17) made rank orders of all sub-attributes of safety and comfort, respectively. He then compared the sub-attributes of equal rank order of safety and comfort to each other: the most important sub-attribute of safety with the most important sub-attribute of comfort, the second most important sub-attributes to each other, and so on. Note that the fact that this is an incomplete pairwise comparison makes the relationship between phases 4 and 5 visible. In a correct pairwise comparison, each sub-attribute has to be compared to every other sub-attribute, thereby making the distinction between homogeneous and heterogeneous weighting superfluous. The distinction has can be useful for an actor, however. It may be easier to find a common denominator (if one is desired) for sub-attributes pertaining to the same attribute than when the sub-attributes pertain to different attributes. However, as noted earlier common

denominators played an insignificant role. It would be interesting to see if in importance assessments performed by experts the distinction between homogeneous and heterogeneous weighting is of more significance.

*Expectation 8 is accepted* for the extreme case. If would seem logical to suppose that the results of earlier phases would be the input for phase 6, but this was not the case in any explicit way, in the sense that there was no relationship between the weights assigned in earlier phases and the weights assigned in phase 6, with the exception of one subject (17). As mentioned above, he compared sub-attributes of safety and comfort pairwise, and concluded, after seeing that each time the sub-attribute of comfort won, that comfort was apparently more important. But, even then, this was a conclusion he drew only after having been questioned about it afterwards, not during the assignment. This was the only clear-cut case of a relationship between homogeneous and heterogeneous weighting.

In the case of one subject (4), a relationship could be established between the weights of the sub-attributes and those of the main attributes, but in this case the weights of the main attributes were assigned first and not further deliberated thereafter.

In various experiments relationships have been observed between the weights of sub-attributes and main attributes. Weber & Borcherding (1993) found that the more detailed the sub-attributes, the higher the weights assigned. Weber, Eisenfúhr & Von Winterfeldt (1988) found that the directly assigned weight of an attribute is lower than the sum of the weights of the sub-attributes if the attribute is decomposed, although this effect can, according to Póyhónen & Hámáláinen (1998) be partly explained by the method of research (computing aggregate average weights instead of weights per individual). Thus, there are reasons to assume a relationship between earlier phases and phase 6. But, even when some subjects, mostly after intervention by the experimenter, gave quantitative weights for both the main attributes and the constituting sub-attributes, this relationship between these weights could not be established and, with the exception of subject 17 (see above) the subjects themselves did not seem to try and establish such a relationship explicitly.

Not a single statement could be found that indicated that the definition of any main attribute was modified based on the weights of sub-attributes. As we saw earlier, no exact definitions were used anyway. The definitions that subjects may have had had in their heads may well have shifted, but not in a way that could be observed unambiguously.

All this is not to say, of course, that the preceding phases had no significance at all. But what this significance was is far from clear. From the protocols it is clear that arguments why certain (sub-)attributes were or were not important used during phase 6 often originated in earlier phases. But would the eventually assigned weights have been any different if the subjects had somehow been prevented from going through phases 2 to 5 and had been forced to weigh safety and comfort immediately after having received the assignment? This is unclear since the subjects were not asked to give preliminary weights after having received the assignment, for fear of stimulating rationalization of those preliminary weights.

### 4.3.3 The weighting cluster: conclusions

In accordance with to expectation 6 to 8, there were no clear systematic relationships between the phases as conducted by the subjects. The results point clearly and unambiguously to the importance assessment process being conducted in an incremental way, at least for the type of actors that we studied. Given the wicked nature of importance assessment problems this was to be expected, but the degree of incrementality is surprising, given the clear structure of the process as a whole in the case of almost every subject and the effort of the subjects to achieve a balanced, well-motivated result. Since all subjects conducted absolute weighting and the majority ended up in comparing safety with comfort (relative weighting), the way in which they proceeded from one phase to another touches the core of the weighting process (comparing incomparable attributes). Hence, this is an area for further research. It would be especially interesting to see whether in importance assessment processes conducted by experts one would indeed observe clear relationships between phases.

## 5 Discussion

We started this chapter with the following problem statement:

> *To what extent and in which way are the phases that actors performing importance assessments go through related to each other, in the sense that the output of each phase is used as input for subsequent phases?*

It was convincingly demonstrated that both in the structuring and in the weighting cluster laymen tackling non-routine importance assessment problems work in an incremental way. The structuring cluster does not provide a good basis for the weighting cluster (no clear definition of ‘importance’ and of the attributes to be weighted). In the weighting phase, no clear, unambiguous relationships between the phases could be observed. The distinction between phases 4 and 5 (homogeneous and heterogeneous weighting, although clearly observable, seems to provide little insight in the importance assessment process.

Our results confirm the relationship between wicked problems and incrementalism found in the literature (see Section 2.2), at least for laymen tackling non-routine importance assessment problems, separated from the evaluation of alternatives. They also show how the theoretical concept of incrementalism manifests itself in a common, practical situation. But they also show the limits of the concept. If we cannot explain what the relationship between the structuring and the weighting cluster is, or the relationships between the phases of the weighting cluster, how can we explain why importance assessment processes come as they do? Rational behavior has a self-explanatory logic that incrementalism lacks. Given that the general structure of the WAM is clearly recognizable, at least with our subjects, in importance assessment behavior, the absence of relationships between phases is somewhat unsatisfying. Even when actors behave incrementally, we could have observed phase 6 (attribute weighting) ‘grow’ from the

preceding phases. But, as stated earlier, the relationship between phase 6 and the preceding phases, if any, remains unclear.
Two phases of the WAM stand out in terms of the effort devoted to them: (sub-)attribute processing and absolute (sub-)attribute weighting (phases 2 and 3, see Table 1). It is in itself remarkable that so much effort is devoted to phases that are merely preparations to the actual task: phase 6 (attribute weighting). Although relations between (sub-)attributes are established in phases 4 to 6, it remains unclear how the results of phases 2 And 3 contribute to these latter phases. There are three possible explanations. The weighting process may be essentially a process of absolute weighting, in which virtually all weighting is done in phase 3 and phases 4 to 6 are merely used to 'translate' the result of phase 3 into the specific weights required by the assignment. Or there is a relationship between phase 3 and subsequent phases in the sense that phase 3 provides the building blocks for the actual weighting in subsequent phases. We found no indications to support this explanation. Finally, it is possible that importance assessment processes are so personal that some aspects cannot be studied by looking at groups of subjects. In-depth research of individual cognitive processes may be in order. This calls for elaborate psychological research that is beyond the scope of this thesis and entails a much more constructivist approach (Easterby-Smith *et al*, 1993) than we followed in this research.

The observation that actors behave highly incrementally during importance assessment processes may shed a new light on the results of studies mentioned in Section 1.1. In many of these studies, irrational behavior is observed (see, for example, Kahnemann & Tversky, 2000 Shafir & Leboeuf, 2002; Shafir, Simonson, & Tversky, 1993). Still, Van der Pligt *et al*. (2000) observe rationality of the subjects. In these studies, laymen were often involved, tackling problems more routine than ours, but still artificial. These studies, like ours, were often conducted in a laboratory and not in real-life situations. If subjects proceed incrementally in such a way that absolute weighting is much more frequent than relative weighting, perhaps some irrational behavior in these studies can be explained by assuming dynamically shifting weights which, taken together in one point in time, may well be inconsistent. We will not dwell on this issue, but it is worthy of further examination. This line of reasoning follows the ideas of Lootsma (2002).

A complication is that it is by no means certain that actors in real-life organizations perform importance assessment processes in the same way as our student (laymen) subjects did. Experts differ from laymen (Chi, Glaser & Farr, 1988; Van der Heijden, 1998) and actors having to deal with organizational constraints differ from subjects studied in a laboratory. These differences can pertain to the way the phases of the WAM are executed (see Chapter 3). For example: more or less effort may be devoted to a phase. It is also possible that the structure of the WAM differs from the structure of importance assessment processes in real life. For example, as yet unidentified phases may play a role. We expect our results to be relevant for real-life situations, but we cannot be sure. This is an area for further research. Aspects of this issue are covered in Chapter 3. And we should remember that, as stated earlier, decisions not always involve weighting. So, our results only apply to situations where attributes are actually weighted and where the weighting is separated from the evaluation of alternatives.

A way to look at the relationships between phases is to study their sequencing in time. Is a phase only started after all preceding phases are completed? Or do actors iterate between phases? As reported in Chapter 3, a cursory look at the protocols showed a certain global phase sequence, but rather wild jumping on a detailed level as well. This problem is not addressed in this thesis, as it is far from straightforward. For example, actors can perform phase 2 of the WAM (attribute processing) for all attributes before proceeding further, but they could also process each attribute and then weigh it (phase 3) before returning to phase 2 for processing the next attribute. What is rational and what is incremental? Can subjects use both methods at the same time and still be rational? Questions like these will be addressed in a future article, pending the completion of the required analysis.

All in all, the WAM seems to be a valid analysis tool, but so far it has merely served to point at the pivotal issues in the quest for understanding the importance assessment process. The incremental behavior identified in this chapter remains a fruitful area for research.

**Literature**


Akin, Ó. (1994). Creativity in design. *Performance Improvement Quarterly, volume 7*, 9-23.

Alpert, M.I. (1971). Identification of determinant attributes: a comparison of methods. *Journal of marketing research*, volume 8, 184-191.

Baarda, D.B. & de Goede, M.P.M. (2001). *Basisboek methoden en technieken*. Groningen: Stenfert Kroese.

Beach, L.R. (1993), Broadening the definition of decision making: The role of pre choice screening of options, *American Psychological science*, volume 4, 215-220.

Beach, L.R. Puto, C.P, Heckler, S.E, Naylor, G & Marble, T.A (1996). Differential versus unit weighting of violations, framing, and the role of probability in Image Theory's compatibility test. *Organizational behavior and Human Decision Processes, volume 65*, 77-82.

Bettman, J.R, Luce, M.F. & Payne, J.M. (1998). Constructive consumer choice processes. *Journal of consumer research*. Volume 25, 187-217

Checkland, P, & Scholes, J. (1990). *Soft Systems Methodology in action*. Chichester [etc]: John Wiley & Sons.

Cooper, D.R. & Schindler, P.S. (2003). *Business research methods, 8th edition.* Boston: Irwin/McGraw-Hill.

Easterby-Smith, M, Thorpe, R. & Lowe, A. (1993). *Management research, an introduction*. London [etc]: Sage Publications.

Ericsson, K.A. & Simon, H.A. (1993). *Protocol analysis: verbal reports as data* (2nd ed.). Cambridge, MA [etc.]: Bradford Books/MIT Press.

Evans, J.R. (1991). *Creative thinking in the decision and management sciences.* Cincinnati: South-Western Publishing Co.

Goldstein, W.M. & Mitzel, H.C. (1992). The relative importance of relative importance: Inferring other people's preferences from relative importance ratings and previous decisions, *Organizational behavior and human decision processes,* Volume51, 382-415.

Harte, J.M & Koele, P. (1995). A comparison of different methods for the elicitation of attribute weights: structural modeling, process tracing and self-reports. *Organizational Behavior and Human Decision Processes, volume 64,* 49-64.

Hicks, M.J. (1999). *Problem solving in business and management; hard, soft and creative approaches.* London [etc.]: International Thomson Business Press.

Jaccard, J, Brinberg, D & Ackerman, L.J (1986). Assessing attribute importance. *Journal of Consumer Research, volume 12,* 463-467.

Janis, I.L. (1971) Groupthink. In: Srtaw, B (ed.) *Psychological dimensions of organizational behavior*. New York: McMillan.

Janis, I.L. (1972). *Victims of groupthink; a psychological study of foreign policy decisions and fiascos*. Boston: Houghton, Mifflin

Janis, I.L. & Mann L. (1977). *Decision making: a psychological analysis of conflict, choice and commitment*. New York: Free Press.

Kahnemann, D. & Tversky, A. (2000) (ed.). *Choices, values and frames*. Cambridge: Cambridge University Press.

Keeney, R.L (1992). *Value-focused thinking.* Cambridge MA [etc.]: Harvard University Press.

Keeney, R.L & Raiffa, H. (1976). *Decisions with multiple objectives, preferences and value tradeoffs.* New York [etc.]: John Wiley & Sons.

Kepner, C.H. & Tregoe, B.B. (1981). *The new rational manager*. Princeton: Princeton Research Press.

Koopman, P.L. (1980). *Besluitvorming in organisaties*. Assen: Van Gorcum.

Lindblom, C.E. (1959. The science of muddling through. *Public Administration Review, volume 19*, 79-88.

Lipschitz, R, & Bar-Ilan, O. (1996). How problems are solved; reconsidering the phase theorem. *Organizational Behavior and Human Decision Processes, volume 65,* 48-60.
Lootsma, F.A. (2002). *A revision of basic concepts in multi-criteria decision analysis and the construction of a framework for decision making*. Unpublished paper.
Póyhónen, M & Hámáláinen, R. (1998). Notes on the weighting biases in value trees. *Journal of Behavioral Decision Making, volume11,* 139-150.
Rittel, H.W.J. & Webber, M.M. (1973). Dilemmas in a general theory of planning. *Policy Sciences,* Volume 4, 155-169.
Rubinstein, A. (1998). *Modeling bounded rationality*. Cambridge (MA), London: The MIT Press
Shafir, E. & Leboeuf, R.A. (2002). Rationality. *Annual Revenue Psychology, volume 53,* 491-517.
Shafir, E., Simonson, I & Tversky, A (1993), Reason-bases choice, Cognition, Volume 43, 11-36.
Simon, H.A. (1960). The executive as decision maker. In: *The new Science of Management Decision*. New York [etc]: Harper and Row.
Simon, H.A. (1976). *Administrative behavior.* New York: The Free Press, London: Collier.McMillan Publishing.
Simon, H.A (1979). Information Processing Theory of Human Problem Solving. In: Estes, W (ed.) *Handbook of Learning and Cognitive Processes* (Vol. 5). Hillsdale: Lawrence Erlbaum Associates Inc.
Skyrms, B. (1990). *The dynamics of rational deliberation*. Cambridge (Ma): Harvard University Press.
Strauss, A. & Corbin, J. (1998). *Basics of qualitative research; Grounded Theory, procedures and techniques.* Thousand Oaks CA [etc.]. Sage Publications.
Teglasi, H. & Epstein, S. (1998). Temperament and personality theory, the perspective of cognitive-experiential self-theory. *School Psychology,* volume 27, 534-550.
Timmermans, D. (1993) The impact of task complexity on information use in multi-attribute decision making, *Journal of behavioral decision making,* Volume 6, 95-111.
Van der Pligt, J, de Vries, N.K, Manstead, A.S.R, & van Harreveld, F. (2000). The importance in being selective; the role of attribute importance in attitudinal judgment. *Advances in Experimental Social Psychology, volume 32,* 135-201.
Van Riemsdijk, M.J (ed). *Dilemma's in de bedrijfskundige wetenschap*. Assen: van Gorcum.
Visscher, K. (2001). Design methodology in management consulting, P.hD-thesis, Enschede: Twente University.
Weber, M. & Borcherding, K. (1993) Behavioral influences on weight judgments in multiattribute decision making, *European journal of Operational research,* Volume 67, 1-12.
Weber, M, Eisenfúhr, F. & Winterfeldt, D. von (1988). The effect of splitting attributes on weights in multi-attribute utility measurement. *Management Science, volume 34,* 431-445.
Wilkie, W.L. & Pessemier, E.A. (1973) Issues in Marketing Use of multi-attribute attitude models, *Journal of marketing research,* Volume 10, 428-441.

# Chapter 5: Bounded rationality in non-routine decision problems
## A study on complexity reduction in attribute weighting[5]


### Abstract

Weighting attributes of alternatives during a decision process is extremely complex. Many attributes and an infinite number of weight values can be considered. Actors confronted with this problem will be boundedly rational; they will take only a limited subset of attributes and weight values into consideration. Our question is the following: is there a system (identifiable pattern) in the attributes and weight values considered in non-routine decisions? In a think-aloud experiment we found that actors generate attributes without apparent regard for comprehensiveness, redundancy and interdependence. They did their best to try and generate large numbers of attributes, but did not explicitly evaluate the chosen attributes. But they consistently followed an (always implicit) system in the choice of weight values taken into consideration. 'Boundedly rational' thus need not mean 'unsystematic'. On the basis of these findings, we propose the development of instruments for improving the weighting process.


### 1 Introduction

When making a decision, i.e. choosing between several alternative objects (goods) or courses of action (Van der Pligt & Koele, 1993), establishing the relative importance of the attributes in which the attractiveness of each of the alternatives is expressed, is a challenging task. If the attractiveness of each alternative is expressed in a utility function (Keeney & Raiffa, 1976), the actor concerned should establish:

1: The attributes (characteristics of the alternatives) to be evaluated;

2: The weight (importance) of each attribute;

3: The score of each alternative on each attribute.

The first two issues are related: should the weight of an attribute be zero or not (issue 1) and, if not, what should the weight be (issue 2)? This thesis does not deal with the third issue, since it is not a weighting but a measurement issue.
A decision often concerns a large number of attributes. Think of, for example, the acquisition of a new minibus for a transport company, the practical example studied in this research. Performance (number of passengers, range, maximum speed), cost (acquisition and operating cost), passenger comfort (legroom, climate control, internal noise), environmental performance (external noise, emissions), there are potentially hundreds of attributes to consider. And each attribute can have an infinite number of

[5] This chapter is based on an article with the same title, which will be submitted for publication to the Journal of Economic Behavior & Organization.

weight values (not important, very important, more or less important than other attributes, values on a scale from 0 to 1 or 0 to 10, continuous or discrete values and so on). So, choosing attributes and assigning weights can be a daunting task, especially in the case of non-routine decisions.

This chapter concerns the ways in which actors participating in organizational decisions handle the complexity involved in assigning weights to attributes in non-routine decisions. The actual assignment of weights is called *weight judgment* or *importance judgment*. The mental processes leading to the importance judgment are together called the *importance assessment* process or *weight assessment process*. This chapter focuses on the latter from the perspective of bounded rationality.

An important limitation of our research is that we confine ourselves to situations where weighting of attributes is performed. In some choice strategies, weighting is not necessary (Bettman, Luce & Payne, 1998, see Section 4.2 of Chapter 1). Also, we specifically look at situations where the weighting of attributes is done before alternatives are evaluated. This need not always be the case (Bettman, Luce & Payne, 1998). When important decisions have to be made within an organizational context, like the acquisition of a capital good, the conditions on which our research is focused are likely to be in order.

**Rationality and bounded rationality: the problem statement**

When taking a decision, a rational actor would, ideally, consider all possible alternatives and their consequences before choosing the best solution (Simon, 1976). When weighting attributes, the alternatives and consequences are combinations of attributes, weights and arguments for the chosen weights) (see Chapter 2 for an elaboration). A decision-maker displaying this behavior is often called ‘Rational Man’ (Rubinstein, 1998). ‘Rational Man’, however, is little more than a theoretical construct. Real life is different. In all but the simplest decision problems only a limited number of alternatives and their consequences are taken into consideration, thus greatly reducing complexity and hence cognitive workload. Simon (1976) called this phenomenon ‘bounded rationality’.

In sum, we define’ rationality’ in this chapter as ‘taking all possible attributes, weights and arguments into account’. Bounded rationality means: ‘taking only a limited subset of attributes, weight and arguments into account’.

This chapter deals with both conscious and unconscious forms of bounded rationality. Furthermore, we look for both explicit and implicit forms of bounded rationality. The former show in statements by actors, the latter in, for example, their behavior or in the end result of the importance assessment process (the importance judgment).
Our problem statement is, therefore:

> *Which specific (explicit or implicit, conscious or unconscious) forms of bounded rationality can be distinguished in importance assessment processes by individual laymen actors concerning non-routine problems within an organizational context, and to what extent do these forms of bounded rationality occur?*

The reasons why we focus on individual laymen and non-routine decisions are explained in Section 4. We choose to focus on organizational decisions, because of their practical relevance for research on management.

**Scientific and practical relevance**

Bounded rationality is more of a general concept than a theory from which specific expectations can be derived (Aumann, 1997). Although the concept has been modeled theoretically in a general way (see, for example, Munier *et al*, 1999; Rubinstein, 1998), this theory has not been tested experimentally on real life importance assessment processes, to our knowledge. Investigating how bounded rationality manifests itself in practice adds to our understanding of this phenomenon. Importance assessment processes are especially interesting because they cannot be solved by logic alone. Personal preferences and values are inevitably involved. This is not the case in many types of highly structured, logical problems that have been received attention of researcher in problem solving and bounded rationality. Furthermore, research on bounded rationality has the potential to explain certain aspects of importance assessment processes, such as their incremental structure (see Chapter 3).

The practical relevance of our study is that, if specific forms of bounded rationality are identified, instruments could possibly be developed for increasing the quality of importance assessment processes. Our assumption here is that either an increase in rationality (less complexity reduction) or a more conscious form of bounded rationality (choosing a way of complexity reduction that is best suited for the problem at hand) will increase the quality of the importance assessment process compared to uncontrolled bounded rationality. We believe that instruments for stimulating rationality can, at the very least, help actors operating within an organizational context in justifying their reasoning behind importance judgments to others. Keeney & Raiffa (1976) stress the use of formal analysis in decision problems in what they call ‘the reconciliation process’ in which several actors try to reach a common decision. Although they are aware that ‘in some circumstances, the more confusion that abounds the easier it is to establish a compromise’, they maintain that ‘familiarization can sometimes facilitate reconciliation’ (p10). We believe the same is true for importance assessments.

The structure of this chapter is as follows. We start with some theoretical observations about bounded rationality, focusing on importance assessments (Section 2) Next we address the research method and briefly describe the think-aloud experiment (Section 3). Subsequently we identify the measured indicators (Section 4). Then we proceed towards the results from two perspectives from which we look at importance assessment processes (Sections 5 and 6). Some concluding remarks are made in Section 7.

# 2 Theoretical background and expectations

## 2.1 Introduction

How could bounded rationality manifest itself in importance assessment processes? Earlier, we identified three elements of a utility function: attributes, their scores and their weights. We also, in Section 1, stated that actors might give arguments for taking attributes in consideration or for assigning certain weights. The weights of the attributes constitute the importance judgment; the scores and the utility function constitute the application of these weights to a decision problem. We confine ourselves to attributes and weights. Two mutually independent perspectives are used.

Firstly, we examine the way in which actors limit the number of attributes taken into consideration, either consciously or unconsciously. The question is: given that actors only take a limited number of attributes into consideration (see Section 1), can a system be observed in those attributes that are taken into consideration?

Secondly, we want to find out in which way actors reduce the number of weight values they take into consideration. So, the question here is: given that only a limited number of weight values are taken into account (see Section 1), can a system be observed in those weight values which are taken into consideration?

The two perspectives are addressed in the next subsections. Section 2.2 is devoted to the first perspective, Section 2.3 to the second. Other forms of bounded rationality can easily be imagined, but they are not addressed in this chapter in order to limit the number of issues covered. For example, limitations of the number of possible utility functions taken into consideration is also a form of bounded rationality and may have influence on the weights assigned, but it is not addressed. Also, we do not look at the reduction of the number of arguments actors may use to justify their weight judgments, even though this is definitely an interesting topic, but too large to cover here.

## 2.2 Bounded rationality and the choice of attributes

### 2.2.1 Comprehensiveness, redundancy and interdependence

When choices need to be made, it is self-evident that all relevant attributes and sub-attributes (all attributes and sub-attributes that influence the probability of any alternative being chosen) should be included in the importance judgments on the basis of which alternatives are evaluated. In other words: the importance judgments should be comprehensive (Keeney & Raiffa, 1976). For convenience, in the rest of this chapter, we as a rule speak of ‘attributes’ and not of ‘sub-attributes’. Comprehensiveness is a concept referring intrinsically to framing of the importance assessment problem with abstraction from the concrete alternatives perceived to be available. Hence the focus is on potential

importance of attributes ignoring for the moment whether such potential materializes in discrimination between the available alternatives. Given the bounded rationality of actors, comprehensiveness is obviously not easy to attain. At first sight comprehensiveness likely increases the number of attributes to be taken into consideration, therefore increasing complexity. An important tool for complexity reduction in relation to comprehensiveness might be the use of mental models. We will come back to that later.

On the other hand, for the sake of reducing complexity there should be no redundancy (Keeney & Raiffa, 1976). That is to say: attributes and sub-attributes that are not important should not be taken into consideration. Redundancy may refer to looking at attributes in view of the available alternatives. For example, if all types minibuses that a transport company can choose from have maximum speeds way over the speed legally allowed in the area where the buses are to be operated, 'maximum speed' is a redundant attribute. It will not influence the attractiveness of any minibus available. But, like comprehensiveness, redundancy is also an issue independent of the available alternatives. For example, the friendliness of the person from whom the minibuses are to be bought may well be considered redundant; even without knowing which types of minibus are on the market. After all, it can be said that none of them will be any more attractive because of a friendly salesperson. Redundancy increases cognitive workload and may thus increase the chance of reasoning errors and thus incorrect weights. If complexity is reduced by not taking all attributes into account (thus possibly limiting comprehensiveness), redundant attributes should logically be the first to go.

Interdependence between attributes refers to a common component in two or more attributes, usually an underlying attribute that influences the interdependent attributes in a correlated way. There are several methods to cope with interdependence, but the simplest way is to avoid it. Hence, in that view attributes and sub-attributes that are interdependent should be partly left out or redefined in order to avoid 'double-counting' (Vincke, 1992). Double-counting increases complexity because of the need to cater for it. For example, if two attributes overlap, three weights are required instead of just two (the weighs for the independent part of each attribute, plus the weight for the common part).
So, actors having to define a set of attributes for a decision have to make sure that:

- all attributes that may influence the decision are taken into account (ensuring comprehensiveness);
- no attributes are taken into account that do not influence the decision (avoiding redundancy);
- attributes that are taken into account are independent of each other.

In the next subsection, we will elaborate on the nature and role of comprehensiveness, redundancy and interdependence. This leads to a discussion about the ways actors can cope with them.

### 2.2.2 The role of comprehensiveness, redundancy and interdependence in importance assessment

How does one know that one has been comprehensive, that one has included all the relevant attributes? There is no objective way of being sure, but one can increase the likelihood of being comprehensive. At the goal level, according to Keeney and Raiffa (1976), those goals should be included that would alter the outcome of the decision if they were omitted (see Hammond, Keeney & Raiffa (1998) and Keeney (1992) for elaborate examples). For attributes, this would mean that a set of attributes is comprehensive if all attributes that can potentially influence the decision (in the case used in our research: the choice of a type of minibus) are included in the set. So, if the number of passengers a minibus can carry makes a difference, it should be taken into consideration and thus get a weight. This attribute seems straightforward when choosing a minibus, but there may be attributes that are relevant and still are forgotten. For example, if the attributes are generated during a heat wave, nobody may think of the ease with which a minibus can start in cold or wet weather as an important attribute.

By looking at the potential effect on comprehensiveness of including attributes, abstracting from the concrete alternatives, the problem is decoupled from the actual scores of alternatives. Different scores on an attribute lead to different utilities of the minibuses, so that it may become clear only when all attributes are identified which attributes influence the eventual choice. Variations in attribute scores may or may not lead to different rankings of utilities (attractiveness) of the various types of minibuses; this cannot be determined beforehand, but in a next stage of the importance assessment process: that of addressing redundancy.

Redundancy of an attribute means that taking the attribute into consideration does not alter the eventual decision. It is thus the complement of comprehensiveness. ‘Redundant’ is not the same as ‘unimportant’. In the example used earlier: ‘maximum speed’ may be potentially important, but it does not alter the decision. If the concrete alternatives as considered show no discrimination with respect to this attribute, since they all fall in the category satisfactory, say in between 100-110 km/h, this makes “maximum speed” redundant. But it is still important, since the buyer may have liked to have a minibus with a higher maximum speed (at or above the legal speed limit of 120 km/h). Analogously, ‘reliability of the garage selling the minibus’ may be redundant because all garages under consideration in the alternatives have an excellent reputation. Another possibility of redundancy is that an attribute is so unimportant that any alternative should have an unrealistically high score on the attribute to even be considered as a choice.

An extreme example of redundancy is the case when there are only two attributes to consider, like safety and comfort in this experiment, and all minibuses are adequate in terms of safety, because of governmental safety requirements. Then it seems logical do decide only on comfort. As another example, ‘friendliness of the person selling the minibus’ may be left out immediately in terms of comprehensiveness because it is unimportant anyway (it does not pertain to the bus itself). However, reliability of the garage selling the minibus might be incorporated due to its potential effect on safety.

Redundancy of attributes will often only become apparent when the alternatives to be chosen from are known. In the example above, ‘maximum speed’ only becomes redundant when it is clear that all alternatives fall in the same category of maximum speed, or that they all have maximum speeds way above the legal speed limit. On the other hand, some attributes can be considered redundant without taking the alternatives into consideration. The friendliness of the person selling the minibus is an example.

Taking more attributes into account than is logically necessary has two consequences. Firstly, avoiding taking redundant attributes into account is a very systematic, effective form of complexity reduction by means of bounded rationality. Secondly, redundancy makes thinking about weights and about interdependence (see below) more complex because more attributes have to be processed.

What is the relationship between comprehensiveness and redundancy?
Comprehensiveness is: presence of relevant attributes. Redundancy is: presence of unnecessary attributes. Comprehensiveness is to be strived for, whereas redundancy is to be avoided. Comprehensiveness is sought without taking the available alternatives into consideration, redundancy may (but need not) depend on the available alternatives. Comprehensiveness should be addressed first, redundancy thereafter. After all, if an attribute is not considered, its redundancy will not be assessed.

Interdependency of attributes implies that their scores cannot be varied independently. According to Vincke (1992), attributes should be chosen that are independent from each other as much as possible by reducing the number of attributes or re-definition of the attributes, so that common causal factors are avoided. Suppose an actor considers only ‘smoothness’ (of the way the minibus moves over perhaps bumpy roads) and ‘seat quality’ as sub-attributes of comfort and he assigns a weight of, say, 0,6 to ‘smoothness’ and 0,4 to ‘seat quality. It is conceivable that ‘smoothness’ also encompasses ‘seat spring quality’ as a sub-attribute, besides an independent sub-attribute, say ‘car suspension quality’. Suppose that their weights are both 0.5. ’Seat quality’ consists of ‘seat spring quality’ and ‘seat cover textile quality’. Again suppose that their weights are both 0.5. All attributes and sub-attributes have potential scores in the range [0,1]. So, ‘seat spring quality’ is a sub-attribute of both ‘smoothness’ and ‘seat quality’ making the two interdependent. One way out of this problem would be to consider the three sub-attributes as independent attributes constituting comfort, with weights for ‘car suspension quality’, ‘seat spring quality’ and ‘seat cover textile quality’ equal to 0.3, 0.5 and 0.2, respectively. (‘Car suspension quality’ gets a weight of 0.5 x 0.6, ‘seat spring quality’ gets a weight of 0.5 x 0.6 + 0.5 x 0.4, and ‘seat cover textile quality’ gets a weight of 0.5 x 0.4). Another way would be to redefine ‘seat quality’ as encompassing all effects of ‘seat spring quality’ plus ‘seat cover textile quality’ and re-defining smoothness as ‘car suspension quality’ only. The respective weights would then be 0.7 and 0.3 (0.5 x 0.6 plus 0.5 x 0.4 plus 0.5 x 0.4 and 0.5 x 0.6) in comfort. The sub-attributes of ‘seat quality’: ‘seat spring quality’ and ‘seat cover textile quality’, would have weights 5/7 and 2/7, respectively (0.5 x 0.6 plus 0.5 x 0.4 and 0.5 x 0.4, both divided by the weight of their main attribute (‘seat quality’) of 0.7).

The clue is that attributes without interdependency are much more clear. There are several ways to deal with interdependence, but at least if interdependence exists it should be identified and the logical way to deal with it should be addressed.

### 2.2.3 Handling comprehensiveness, redundancy and interdependence: expectations

Assuming subjects are aware of comprehensiveness, redundancy and interdependence issues (and we already stated that being aware of these issues is relevant in the context of bounded rationality), they can handle causal and overlapping relationships between attributes by making 'mental models'. Making some sort of model can help in achieving comprehensiveness without having to consider an infinite number of attributes, most of which will likely turn out to be redundant.

De Boer (1998) explores several more or less formalized methods to make such models. The simplest is 'cognitive mapping' (Warren, 1995). According to De Boer, 'In its most simple form, a cognitive map consists of a network of cause-and-effect relationships between factors in the problem situation on hand' (De Boer, 1998, p 108). A cognitive map may concern consequences of possible alternatives but also the factors (attributes of the alternatives) that cause these consequences. De Boer discusses several other methods, such as Strategic Choice and the Influence Diagram. The basic idea remains the same. What matters is whether actors build models of the relationships between attributes or not, thereby ensuring a maximum level of comprehensiveness while reducing complexity by taking a limited number of attributers into account. Figure 1 illustrates a hypothetical cognitive map, based on de Boer (1998), but adapted for the subject of this thesis. It displays some assumed causal relations between the environment and the importance of safety versus comfort.

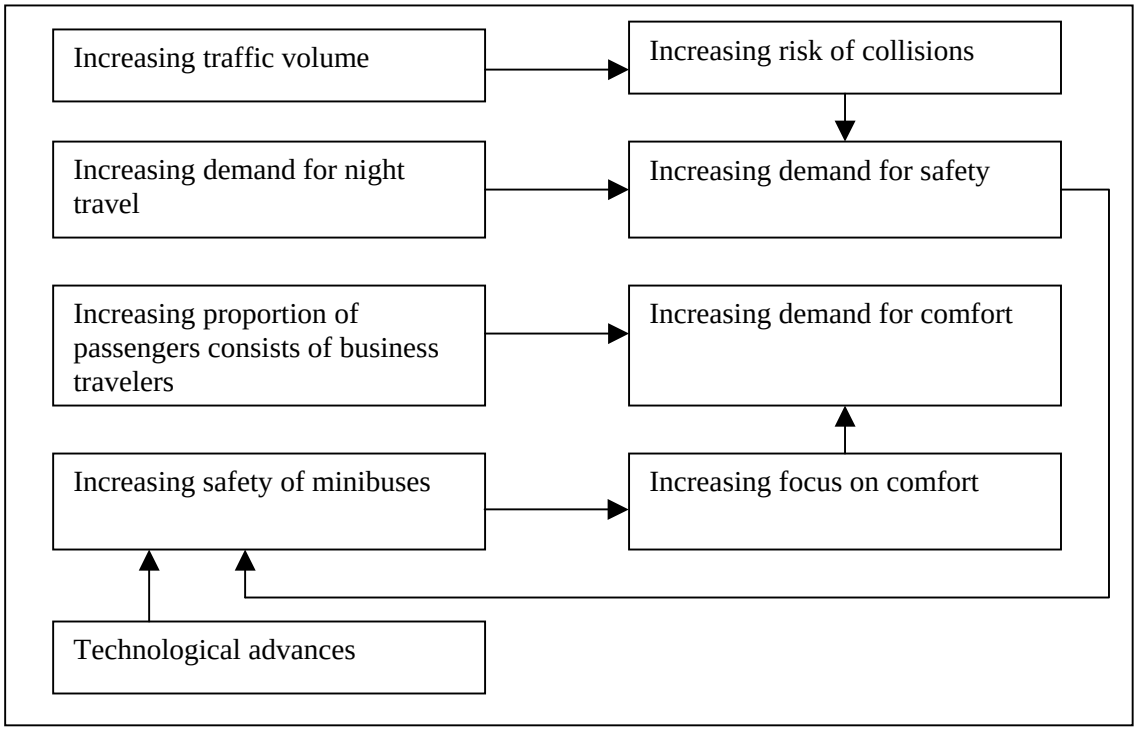



*Figure 1: Example of a cognitive map: the relation with the environment.*

Until now, we have either stated that actors should be aware of the necessity to consider comprehensiveness, redundancy and interdependence, or that they handle these issues without being aware of them, intuitively. This is an assumption, based on a normative notion. But is it a reasonable expectation to assume that they are actually aware of these phenomenons? Given the lack of structuring of the importance assessment process (see Chapters 3 and 4), we would expect not. Also, many authors have given reasons for actors' lack of rationality (see, for example, Simon, 1976, Arthur, 1994, Kahneman, 1994, Loewenstein, 1996, Simonson, 1990, Shafir, Simonson & Tversky, 1993, Levine, Halberstadt & Goldstone, 1996, Kaufmann, 1999, Shafir & LeBoeuf, 2002; Schwartz, 2000, Toda, 1980 and Damasio, 1995). Reasons that could also pertain to a lack of system within bounded rationality include: a lack of information-processing capacity, a lack of consistency in reasoning, possibly because of difficulties to verbalize elements of the problem, a limited ability for deductive reasoning, and environmental and emotional factors. Although, to our knowledge, these reasons for a lack of rationality have never been used to study to what extent there are systematic patterns in attributes and weight values taken into consideration in boundedly rational behavior, is seems reasonable to expect that they play a role.

On the other hand, given the fact that actors devote a disproportional amount of attention to the identification of attributes (see Chapter 2), it is not unreasonable to expect that they will, even if by accident, be confronted with issues of comprehensiveness, redundancy and interdependence. Literature indicates that actors can make several sorts of cost-benefit analysis when searching for information (see, for example, Saad & Russo, 1996).

It is entirely possible that actors identifying attributes perceive a need for some sort of stopping criteria, and hence aim for a sufficient level of comprehensiveness before deciding enough attributes have been identified. Or they may feel that if they, after a while, identify only new attributes that are either redundant or interdependent, it is time to stop identifying attributes. The limited information capacity of actors (see, for example, Miller, 1956) can be eliminated as a factor by assuming that actors can make notes during the importance assessment process.
The effort devoted to the identification of attributes could be seen as a form of 'framing'; deciding what to attend to and what to ignore (Schón, 1985, as cited by Akin (1994). Framing can take many forms, such as re-formulating problems, changing the 'rules of the game' so that problems with no logical solution become solvable, interpreting concepts in a certain way, like 'win' or 'loss' and looking at a problem from a certain disciplinary or normative point of view. At least the third form has also been observed even with laymen (see, for example, Beach, 1993; Beach, Puto, Heckler, Naylor & Marble, 1996;; Highhouse, Paese, & Leatherberry, 1996; Kahneman & Tversky, 2000; Polzer, Stewart, & Simmons, 1999). So it seems reasonable to expect framing by actors performing importance assessment processes. We do not know whether this framing will take the form of attending to comprehensiveness, redundancy and interdependence, but it certainly is a possibility.

Framing is especially important with respect to comprehensiveness. Redundancy and interdependence of attributes can to a large extent (although not completely) be logically determined as long as the definitions of the attributes are clear (which in itself involves framing). However, whether a set of attributes is comprehensive does not only depend on the definitions of the attributes, but also on the frame of reference of the actor; the concept of, in our case, 'safety' and ' comfort'. For example, if an actor has a narrow idea of the concept of 'comfort', in the sense of 'comfort of the trip in the minibus' the attribute of comfort that a client can be picked up at his front door (which a train cannot do) need not be taken into consideration. But if an actor has a wide concept of 'comfort', not taking into account this attribute can be seen as a lack of comprehensiveness. After all, some buses may be so large that they cannot reach the homes of some clients, while others can.

A third reason for expecting actors to be aware of the issues of comprehensiveness, redundancy and interdependence is that they tend to display high confidence in their importance judgments, even if they are not experts in the field under consideration (see Chapter 3). One of several possible explanations for this could be that they are convinced of the soundness of their reasoning. 'Soundness' need not mean 'rationality', but it can also mean internal consistency and proper argumentation, for example for not taking certain attributes and weight values into consideration. Of course, other explanations for the observed high confidence of actors performing importance assessment processes are also possible (see Chapter 3).

So, is the glass half empty or half full? There is no need to make a judgment call at this point. We choose to formulate some expectations in an affirmative way (actors take comprehensiveness etc. into consideration), for one simple reason. Looking for

something that is there seems a more natural way of enquiry than looking for something that you don't expect to be there. We elaborated on this philosophy in Chapter 4. So, we look for signs that actors take comprehensiveness etc. into account, and not for a lack of those signs. In order to avoid misunderstandings: if an actor says that comprehensiveness is irrelevant, he or she takes it into consideration; it is not neglected. However, we will also formulate the complement of each expectation, in order to do justice to the arguments against identifiable patterns in bounded rationality as addressed above. If expectations of this sort are confirmed this can be interpreted as a stronger form of the incremental approach as encountered already in Chapter 4.

Based on the above, we propose the following expectations.

*Expectation 1A:* *Given their bounded rationality, actors try to make certain (explicitly or implicitly) that the set of attributes they use is comprehensive.*

*Expectation 1B:* *Given their bounded rationality, actors are not concerned about the comprehensiveness of the set of attributes they use.*

*Expectation 2A:* *Given their bounded rationality, actors try to make certain (explicitly or implicitly) that the set of attributes they use is non-redundant.*

*Expectation 2B:* *Given their bounded rationality, actors are not concerned about the redundancy within the set of attributes they use.*

*Expectation 3A:* *Given their bounded rationality, actors try to make certain (explicitly or implicitly) that the attributes they use are independent.*

*Expectation 3B:* *Given their bounded rationality, actors are not concerned about the interdependence of the set of attributes they use.*

The handling of attributes was the first of our two perspectives concerning bounded rationality in importance assessment processes. The confirmation or non-conformation of these expectations is discussed in section 5. The second perspective, reducing the number of weight values, is discussed in the next section.

### 2.3 Bounded rationality and the choice of weight values

Weights can have any positive value imaginable. So, there are a limitless number of weight values to choose from. This is a humanly impossible task. We take for granted

that actors seek means of reducing the number of weight values to be taken into consideration. This is in line with Simon's (1976) concept of bounded rationality.

If actors reduce the number of weight alternative, then the question remains whether this is done in a systematic fashion. Do subjects consider weight values that they arrive at by accident or do they, perhaps without knowing it, have a pattern of weight values to consider? Patterns could be, for example: a limited range of values, or only specific values, like increments of 1, 0,1 and the like. Perhaps some subjects resort to anchoring; taking a certain value (which might have nothing to do with the problem at hand) as a reference figure. The phenomenon of anchoring has been observed in various estimation processes (see, for example, Chapman & Johnson, 1999 and Wong & Kwong, 2000).

Another way to reduce the number of weight values is to lower the measurement level of the weights to the binomial or ordinal level. At the binomial level, an attribute is either important or not important. At the ordinal level, attributes may be slightly important, rather important, very important, more or less important than other attributes, and the like. The number of ordinal qualifications can be quite large, but for all practical intents and purposes it will not be infinite. For efficiency reasons this issue is discussed only briefly in this chapter (see Section 6.2).

As in the previous section, subjects can be explicitly or implicitly boundedly rational. The arguments in favor of and against an identifiable pattern of weight values are the same as those mentioned in the previous section, so here we also provide an expectation for both possibilities.
Taking for granted that bounded rationality is in order (a limited number of all possible weight values is taken into account) our expectations are:

*Expectation 4A:* *There is an identifiable structure in the set of weight values that subjects take into consideration.*

*Expectation 4B:* *There is no identifiable structure in the set of weight values that subjects take into consideration.*

What is 'identifiable' will become clear in Section 4.2. As for structure we distinguish between two aspects: the scale as used and the method for positioning a weight on that scale.

If an identifiable structure for the set of weight values is used it is interesting to analyze in more depth whether some systematic approach is underlying the choice of the set of weight values. Here we address analogous issues for the set of weight values as encountered in the previous sub-section for the choice of the set of attributes. The counterparts of attribute comprehensiveness, attribute redundancy and attribute interdependence will be weight scale comprehensiveness, weight value redundancy and interdependence of identifiable weight value structures. These concepts are defined in analogy to their counterparts, as follows:

| | |
|---|---|
| Weight scale comprehensiveness | Evaluation of the discriminative power of the potential set of weight values in view of their purpose: sufficient power to describe the importance of attributes, while bounding the potential accuracy to what is assessed to be necessary given the general context |
| Weight value redundancy | Leaving out weight values from the potentially useful set in view of the actual scores of the attributes as used, in a systematic way (e.g. in relation to uncertainty in actual scores) |
| Interdependent weight value structures | More than one identifiable weight value structure is used and the combination of the various structures can be described in a systematic way |

*Table 1: Characteristics of weight values*

Now, analogous to the expectations formulated for the choice of the set of attributes we formulate the following expectations for the choice of the set of weight values:

*Expectation 5A:* *Given their bounded rationality, actors try to make certain (explicitly or implicitly) that the set of weight values they use is comprehensive.*

*Expectation 5B:* *Given their bounded rationality, actors are not concerned about the comprehensiveness of the set of weight values they us.*

*Expectation 6A:* *Given their bounded rationality, actors try to make certain (explicitly or implicitly) that the set of weight values they use is non-redundant.*

*Expectation 6B:* *Given their bounded rationality, actors are not concerned about the redundancy within the set of weight values they use.*

*Expectation 7A:* *Given their bounded rationality, actors try to make certain (explicitly or implicitly) that combinations of identifiable weight value structures (scale and perhaps positioning method) are independent.*

*Expectation 7B:* *Given their bounded rationality, actors using various identifiable weight value structures are not concerned about the interdependence of the structures of weight values they use.*

The confirmation or non-conformation of these expectations is discussed in section 6.

## 3 The research method

Only a general outline of the research method is given. The method is extensively discussed in Chapter 2.

Obviously, our research should be seen as theory-building, not theory-testing. Therefore, we needed a research method with which we could investigate many variables, without knowing precisely which variables would turn out to be relevant for our problem statement, and what the relationships between the variables might be. Also, we needed a method in which we could 'look into the subjects' minds'. In order to gain insight in cognitive processes of our subjects, we used the think-aloud method. This is a good method for analyzing cognitive processes (Ericsson & Simon, 1993). Methods such as choice experiments and process tracing show the results of cognitive processes but not the processes themselves, while retrospective reporting methods, such as interviews and diaries, leave too much room for interpretation of the cognitive processes by the subjects themselves and are vulnerable to lapses of memory. Besides, because there was no model of the importance assessment process, the formulation of interview questions would have been very difficult.

### Sample and assignment

18 undergraduate students of the University of Twente were given an individual assignment based on a fictional case. University students may be assumed to have enough analytical abilities to perform the assignment satisfactorily, without having the knowledge and skills in this area that would enable them to rely on previous experience of importance assessments. Hence, the danger that they give weights based on previously obtained knowledge is minimized.

The assignment consisted of supporting the acquisition process of new minibuses by a local company. The subjects were asked to establish the importance of two characteristics of the to-be-acquired minibuses *vis-à-vis* each other, and were asked to imagine that they would be advising the management team during the acquisition process. The attributes, *safety* and *passenger comfort*, were chosen to prevent comparability by some readily available algorithm or heuristic or easy expression in a common denominator such as money. The information that was supplied included a brochure of the company, a leaflet explaining the decision context and two brochures on minibuses; one on a Volkswagen and one on an Opel. The latter enabled the subjects to get familiar with the specific capital good to be acquired. It was made clear that these examples of minibuses did not mean that the subjects had to make a choice between them.

**Procedure**

The respondents were asked to think aloud during the assessment process. The general guidelines for think-aloud studies given by Ericsson & Simon (1993) were followed, including a practice session to familiarize the subjects with the think-aloud strategy. All verbal information given by the respondent was recorded and typed out literally. After completion of the assignment, a short interview was conducted. In total, each session lasted for a maximum of two hours, for which the subjects were paid 20 Euro. Two pilot sessions were conducted, which led to some minor adjustments of the assignment.
Two kinds of analyses have been performed using the typed out protocols:

1: A largely qualitative analysis according to the general rules of the 'Grounded Theory' approach (Strauss & Corbin, 1998);

2: A more quantitative analysis based on a formal coding scheme that was designed on the basis of the qualitative analysis. Two coders performed the coding activities. Although they worked independently of each other, during the coding of the first six protocols weekly meetings were held to discuss general coding issues in order to enhance its reliability. The refinements to the coding scheme were retrospectively applied independently by the coders. The overall Cohen's Kappa (Baarda & De Goede, 2001) for inter-rater consistence was .97 over a total number of verbal segments of 7253.

The expectations formulated in this chapter are partly analyzed in quantitative terms. Note that we talk of expectations and not of hypotheses, since the number of subjects was too small to test hypotheses with any statistical validity. For expectations pertaining to the frequency of behavior to be observed, we use two thresholds. The most rigorous threshold concerns the behavior of the entire research population. This assumes that all subjects are more or less identical and behave in the same way. In this 'extreme case', we accept an expectation if 80% of the subjects (15, rounded off upwards) behaves accepting the expectation. This is a somewhat arbitrary threshold, but with the small number of subjects involved, 80% seems an ambitious but realistic threshold.

It is obvious that generally we cannot expect all subjects to always behave in identical ways. Some will behave according to our expectations, some will not. Still, in this case we want to know what the general trend is. Does the majority behave as we expect? We accept a majority expectation if 12 or more subjects behave according to the expectation (a margin of three subjects above the lower limit of 9 for a draw and also three below the acceptance limit of 15 for the 'extreme case'. This procedure avoids a nasty problem. In the extreme case, an expectation is accepted if 15 to 18 subjects behave according to the expectation. But this reasoning means that the complement of the expectation can only be accepted if 15 to 18 subjects show behavior compliant with this complementary expectation, Thus if more than 4 but fewer than 15 subjects behave according to the expectation or its complement, neither expectation can be accepted. Out of 19 possible results (0 to 18 subjects can behave according to an expectation), 10 values do not lead to a clear conclusion. In the majority case, this number is only four (if 8 to 11 subjects display a certain behavior), but the conclusion is weaker, of course.

For all expectations we use several indicators. While all indicators taken together determine whether an expectation is accepted or rejected, the 'weakest' indicator determines the qualification (extreme or majority case). In this way, we minimize the risk of unjustified acceptation of the expectations

**Limitations**

The limitations of this research were discussed in Chapters 1 and 2 and summarized in Section 4 of Chapter 3. We will not repeat them here. Suffices it to say that our data do not readily support our conclusions being extended to real-life contexts. Our aim was to establish the structure of the importance assessment process uncontaminated by factors like group dynamics and previous experience of our subjects. The results can be used as a basis for future research introducing factors that manifest themselves in real-life situations. We expect our results to be mirrored in real-life situations to some extent. After all, it is not uncommon for individuals (albeit often with a certain degree of expertise) to make importance assessment processes under circumstances similar to those in our research. But we did not research factors such as 'contamination' by previous importance judgments, which may alter the importance assessment process. So, even if we expect our results to be applicable to real-life situations, we cannot say to what extent and in what form.
Our research pertains only to decision contexts where there is explicit weighting and where the importance assessment process is separated from the evaluation of alternatives. Whenever we make propositions about importance assessment processes in real-world contexts, these propositions can only be validated by further research.

## 4 Measured indicators

All indicators are essentially thoughts (cognitions) verbalized by the subjects in our research. The way subjects were induced to verbalize thoughts is explained in Section 3.

### 4.1 Comprehensiveness, redundancy and interdependence of attributes

*As for comprehensiveness* we can merely observe whether subjects concern themselves, explicitly or implicitly, with comprehensiveness. We also look at the number of subjects that state explicitly that they want to define safety and comfort (regardless of whether they actually did so or tried to do so), assuming that comprehensiveness as the subjects see it is served when explicit definitions are used. In Chapter 4 we saw that definitions were hardly used. Now we look not only at whether definitions were actually used, but also whether subjects saw the relevance of using definitions, even if they got no further than good intentions. Furthermore, we observe whether they give an indication as to why they think, at a certain moment, that they have named enough sub-attributes for safety or comfort to be able to start assigning weights. We leave out statements like 'I think I have found enough attributes', because it is seldom clear what they mean: They might mean: 'I have spent enough time spent on searching for attributes', 'I have enough attributes to

work with', 'I can't think of anything else for the moment', or even 'I am fed up trying to find attributes'. So, the statements we take into consideration have to have an explicit or implicit reference to comprehensiveness.
For *redundancy*, we have two indicators. Firstly, we look for any statements that indicate that the subject is concerned with redundancy. For example, a subject might say: "I have taken care that I don't waste time on attributes that do in the end not alter my decision". Secondly, we measure if any attributes have been left out because they were redundant. There can be several reasons for this. For example: all likely alternatives would probably have equal or similar scores on the attribute, an attribute did not achieve the minimum level of importance, or other reasons. This indicator had to appear more than just incidentally. If a subject mentioned only once or twice that a certain attribute was not included for reasons that pertain to redundancy, we do not see this as serious consideration for redundancy.
In order to establish whether subjects' attributes are truly *independent*, we should know the exact definitions of all attributes used. Since subjects were quite happy decomposing safety and/or comfort in twenty or more sub-attributes, some of them re-formulated several times, it was practically impossible to go over all the definitions used. Besides, doing so would likely have resulted in a high degree of rationalization, since it would have had to be done after the experiment. So, we look at more intuitive indicators of independence. Such an indicator could well be: systematic decomposition of attributes. An example is: splitting 'safety' in 'active safety' en 'passive safety'. In Chapter 4 we looked at the actual use of systematic decomposition and saw this hardly ever occurred. So we take this result into account without elaborating on this indicator. We look at statements concerning the desire for systematic decomposition, however, even if they do not go beyond good intentions. Secondly, we look for statements by the subjects that indicated that they were aware of unwanted interdependence. A third indicator is the extent to which subjects have made mental models i.e. cognitive maps. This could be readily checked because the subjects were allowed to make notes. The experimenter retrieved these notes. This indicator may also give insight in the way comprehensiveness is achieved and redundancy is avoided.

We already stated that comprehensiveness; independence and redundancy are closely related. It could be argued that the above-mentioned indicators sometimes pertain to more than one variable. This proved not to be a problem, as is explained in Section 5.2.

We fully realize that (bounded) rationality may manifest itself implicitly, as mentioned in Section 1. For example, subjects may, while summarizing the attributes to be weighted, pay attention to comprehensiveness, redundancy or interdependence 'at a glance' without mentioning it, perhaps even without being aware of it. If bounded rationality is implied in statements the subjects make, or in their behavior, the researcher may recognize it, but obviously chances are he will not. This cannot be helped. It is a limitation of the think-aloud method, but we are not aware of any other method that would address this problem and still yield the rich data that the think-aloud method provides. In the discussion of the results we will pay attention to the possibility of implicit rationality, even if no 'hard' indicators are available. And we have one indicator for implicit attention to interdependence: the degree of interdependence manifest in the attributes eventually

weighted. This is relatively easy to check, although we cannot be sure that dependencies we see were also perceived as such by the subjects. Unfortunately, such an indicator is not possible for comprehensiveness and redundancy, for these issues depend too much on the perceptions (frames) of the subjects and on the (nearly always implicit) definitions of the attributes they use.

## 4.2 The weight values taken into consideration

The first indicator is whether the subjects make statements that explicitly or implicitly reduce the number of weight values under consideration. A subject might say, for example: "I want to assign weights on a scale from 0 to 1. Now I know that safety is more important than comfort, but I don't know yet how much more important it is. So, I only have to consider weights between 0.5 and 1". In this case, it is clear that the range of weight values has been reduced to the interval 0.5-1.0.
The second indicator is: the weight values that are actually considered. If, for example, a subject only mentions weight values of 0.1, 0.2 and 0.3, then we assume that he or she only considers weights at 0,1 intervals and not in between. We set a minimum of three mentioned weight values before we assume a systematic reduction in the number of weight values, and no weight values violating the system in the mentioned weight values should occur. In the end, one subject mentioned only two weight values and hence was not analyzed further; two subjects mentioned 3 weight values for a given scale and two mentioned four weight values. The rest either mentioned more than four weight values or did not mention any quantitative weights at all. No weight values were mentioned at any moment that violated the system that could be observed in the mentioned weight values.
The use of the above-mentioned indicators not only showed whether subjects limited the number of weight values to consider, but also identified systems the subjects used in doing this.
A third indicator is the mentioning by subjects of rules-of-thumb or other heuristics that have nothing to do with the setting of weights but provide anchoring values. An example is the famous 20-80 rule for relating concepts "not so important" and "important" to numbers 0.2 and 0.8 respectively.. Of course, this indicator can only be detected if the subject names the rule or heuristic associated with the anchor. Otherwise, we have no way of knowing whether a stated weight value is coincidental or a result of anchoring.
Statements that pertain to ordinal weights are only considered as indicators if they pertain to an identifiable scale. Subjects can use many ordinal weights like 'more important', 'really more important', 'much more important' and the like. As long as it is not clear that these qualifications are used with a minimum of consistency (i.e. that 'much more important' indeed generally means 'more important than just 'more important'', we will not use them as indicators. No situation occurred in which this indicator could be used with any guarantee of reliability.
We also excluded the weighting of the two main attributes on an ordinal scale. If the outcome of a weighting process was not more than 'safety is more important than comfort' or the other way around, without it being clear how much more important, we consider this so imprecise that it does not qualify as a meaningful way of weighting.

Anyway, it became apparent that the number of subjects behaving according to expectation 4A would have increased by only 1 if we had included this indicator.
We only consider the weight values for those attributes that are actually weighted. Some subjects indicated that some attributes were relevant (so they were not excluded) but not so important that they received weights. One could argue that for these attributes, which would have to be weighted, an infinite number of weight values is still possible. But this point is merely academic.

As for measurements of comprehensiveness of the weight scale, redundancy of the weight values and interdependency of a combination of weight value structures we use the same reasoning as discussed for comprehensiveness, redundancy and interdependence of attributes. The indicators are:

**Comprehensiveness**

1: Statements indicating concern for the power of weight scales used for expressing importance, as far as range and precision of the scale are concerned. Incidental remarks on individual weight values like 'I don't think that 'quite important' is precise enough', do not count; they say little about the used scale as a whole;

2: The range and precision of the scales actually used. Then it is inferred whether this indicates concern for range and precision.

**Redundancy**

3: Statements indicating that certain parts of the range of a used scale are not to be taken into consideration (excluding the incidental mentioning of individual weight values);

4: Statements indicating that certain categories of weight values within the range of the scale(s) used are not to be taken into consideration.

**Interdependence**

5: Statements indicating that certain combinations of scales and positioning methods are not to be taken into consideration;

6: Indications of combinations of methods used. These combinations are then evaluated with regard to interdependence.

## 5 Results: comprehensiveness, redundancy and interdependence of attributes

We will first discuss the results on the basis of verbalized thoughts (the think-aloud protocols). After that, we will look at more implicit indicators of bounded rationality.
The results are summarized in Table 2.

| Indicator | Number (percentage) of subjects where indicator was observed during assignment and in exit interview |
|---|---|
| **Indicators common for comprehensiveness, interdependence and redundancy** | |
| 1: Intention of systematic decomposition | Assignment: 3 (17%), interview: 0 (0%) |
| 2: Making a mental model | Assignment: 0 (0%), interview: 0 (0%) |
| **Indicators common for comprehensiveness and interdependence** | |
| 3: Explicit definition of attributes or stating the desire to do so | Assignment: 5 (28%), interview: 7 (39%) |
| **Indicators pertaining to comprehensiveness** | |
| 4: Statements concerning comprehensiveness | Assignment: 0 (0%), interview: 0 (0%) |
| 5: Statements concerning why enough attributes have been generated, explicitly related to comprehensiveness | Assignment: 0 (0%), interview: 0 (0%) |
| **Indicators pertaining to interdependence** | |
| 6: Relevance of avoiding interdependence | Assignment: 0 (0%), interview: 0 (0%) |
| **Indicators pertaining to redundancy** | |
| 7: Statements concerning a concern for redundancy of attributes | Assignment: 0 (0%), interview: 0 (0%) |
| 8: Leaving out attributes because they don't pertain to the decision (choice of a bus), for example, behavior of the driver | Assignment: 6 (33%), interview: 1 (6%) |
| 9: Leaving out sub-attributes that do not pertain to the assignment (for example, cost) | Assignment: 5 (28), interview: 1 (6%) |
| 10: Leaving out sub-attributes that are not important to the customer (but only to, for example, the driver) | Assignment: 1 (6%), interview: 0 (0%) |
| 11: Leaving out sub-attributes that do not attain a minimum level of importance | Assignment: 2 (11%), interview: 1 (6%) |
| 12: Leaving out an attribute because the scores for all alternatives (types of minibuses) are similar. | Assignment: 4 (22%), interview: 0 (0%) |
| 13: Leaving out sub-attributes of safety because the client doesn't notice anyway | Assignment: 1 (6%), interview: 2 (11%) |

*Table 2: Number of subjects with whom indicators concerning comprehensiveness, redundancy or independence could be observed. A distinction is made between the behavior observed during the execution of the assignment and activities mentioned in the interview afterwards (as extra information, nit used in testing the expectations).*

As is clear from the table expectations 1A, 2A and 3A have to be rejected and expectations 1B, 2B and 3B can be accepted. We will now discuss these results in more detail. The fact that subjects did not pay much attention to comprehensiveness, redundancy and interdependence while it is obvious that it would have been reasonable to do so should of course not be interpreted as a condemnation of the work of the subjects.

Our research is descriptive, not normative. The human brain works as it does, and if intelligent subjects, as university students can be assumed to be, do not pay much attention to the above-mentioned issues, it has to be taken as a given. There dedication was not in doubt; they made every effort to conduct the importance assessment process as well as possible

**Comprehensiveness**

No subject referred in any significant and recognizable way to the issue of comprehensiveness (indicator 4). No clear reason could be found why subjects stopped looking for more sub-attributes (indicator 5). Saad and Russo (1996) discuss several ways in which people who have to choose between alternatives can decide when to stop acquiring more information and make a choice. Unfortunately, these models could not be tested, for the following reasons. Firstly, the information acquisition task in our experiment was limited. There was no guarantee that all the required information could be found in the documents provided. Indeed, most subjects derived sub-attributes at least partly by creative thinking and association. Secondly, there was no identifiable moment at which the subjects stopped generating sub-attributes. Many subjects generated new sub-attributes long after they had shifted emphasis from identifying sub-attributes to weighting them. This is in itself an indication that there was no finite moment at which the subjects took the time to evaluate the comprehensiveness of their sets of sub-attributes.

8 subjects stated the desire to define ‘safety’ or ‘comfort’, either during the assignment or during the interview (indicator 3), but only one got close to actually doing so, by describing ‘safety’ as’ the number of deaths and wounded per year’. But he then proceeded to treat this definition as merely an attribute of safety, along with, amongst others, adjustable headrests (a determining factor of the number of deaths and wounded, and hence a case of interdependence). During the interviews, subjects frequently stated that they had started the assignment by defining ‘safety’ and ‘comfort’ or by naming elements that constituted these attributes, but in no occasion was any mention made of the need or desire to assure comprehensiveness or avoid interdependence.
A complication for assessing comprehensiveness is that subjects might take sub-attributes into consideration without mentioning them because they were self-evident. One subject said this explicitly, but others might have done the same. But the potential consequences of this are likely to be small, for no checks for comprehensiveness were performed anyway. And those attributes would have to be assigned weights all the same, so at some point they would have to be made explicit. So the results concerning the expectation are not influenced.

As we already concluded from Table 2, none of the indicators was observed with 12 or more subjects, so expectation 1A has to be rejected. The complementary expectation 1B: subjects are *not* concerned that the set of attributes is comprehensive, can be *accepted for the majority case*. The complement of the weakest indicator (a desire for systematic decomposition) occurs in 13 of the 18 cases during the assignment.

This result is somewhat dissatisfying. Even if not familiar with the concept of comprehensiveness, one would expect the subjects to wonder whether all, or a sufficient number of, relevant attributes were identified. Reasons for this expectation were given in Section 2.2. It is possible that the subjects considered comprehensiveness implicitly, and that they stopped looking for more attributes when they were satisfied with the attained comprehensiveness. The disproportional amount of effort that they devoted to identifying attributes points in this direction. One could view this as a parallel to the satisficing strategy in choice between alternatives (Bettman, Luce & Payne, 1998, see Section 4.2 in Chapter 1). But there are no clear indications that this was the case, and anyway unsystematic generation of attributes is at best partly effective in assuring comprehensiveness. It is likely to identify the main areas where attributes have to be sought, but large gaps may remain within those areas. If framing was in order, it did not take the form of explicitly stating systems boundaries. Other forms of framing were observed later in the importance assessment process, like taking the supposed preferences of various stakeholders (clients, the management of the transport company, the drivers of the minibuses) into consideration, but these frames of reference did not pertain to comprehensiveness of attributes. Which, if any, of the reasons for lack of rationality mentioned in Section 2.2 can explain the subjects' behavior cannot be said on the basis of our results. It is remarkable that subjects did not use readily available methods to reduce cognitive workload, like making models on paper. This would suggest that they either were not aware of the issue of comprehensiveness, or that they did not consider it to be a problem, perhaps because they were more focused on the number of attributes generated than at their quality. But we cannot be sure.

As stated earlier, the fact that subjects generally had high confidence in the adequacy of the weights they assigned (see Chapter 3) suggests some implicit concern for comprehensiveness redundancy and interdependence. But we should not confuse self-confidence with objective quality. Subjects usually generated large numbers of attributes (Chapter 3) and may have mistaken the number of attributes identified as an indicator of comprehensiveness (it can also be an indicator of redundancy). But nobody said so. Besides, the subjects knew that they had to justify their decision, so if part of their justification was that they were comprehensive, it would likely have shown. It should be stressed that 11 subjects used two sources to identify attributes: the brochures about two types of minibuses given to them, and their own common sense. This can be seen as an attempt to be comprehensive. But again, this is not enough within an organizational context. So it seems that comprehensiveness, rightly or wrongly, was not seen as a problem.

So, while subjects had high confidence in their work, we can only assume this as an indicator of the quality of their work (and hence of their attention to comprehensiveness) if we assume their self-confidence was justified. Given the incremental nature of the weighting process as observed in our study (see Chapter 4) and the lack of expertise of the subjects (see Chapters 2 and 3) we see no reason to do this. Other explanations of the confidence discussed in Chapter 3 may be more credible. The same reasoning applies to the other two aspects of handling attributes: redundancy and interdependence.

**Redundancy**

Subjects hardly paid attention to redundancy. No statements concerning redundancy were made (indicator 7). Leaving attributes out because they don't pertain to minibuses or the assignment, are not important enough in general or for the clients of the minibus company, are expected to have similar scores for all alternative minibuses or cannot be noticed by clients were methods of avoiding redundancy (indicators 8-13) used by 10 subjects (56%) during the assignment and mentioned by 4 subjects (23%) during the interview. Again, the rules mentioned in the table were not used systematically but incidentally; they were used only with a small number of eliminated attributes.

No subject weighted all the sub-attributes that he or she generated. So, logically, many sub-attributes were excluded during the weighting process. In many occasions this was done implicitly; sub-attributes were simply not mentioned any more after they had been generated. In some occasions, subjects, after having generated the sub-attributes, went through them and stated whether they were important. But that was not done in a consistent way. For example, an attribute that was 'not really important' might or might not turn out to have been eliminated. Only in the occasions mentioned in Table 2 was explicitly or implicitly stated that a sub-attribute was eliminated because it was not important enough. This can be seen as a parallel of the Elimination-By-Aspect choice strategy (Bettman, Luce & Payne, 1998, see Section 4.2 of Chapter 1). In most cases, no reason was given why this was so. Subjects just, 'checked off' the sub-attributes and eliminating was a yes/no question. If reasons were mentioned, they were varied; from not knowing what the sub-attribute actually meant to the sub-attribute failing to give any meaningful contribution to safety or comfort. Eliminating attributes implicitly can in itself be seen as an indicator of bounded rationality. It was not used as such in this study because we took for granted that our subjects would be boundedly rational (see Section 1). It is the structure of bounded rationality we are after.
A reason for eliminating sub-attributes that would be expected to be used often is that every minibus has the same or a similar score on a sub-attribute. But in practice this reason was mentioned explicitly only 5 times. An explanation is that sub-attributes on which all types of minibuses were expected to have similar scores were mostly not excluded but were give a lower weight.

The results show no significant explicit concern for or handling of redundancy. Expectation 2A has to be rejected. Its complement, expectation 2B can be accepted for the majority case.

Still, all subjects practiced an implicit form of dealing with redundancy. As stated, the minimum number of (sub-)attributes generated was 22, excluding those attributes rejected outright. The average was 19.6 for safety and 24.4 for comfort (Chapter 3) The maximum total number of (sub-)attributes eventually weighted by any subject was 17. So, far fewer attributes were weighted than were initially generated. However, the maximum number of (sub-)attribute originally taken into consideration and later explicitly rejected was 3 (for 2 subjects). Many (sub-)attributes were eliminated implicitly. This, however, cannot serve as an indicator for consideration of redundancy. It

is possible that they were consciously but implicitly eliminated, but it is equally well possible that, for example, many attributes were simply forgotten after they had been mentioned. And handling a large number of sub-attributes may make the weighting task too complex. At least one subject recognized this danger. This clearly is bounded rationality, but not necessarily concern for redundancy.
As with comprehensiveness, the reasons for the lack of attention for redundancy are unclear. The possibilities offered in Section 2.2 may be in order, but since so little was said concerning redundancy, it is not possible to draw any conclusions. It is possible that subjects were not aware of the issue of redundancy, or did not think it relevant. Also, the high level of confidence does not provide an explanation, as discussed in the case of comprehensiveness.

**Interdependence**

Three subjects performed some form of systematic decomposition (indicator 1). But the figures belie the fact that in every single case systematic decomposition was incidental (see Chapter 4). For example: a subject might decompose ‘safety’ in ‘active safety’ and ‘passive safety’ (a comprehensive decomposition) but then go on decomposing ‘safety’ in other attributes like ‘presence of safety straps’ and ‘anti-skid brakes’ without grouping these attributes under ‘active safety’ and ‘passive safety’. In this way, any benefit of systematic decomposition is lost.
While 3 subjects stated, either during the assignment or during the interview, the general desire to work systematically or in a structured fashion, no one specifically referred to or hinted at decomposition to avoid interdependence (indicator 6).
Mental models were not made, at least not on paper (indicator 2). Lists of sub-attributes were frequently made, but the establishment of causal relationships between (sub)-attributes was an exception. Subjects frequently referred in passing to statistical or causal relationships between attributes, but this hardly ever resulted in attributes being excluded and not much concern for causal relationships was shown anyway (Chapter 4). It led merely to the adjustment of the weights of attributes. Only on one occasion did a subject observe that an attribute (stopping distance) was directly, one could say deterministically, influenced by another attribute (weight of the minibus), but he then proceeded to include both attributes in the weighting process. The only clear case of concern for interdependence was when a subject stated that, of three attributes with the same function, only one of those needed to be taken into consideration when judging a minibus.

As stated in Section 3.1, we also looked at an implicit indicator of concern for independence: the degree in which the attributes eventually weighted were indeed independent of each other. This was only possible with the five subjects that gave as the end product of their importance assessment weight to sub-attributes, the so-called reductionists. The other subjects weighted only safety and comfort, which we consider independent of each other within the context of the assignment. For the five reductionists, we established how many pairs of sub-attributes could in our own view be considered interdependent (admittedly, a subjective judgment) as a percentage of the total number of possible pairs of sub-attributes for safety and comfort, respectively. The number of

possible pairs when *n* attributes are mentioned is *n(n-1)/2*. We consider a value of more than 10% a case of significant interdependence. With this procedure, we found that two of the five reductionists (40%) showed significant interdependence (maximal values were 40% and 50%, respectively). So, the majority seems to take care of interdependence (although not enough for the majority rule), which is contrary to the values on the other indicators. A likely cause seems to be that the number of eventually weighted sub-attributes was so small relative to the number of initially generated attributes that interdependence seems unlikely, even when not trying to avoid it. In four of the five cases (including the three cases where there was no interdependence) it was easily possible for us to weigh at least 25% more of the initially generated attributes without having any trouble to avoid interdependence. So, we conclude that the low level of interdependence of attributes is more of a by-product of lack of concern for comprehensiveness then the result of concern for interdependence.

In sum, our indicators have such low values that we reject expectation 3A and accept its complement *3B for the majority case*. This is in line with some other observations we made. Firstly, all subjects, not just the reductionists, devoted much more effort to weighting (sub-)attributes in isolation ('safety is important') than to weighting (sub-)attributes against each other ('safety is more important than comfort'). For further elaboration, see Chapter 3. This indicates that subjects were not deeply concerned with relationships between (sub-)attributes. Further proof for this assumption is provided in Chapter 4, where we show that there are no clear quantitative or qualitative relationships between sub-attributes and their main attributes. Also, some subjects indicated that sub-attributes could be used as conjunction thresholds ('if a minibus has no air conditioning I will not even consider it'). In that case, interdependence between sub-attributes is irrelevant; below the conjunction threshold the weight of the sub-attribute is infinite. All in all, we see no reason to believe that subjects were concerned about interdependence. We found no ways of framing concerning interdependence to any significant extent. As with comprehensiveness and redundancy, the high level of confidence of the subjects provides no explanation.

We now know how bounded rationality manifests itself during the generation of attributes. This was the first perspective as outlined in Section 2. But the attributes taken into consideration have to be weighted; that is what the importance assessment process is all about. In what way does bounded rationality manifest itself in the choice of weight values to be taken into consideration? In the next section we'll investigate this question.

## 6 Results: the structure of the set of weight values taken into consideration

### 6.1 Weight values used

Of the 18 subjects, 17 (94%) limited the range of weight values with one or more methods. Should we choose to include ordinal weighting for the two main attributes ('safety is more important than comfort'), then the number would be 18 (100%). This is,

inevitably, a clear case of structure in bounded rationality.

The results are summarized in Table 3.

| **Scale** | **Number (%) of subjects employing it** |
|---|---|
| 1: Limited number of ordinal categories | 4 (22%) |
| 2: Limited number of values on a scale from 0 to 1 or from 1 to 10 | 9 (50%) |
| 3: Limited number of ratio values (1:1, 1:2 1:3 etc.) | 4 (22%) |
| **Positioning method** | **Number (%) of subjects employing it** |
| 4: Starting with extreme value, working towards the middle | 3 (17%) |
| 5: Starting from the middle, working towards extreme values | 6 (33%) |
| 6: Anchoring | 1 (5%) |

*Table 3: Ways of reducing the number of considered weight values employed.*

Some of the subjects used more than one scale or positioning method. This will be discussed later in this section. In some cases, especially with positioning methods 4 to 6, subjects were explicit in their way of working. One of the subjects, for example, explicitly stated that he arrived at a weight of 80% for safety because he thought about the 20-80 rule. In most cases, however, subjects were implicit, and the structure in the weight values was inferred from the weights actually mentioned. All subjects used ordinal qualifications, but we did not see those as constituting scales (see Section 6.2).

We conclude that for those subjects that limited the range of considered weight values, in each case there was an identifiable structure in the sets of weights considered (increments in a scale from 0 to 1, ratios, reference points). This means that expectation 4A can be accepted and 4B convincingly rejected. Subjects were not only boundedly rational in taking only a small subset of all possible weight values into consideration (we assumed that from the start), but there also was a system in the weight values they took into consideration, even though in many cases there are no indications that the subjects were aware of this system.

An interesting question is why we found a system in the choice of weight values but not in the choice of choice of attributes. It will hardly be a surprise that, as in the case of attributes, a systematic approach with respect to comprehensiveness of the weight scale or redundancy of the weight values is not the reason. Again, the expectations 5A and 6A have to be rejected and *5B and 6B can be accepted even for the extreme case*. We will come back to this later. We have no definite answer, but several possible explanations for the acceptance of 4A come to mind. The first pertains to the reasons for lack of rationality mentioned in Section 2.2. Choosing a limited number of weight values probably provides a much more manifest reduction of cognitive workload than choosing

certain attributes. After all, in the latter case actors may not realize they lave out certain attributes while this is obvious with choosing a set of weight values. This is in itself no reason for choosing a system, but choosing a system reduces workload even further. After having chosen 'intervals of 0.1' the work is done. Besides, actors may be familiar with certain frames for weight values from popular publications such as the Consumers' Guide in the Netherlands. And thinking in rough percentages is rather common. Anchors also provide frames. A case in point is the 'fifty-fifty' case (see, for example, Bruine de Bruin, Fishhoff, Milstein & Halpers-Felsher, 2000; Fishhoff & Bruine de Bruin, 1999). From the think-aloud [protocols, we definitely got the impression that some 'ready-made' frames, like 0.1 intervals, were common knowledge amongst our subjects. Also, it was clear from the assignment that weight values had to be chosen during the course of the assignment, whereas this was not at all clear for the attributes (the subjects could have stayed with the two main attributes in the assignment: safety and comfort). So, the need to formulate weight values combined with the availability of 'ready-made' frames that reduces cognitive workload may have encouraged the subjects to apply a system in their choice of possible weight values, albeit perhaps unconsciously.

As stated, expectations 5B and 6B are accepted: subjects are by no means concerned about comprehensiveness and redundancy. We have no definite explanation for this, but looking at the think-aloud protocols in a global way and keeping the results of the exit interviews in mind we can propose several possible reasons that could be investigated further:

1: Subjects almost never deliberated about the scale to use. They either picked one without seemingly thinking about it (often starting with unstructured ordinal qualifications) or picked a scale that followed from a weighting method. For example, if they performed pairwise comparison, some sort of ranking order was the obvious scale to use. Sometimes subjects thought about the way attributes should be compared on importance, but that was mostly done on the concrete level of comparing two or more specific attributes, not on a general scale level. One subject consciously chose a scale; one that her employer used. It may be that she was more familiar with weighting problems than the other subjects because she was the only one with working experience;

   All this does not preclude deliberations on how to use the chosen scale. One subject (4) elaborately evaluated which minibus would be chosen using, amongst other things, various conjunction levels. But the 0 to 1 interval weight scale and the weights of the attributes 'safety' and 'comfort' stayed in place. Some subjects interpreted the 'weighting' as mentioned in the assignment as some sort of scale (for example 'giving a mark'. If subjects were concerned about a scale, it was about how to apply it (for example, how to reduce the complexity of pairwise comparison), and not about the range or precision of the scale;

2: None of the subjects wondered how precise the weighting had to be (Chapter 4). This eliminates a potential motive to consider comprehensiveness;

3: Relatively, much effort was spent on identifying attributes and weighing them in isolation (not against each other). This issue is covered in Chapter 3. For this so-called absolute weighting, it seems logical to use simple scales like 1 to 10 scales or ordinal qualifications ('very important', 'not important') without giving the scaling issue much further thought. Why think about scales if qualifications we use daily are available and, at least at first sight, adequate? When comparing attributes the scaling issue may, of course come into view again. But the following points show that there are other reasons not to be concerned about scales;

4: To the extent that subjects used scales, they mostly used rather straightforward scales like 0 to 1 intervals, 0 to 10 intervals and importance ratios (safety is X times as important as comfort). All these scales are well known, from, for example, test reports published by the Dutch Consumers' Association. It is quite possible that subjects accepted these scales as valid, as far as, for example, range and precision are concerned. As stated, one subject used a scale used by her employer, and hence had little reason to deliberate on the validity of the scale;

5: Range and precision, as manifestations of comprehensiveness, do not seem to have generated problems with scales in the way they may do with attributes. Even after a scale has been chosen the range can be extended if desired (for example, the interval can be enlarged) and most often, new weight values can be inserted in between existing values. This is true for both qualitative (ordinal) and qualitative scales. Also, the required range is straightforward in advance: from 'not important at all' to 'this attribute is the only one to take into consideration'. So there seems little need to deliberate. Furthermore, if a scale is defined, all possible values are usually clear even if not made explicit. Every adult of even modest intelligence can generate an indefinite number of values between 0 and 1, for example. With striving for comprehensiveness of attribute it is different. Naming one, or any number of, attributes need give no clue as to which attributes are still missing;

6: Redundancy of a scale is also not a big issue. Values that are redundant can simply be ignored, either incidentally or systematically. 17 of the 18 subjects did the latter, either by systematically working with a limited subset of weight values or by using positioning methods. Our strong impression was that this was so straightforward for the subjects that deliberations about the redundancy of (parts of) scales were superfluous. When subjects were considering weight values, they could do this without any need for considering the redundancy of scales;

7: Given the flexibility of scales as illustrated above, concern about general range and precision issues would likely have had no advantages but would only have increased complexity. Why worry about comprehensiveness or redundancy of scales if ready-made scales are available and can be adapted almost at will?

The above provides no reason why subjects should have felt that comprehensiveness and redundancy were no issues outright. Most subjects had preferences for ordinal scales. 13 of the 18 subjects gave ordinal judgments at first and 7 persisted even after having been

asked by the experimenter to give quantitative weights. With ordinal scales, comprehensiveness and redundancy are issues to be at least looked into. Also, subjects could have wondered whether, for example, 0.1 intervals on scales from 0 to 1 were precise enough. Furthermore, all subjects gave many qualitative weights to attributes during the phase of absolute weighting that could not be seen as values on a scale. Thinking about the range of qualifications available would have been reasonable. So, if hat subjects did not consider comprehensiveness and redundancy it was not because there was no reason at all for it.

All in all, we feel that choosing and working with scales was not experienced as a problem. The problem seems to have been much more for the subjects to get clear to themselves how important they thought attributes were, and then to express it on a scale. The opinion about the importance of attributes that subjects had was much less precise than the scales they used. Note, however, that these reasons offer no explanation for the fact that, within the scales used, only a limited number of values were considered.

## 6.2 Combinations of scales and positioning methods

Now that the expectation 4A has been accepted, we can go one step further. Let us investigate the expectations concerning the possible use of interdependent weight value structures formulated in 7A and 7B. The total number of subjects mentioned in Table 3 indicates that some subjects must have used more than one method. This is indeed the case, as is shown in Table 4. This combination of methods makes possible a further reduction of the number of weight values taken into consideration (see below).

| **Scale** | **Subjects using the scale** |
|---|---|
| 1: Limited number of ordinal categories | 2, 3, 16, 17 |
| 2: Limited number of values on a scale from 0 to 1 or from 1 to 10 | 1, 4, 5, 8, 9, 10, 13, 14, 18 |
| 3: Limited number of ratio values (1:1, 1:2 1:3 etc.) | 1, 7, 12, 18 |
| **Positioning method** | **Subjects using the method** |
| 4: Starting with extreme value, working towards the middle | 4, 7, 18 |
| 5: Starting from the middle, working towards extreme values | 6, 10, 11, 13, 14, 18 |
| 6: Anchoring | 14 |

*Table 4 Use of range-limiting and positioning methods by subjects*

Although not immediately apparent from the table, we will show shortly that expectations 7A and 7B can be neither accepted nor rejected on the basis of our data.
The table shows that the subjects who used a limited number of ordinal categories (scale 1) did not use any other scale or positioning method. This is perhaps not surprising, for

these subjects used scale 1 for the ordering of a large number of sub-attributes. In that case, methods 4 and 5, the only two methods suitable for ordinal weights, are perhaps not so informative, since these methods are especially useful in putting two weight vales against each other and then searching for the right balance at interval or ratio level.
The scales 2 and 3 were only in two cases used together by two subjects. If weights are to be meaningful, they have to be comparable and hence have the same scale. Therefore, it is not logical to use both scales at the same time. The conversion between the scales as mentioned is rather straightforward. The subject 18 using two positioning methods used various scales before settling for a definitive one.

Only one subject (18) used three methods (3, 4 and 5). Subject 18 explicitly converged to the eventual value, going from 1:1 through 2:1, 3:1, 4:1 and 5:1 back to 3:1 and 4 to 1 in favor of safety.
The combination of scale 2 or 3 with positioning method 4 or 5 can give a significant reduction of possible weight values. If 0.1 increments on a 0-1 scale are used and the weighting is started from the middle, there are only five values to be considered. None of the subjects got this far. Still, we can assume that the limitations of weight values achieved by the scales and / or positioning methods played a role of some importance in making the assignment manageable for the subjects by reducing the number of alternative solution to consider and thereby the complexity of the assignment. This fact serves as a practical example of one aspect of the bounded rationality theory: the fact that actors limit the options to choose from in order to make the decision process manageable.

Altogether it is clear that interdependence was not a problem that the subjects faced. All but two subjects use one scale and the two using different scales use scales that can be easily converted, even if the subjects did seldom do so. Only in two cases a subject used more than one positioning method, but these methods can be seen as complementary (mutually independent) anyway. This makes expectations 7A and 7B (about the subjects' concern for interdependence) irrelevant. The subjects may not have been concerned about interdependence, but there was no reason to be concerned, given the scales and positioning methods used. Now one could propose that the scales and positioning methods were chosen with the aim of independence, but nothing at all in our data supports this assumption. All this means that expectations 7A and 7B can neither be accepted nor rejected on the basis of our results. Comprehensiveness and redundancy were issues that the subjects could have considered, at least before choosing scales and positioning methods. So expectations 5A, 5B, 6A and 6B are relevant, although not very much so. But once scales and positioning methods had been chosen, interdependence became, to all practical intents and purposes, irrelevant. So expectations 7A and 7B lost their explanatory power. Of course, with a different assignment or with a different kind of subjects (for example experts) the scales used could make interdependence relevant.

One issue may need clarification. As already said, all subjects gave many qualitative weights to attributes, like 'not so important', quite important' and the like. Should these qualifications be considered values on a scale, then interdependence would be a serious issue because in that case all subjects would have used overlapping scales without any identifiable conversion routine between the qualifications and the scales mentioned in

Tables 3 and 4. But the qualifications did, in our view, not approach the character of scales, so redundancy was not a problem. We have the following reasons for this interpretation:

1: There was no identifiable structure in the qualifications. For example, it often was unclear whether 'rather important' was more or less important than 'quite important' or even whether the order of importance was maintained during the importance assessment;

2: The qualifications were mostly aimed at and used during absolute weighting (weighting attributes in isolation instead of against each other). Then it is, for example, difficult to judge how important safety and comfort are relative to each other if safety is 'rather important' and comfort is 'quite important'. Qualifications like 'safety is more important than comfort' were used during relative weighting, but not with any identifiable consistency and sometimes intermittent with absolute weighting. Add to this that many sub-attributes were weighted and it becomes clear that the qualifications cannot reliably be related to scales. The danger of absolute weighting is, of course, that every attribute is judged 'important' or ' very important', so that the word 'important' loses its meaning. Only during relative weighting using the scales mentioned in Tables 3 and 4 did it become clear that, in terms of Orwell 'all attributes are important, bit some are mote important than others'.

So, there seems to be no reason to see the many unstructured qualifications attribute importance given as indications of scales. Thus, they have no influence on interdependence of scales and positioning methods.

## 7 Bounded rationality in importance assessments: conclusions

In this chapter, we investigated the concept of bounded rationality in assigning weights. To this end, we tested a series of expectations. The results are summarized in Table 5. In the remainder of this section, we discuss the way in which these results relate to each other, and what suggestions for improving weighting processes might be derived from them.

| Expectations | Accepted/rejected |
|---|---|
| Expectation 1A:<br>Given their bounded rationality, actors try to make certain (explicitly or implicitly) that the set of attributes they use is comprehensive. | Rejected |
| Expectation 1B:<br>Given their bounded rationality, actors are not concerned about the comprehensiveness of the set of attributes they use. | Accepted for the majority case |
| Expectation 2A:<br>Given their bounded rationality, actors try to make certain (explicitly or implicitly) that the set of attributes they use is non-redundant. | Rejected |
| Expectation 2B:<br>Given their bounded rationality, actors are not concerned about the redundancy within the set of attributes they use. | Accepted for the majority case |
| Expectation 3A:<br>Given their bounded rationality, actors try to make certain (explicitly or implicitly) that the attributes they use are independent. | Rejected |
| Expectation 3B:<br>Given their bounded rationality, actors are not concerned about the interdependence of the set of attributes they use. | Accepted for the majority case |
| Expectation 4A<br>There is an identifiable structure in the set of weight values that subjects take into consideration | Accepted for the extreme case |
| Expectation 4B:<br>There is no identifiable structure in the set of weight values that subjects take into consideration. | Rejected |
| Expectation 5A:<br>Given their bounded rationality, actors try to make certain (explicitly or implicitly) that the set of weight values they use is comprehensive. | Rejected |
| Expectation 5B:<br>Given their bounded rationality, actors are not concerned about the comprehensiveness of the set of weight values they use. | Accepted for the extreme case |
| Expectation 6A:<br>Given their bounded rationality, actors try to make certain (explicitly or implicitly) that the set of weight values they use is non-redundant. | Rejected |
| Expectation 6B: Given their bounded rationality, actors are not concerned about the redundancy within the set of weight values they use. | Accepted for the extreme case |
| Expectations 7A and 7B about concern for interdependence | Neither accepted nor rejected |

*Table 5: The acceptance or rejection of expectations*

We conclude that expectations 1A, 2A and 3A have to be rejected. Comprehensiveness, non-redundancy and independence clearly were not issues of any importance for the subjects. The fact that all expectations were rejected because no single type of statement was observed with any consistency with 12 or more subjects makes the fact that some indicators might be related to more than one variable academic. Subjects were boundedly rational (they took only a limited number of attributes into consideration), but there was no (explicit or implicit, consciously or unconsciously created) system in the way they were boundedly rational.

All in all, there is no reason to believe that subjects were concerned with the need to reduce the complexity of the assignment by reducing the number of attributes under consideration in a conscious, systematic way. They may have been aware of the complexity of the problem facing them, but that did not result in concerted action. One could say that the subjects were implicitly boundedly rational by simply not worrying about rationality: not devoting attention to comprehensiveness, redundancy and interdependence, and eliminating (sub-)attributes from consideration without bothering to explain. The lack of system in the identifying of attributes to be taken into consideration, however, points at a poor structuring of the importance assessment process. This issue, and its consequences for the importance assessment process as a whole, is discussed in Chapter 4. Implicit (bounded) rationality is of limited use in organizational decisions. The quality of implicit rationality cannot be assessed or controlled. Furthermore, implicit rationality as it occurred in our study concerned those elements of the importance assessment process that the subjects could have made explicit had they wanted to, for example by making a mental model. So the question remains why the subjects didn't do so. That the eventual assignment of weights was largely implicit or intuitive (see Chapter 4) is understandable; weighting is partly a personal affair. But it still seems logical that, within the (simulated) organizational context in our experiment, subjects would make explicit those elements that can be made explicit.

The arguments against an identifiable pattern of bounded rationality mentioned in Section 2.2 do not provide an explanation of the behavior of the subjects at this point. They merely provide possible clues for explanations that could be pursued in future research.

Is there a system in the pattern of acceptance and rejection of expectations? The rejection of expectations 1A, 2A, 3A, 5A and 6A indicates that the subjects were not boundedly rational in a systematic way. We saw, however, that they indeed eliminated many attributes along the way and used a restricted set of weight values. This had the effect of bounded rationality, even if it was not intended.

The acceptance of expectation 4A can be interpreted as an indication that subjects reduced mental workload, not by using a random number of weight values to be considered, but by using a system. This makes the weighting process easier, while slight variances in attribute weights will, in reality or in the perception of the decision maker, often not alter the decision eventually taken. After all, even if the total attractiveness of alternatives may change slightly, the rank order of attractiveness will often remain the same. Then it makes no sense to take all weights within those variations into

consideration. This idea is under certain circumstances borne out in a study of decision making with incomplete information (Kirkwood *et al*, 1993) and is in line with Simon's (1976) concept of satisficing. The use of scales was so straightforward that deliberations about their comprehensiveness and redundancy would merely have increased complexity. Hence, it is understandable that expectations 5B and 6B are accepted. Expectations 7A and 7B could neither be accepted nor rejected. This may be due to the character of the assignment (which may induce the use of certain scales) or to that of the subjects (laymen).

All in all, it is clear that 'bounded rationality' does not necessarily mean 'working in a haphazard way'. Bounded rationality can be unsystematic (expectation 1A, 2A and 3A rejected, but it need not be (expectation 4A accepted). This is an interesting area for further research.

It should be stressed, however, that the subjects were not asked to explain why they displayed the behavior that led to the rejection or acceptance of the expectations. So, while the way most subjects worked had the effect of reducing mental workload, we cannot prove that this was also the aim of those subjects. On the contrary, except for general remarks like: "giving numbers to weight values is very difficult" and statements limiting the number of sub-attributes weighted, there are no clues in the think-aloud protocols that subjects explicitly or consciously sought ways to reduce mental workload concerning the issues in the expectations, although there were a few references to other methods of reducing workload. Still, the effect of reduced workload is there, which is in concurrence with the concept of bounded rationality. The results concerning expectations 1A, 2A and 3A and, to a lesser extent, 5A and 6A are in accordance with – and a partial explanation for – the inadequacy of the structuring of weight assessment problems we discussed in Chapter 4. When actors are boundedly rational but do not systematically explore those attributes that they do take into consideration, it should come as no surprise that the attributes eventually weighted form a more or less unstructured collection.

The fact that subjects weighted far fewer attributes than they initially considered may be a form of bounded rationality, but it may also simply point to an unstructured way of working whereby attributes were, for example, simply forgotten. This is in line with the lack of expertise of our subjects in the field of the assignment, a topic covered in Chapter 3. On the other hand, it draws attention to an important limitation of our research design. By primarily looking for bounded rationality on the basis of verbalized thoughts of subjects, we are in danger of only recognizing those forms of bounded rationality that are in any way expressed in the protocols. So, either the subject should consciously address bounded rationality, or the researcher should recognize implicit signs of it. According to Ericsson & Simon (1993), the general structure of cognitive processes is maintained if the think-aloud method is used properly, which we took care to do (see Chapter 2). But we cannot be sure that there was no implicit or unconscious rationality that went undetected. On the other hand, nearly all subjects were so elaborate in their deliberations that we are reasonably sure that we did not miss any significant form of explicit bounded rationality. It seems likely that those forms of implicit bounded rationality that we missed are so transparent to the subjects themselves, and possibly so mixed with, for example, simply

forgetting to weigh attributes, that they will not have a major and systematic influence on the general structure of the important assessment process. For example: if subjects would be asked to make the same importance assessment twice, they may weigh slightly different sets of attributes but those attributes that are explicitly evaluated (and not implicitly left out) are likely to be the same in both cases. Of course, we do not have proof for this, so an interesting area of research would be: exploring the nature of implicit bounded rationality and its influence on the importance assessment process. Obvious ways to set about this research are: asking subjects, after completion of the importance assessment task, to explain why certain attributes were not weighted, or presenting them with pairs of the weighted attributes and asking them to identify interdependences. Also, subjects could be asked to choose from a list of attributes the set of attributes they want to weigh. Interdependences within the chosen sets could be studied, as could the number of chosen attributes as a function of, for example, the number of attributes initially presented, the abstraction level of the attributes, the complexity of the relationships between attributes and so on. Hopefully, explanations for the subjects' behavior can be found in the theoretical perspectives outlined in Section 2.2.

Obviously, the area of implicit bounded rationality in importance assessment processes is a fruitful area for research. But even the limited work we have done yields benefits for management and decision making in practice. As a conclusion of this chapter, we give a few ideas for improvements of the importance assessment process by individual actors within an organizational context where the weight values will have to be justified to other actors at some point. It should be realized, however, that our results apply only to laymen performing non-routine importance assessments, separated from evaluation of alternatives, in situations where attributes are actually weighted. So, we cannot be sure exactly which role the behavior the suggestions are meant to stimulate or discourage plays in real-life situations where experts will often be the ones making importance assessments, intertwined with evaluation of alternatives. Two approaches are possible to counter this problem. First: establish whether the behavior we observed shows in real-life situations and adjust our suggestions accordingly. Second: see how the solutions work out in practice and improve them incrementally.

The suggestions discussed here are based on the assumption that any method that helps individual actors in weighting attributes should follow as closely as possible the natural thinking process. In this way, the actor is left as free as possible to work in his or her own way while at the same time compensating for limitations in human thinking. After all, assigning weights is a highly personal process, comparing the incomparable. So the actor is probably best left free as much as possible.

The suggestions are the following:

1: Make actors aware of the option to reduce the complexity of the task facing them by reducing the number of attributes to be taken into consideration. The same advice may be given with regard to the number of weight values. Subjects tend to reduce the number of attributes and weight values, but encouraging them to do it more consciously may be of help to them. It could at least increase their efficiency.

How can this suggestion be applied in practice? This is addressed in suggestions 2 and 3;

2: Encourage the actor to make a cognitive map or some other mental model (see section 2.2). This has several advantages:

- it may stimulate the examination of the attributes for independence, redundancy and completeness;
- it stimulates a systematic examination of the relationships between (sub)-attributes, therefore giving the possibility to assign weights based on the relative contribution of a sub-attribute to the utility of the corresponding main attribute, in the process facilitating later justification;
- it may encourage systematic decomposition (see the next point).

3: Encourage systematic decomposition of attributes. This may contribute to awareness of the issues of independence, redundancy and comprehensiveness.

With these instruments, the option exists to let actors work in their own way first, and only when they have completed certain activities, give them the option of reflecting on their work in a systematic way. In this manner, the proposed instruments do not unduly limit the creativity of actors or negate the strengths of their personal way of working.

It is, of course, not clear from our experiment whether the quality of the weighting process will be improved in this way. But it seems likely that at least the discussions about the weight values with other actors, almost inevitable during organizational decision processes, will be more productive because more elements of the individual weighting process can be made explicit. Will this increase the quality of weighting and, eventually, decision making? We believe that having the option to make the weighting process more structured increases the number of alternative ways of working available to actors, which is, in our view, a good thing. Evaluating the effects of our suggestions will be an interesting area for future research.

## Literature


Akin, Ó. (1994). Creativity in design. *Performance Improvement Quarterly, volume 7,* 9-23.

Arthur, W.B. (1994). Inductive reasoning and bounded rationality, *Complexity in Economic Theory,* Volume 84, 406-411

Aumann, R. (1997). Rationality and bounded rationality. *Games and Economic Behavior, volume 21,* 2-14.

Baarda, D.B. & de Goede, M.P.M. (2001). *Basisboek methoden en technieken*. Groningen: Stenfert Kroese.

Beach, L.R. (1993), Broadening the definition of decision making: The role of pre choice screening of options, *American Psychological science,* volume 4, 215-220.

Beach, L.R. Puto, C.P, Heckler, S.E, Naylor, G & Marble, T.A (1996). Differential versus unit weighting of violations, framing, and the role of probability in Image Theory's compatibility test. *Organizational behavior and Human Decision Processes, volume 65,* 77-82.

Bettman, J.R, Luce, M.F. & Payne, J.M. (1998). Constructive consumer choice processes. *Journal of consumer research*. Volume 25, 187-217.

Bruine de Bruin, W, Fischhoff, B, Millstein, S.G. & Halpern-Felsher, B. (2000). Verbal and numerical expressions of probability: "It's a fifty-fifty chance". *Organizational Behavior and Human Decision Processes, volume 81.* 115-131.

Chapman, G.B. & Johnson, Eric. J. (1999). Anchoring, activation and the construction of values. *Organizational behavior and human decision processes, volume 79,* 115-153.

Damasio, A.R. (1995) *De vergissing van Descartes – gevoel, verstand en het menselijk brein* (translated from English into Dutch by Liesbeth Teixeira de Mattos. Amsterdam: Wereldbibliotheek.

De Boer, L. (1998). *Operations research in support of purchasing; design of a toolbox for supplier selection.* Enschede: Universiteitsdrukkerij.

Ericsson, K.A. & Simon, H.A. (1993). *Protocol analysis: verbal reports as data* (2nd ed.). Cambridge, MA [etc.]: Bradford Books/MIT Press.

Fischhhoff, B, & Bruine de Bruin, W. (1999). Fifty-fifty = 50%? *Journal of behavioral Decision Making, volume 12,* 149-163

Hammond, J, Keeney, R & Raiffa, H (1998). Gelijk oversteken: ruilen in natura als basis voor complexe beslissingen. *Nijerode Management Review, issue 12,* 63-75.

Highhouse, S., Paese, P.W. & Leatherberry, T. (1996) Contrast effects on strategic-issue framing, *Organizational behavior and human decision processes,* Volume 65, 95-105.

Kahneman, D. (1994) New challenges to rationality assumption, *Journal of institutional and theoretical economics,* Volume 150, 18-36.

Kahnemann, D. & Tversky, A. (2000) (ed.). *Choices, values and frames*. Cambridge: Cambridge university Press.

Kaufman, B.E. (1999). Emotional arousal as a source of bounded rationality. *Journal of Economic Behavior and Organization, volume 38,* 135-144.

Keeney, R.L (1992). *Value-focused thinking.* Cambridge MA [etc.]: Harvard University Press.

Keeney, R.L & Raiffa, H. (1976). *Decisions with multiple objectives, preferences and value tradeoffs.* New York [etc.]: John Wiley & Sons.

Kirkwood, C.W. & Corner, J.L. (1993). The effectiveness of partial information about attribute weights for ranking alternatives in multiattribute decision making. *Organizational Behavior and Human Decision Processes, volume 54.* 456-476.

Levine, G.M., Halberstadt, J.B., Goldtsone, R. (1996) Reasoning and the weighting of attributes in attribute judgments, *Journal of personality and social Psychology,* Volume 70, 230-240.

Loewenstein, G. (1996) Out of control: Visceral influences on behavior, *Organizational behavior and human decision processes,* Volume 65, 272-292.

Miller, G.A. (1956). The magical number seven, plus or minus two. Some limits on capacity of processing information. *Psychological review*, volume 63, 81-87.
Munier, B et al. (1999) Bounded rationality modeling, *Marketing letters,* Volume 10, 233-248.
Polzer, J.T, Stewart, C.J. and Simmons, J.L. (1999). A social categorization explanation for framing effects in nested social dilemma's. *Organizational Behavior and Human Decision Processes, volume 79*, 154-178.
Rubinstein, A. (1998). *Modeling bounded rationality*. Cambridge (MA), London: The MIT Press
Saad, G. & Russo, J.E. (1996). Stopping criteria in sequential choice. *Organizational Behavior and Human Decision Processes, volume 67,* 258-270.
Schwartz, N. (2000. Emotion, cognition and decision making. *Cognition and Emotion, volume 14*, 433-440.
Schón, D. (1985). *The Design Studio*. London: RIBA Publications Ltd.
Shafir, E. & Leboeuf, R.A. (2002). Rationality. *Annual Revenue Psychology, volume 53*, 491-517.
Shafir, E., Simonson, I & Tversky, A (1993), Reason-bases choice, Cognition, Volume 43, 11-36.
Simon, H.A. (1976). *Administrative behavior.* New York: The Free Press, London: Collier.McMillan Publishing.
Simonsos, I (1990). The effect of purchase quantity and timing on variety-seeking behavior. *Journal of marketing research*, volume 27, 150-162.
Strauss, A. & Corbin, J. (1998). *Basics of qualitative research; Grounded Theory, procedures and techniques.* Thousand Oaks CA [etc.]. Sage Publications.
Toda, M (1980). Emotions and decision making. *Acta Psychologica, volume 45*, 133-155.
Van der Pligt, J. & Koele, P. (1993). Beslissen en beoordelen vanuit psychologisch perspectief. In: Koele, P & van der Pligt, J. *Beslissen en beoordelen*. Meppel: Boom.
Vincke, P. (1992). *Multiple Decision Aid.* Chichester [etc]: John Wiley & Sons.
Warren, K. (1995). Exploring competitive futures using cognitive mapping. *Long range planning*, volume 28, 10-21.
Wong, K.F.E. & Kwong, J.Y.Y. (2000). Is 7300m equal to 7.3 km? Same semantics but different anchoring effects. *Organizational Behavior and Human Decision Processes, volume 82*, 314-333.

# Chapter 6: Wrapping up

The aim of this thesis was to build a model of the importance assessment process as performed by individual actors within an organizational context. This aim has been achieved in the sense that a model was built that proved capable of describing the behaviour of our subjects qualitatively and to e certain extent quantitatively. The model was used in testing expectations concerning importance assessment processes. The testing of these expectations gave insights in various aspects of the importance assessment process, such as the interrelations between phases, bounded rationality and possible differences between laymen and experts. We believe that the internal and content validity of our research are adequate. Due to the lack of literature on importance assessment processes available, construct validity is less than desired. External validity is rather low, as a consequence of our research method, namely: think-aloud sessions with students in a laboratory setting. But the relatively high internal validity this approach yielded makes this price worth paying. In research on a phenomenon about which little is known, internal validity is at least as important as external validity.

We started the research with the following problem statement:

> *What is the structure of the thinking process by which layman actors involved in non-routine decision-making processes assess the importance of attributes of the various alternatives under consideration?*
>
> *In particular:*
>
> *1A: In terms of what mental activities can the importance assessment process be described?*
>
> *1B: What characterizes laymen's performance of importance assessment processes?*
>
> *2: What quantitative and/or qualitative relationships exist between these activities?*
>
> *3: To what extent, and how, do actors limit the cognitive workload performing the activities of an importance assessment?*

Before we assess to what extent the problem statement have been answered, we should remember the limitations in our research. As explained in Chapters 1 and 2, we used a small sample, in a laboratory context. As subjects we used not experienced decision makers, but students. Looking back, we think that out choices are justified by the depth and internal validity of our research. Using small samples is not uncommon in think-aloud experiments. We were able to analyse the importance assessment process, uncontaminated by subjects' previous experience and importance judgments, in an environment free from, for example, group dynamic pressures. But this has a price.

Statistical validity is low. And our results cannot be directly related to importance assessment processes in 'real world' contexts. Hence, propositions we made about, for example, the way experts go about importance assessment processes in practical decisions, are just that: propositions, not facts based on research. Our propositions have yet to be tested in future research.
The same sample was used for all studies within this research. So, the results of Chapters 3, 4 and 5 are based on the same sample of 18 students. This means that we could not test the influence of variables by comparing experimental and control groups, and that we did not test and refine our findings with new samples. The in-depth analysis of the results obtained with our sample was so time-consuming that follow-up studies were not possible. Our in-depth analysis enabled us to study our subjects from various angles, as reflected in Chapters 3 to 5. In other words, if no research with control groups could be done, the fact that the results of Chapters 3 to 5 are based on the same sample makes them more easily comparable. In future research, selected variables could be studied in a more quantitative way, with larger samples. It will then often be necessary, however, to operationalize those variables in a way suitable for quantitative research. Most of the indicators we used required detailed text analysis, not suitable for large samples.
We only looked at situations in which weighting actually takes place (some choice strategies require no weighting) and where weighting is separated from the evaluation of alternatives. This may well be the case with important organizational decisions about, for example, the acquisition of capital goods. But often, the conditions we assumed in our research will not prevail, this limiting the external validity of our work.

Within these limitations, the main findings of our research can be summarized as follows:

**1A In terms of what mental activities can the importance assessment process be described?**

The way in which individual laymen go through importance assessment processes concerning non-routine problems within an organizational context can be described in a model consisting of seven phases and a number of auxiliary activities (Chapter 3). This model is called the Weight Assessment Model (WAM). The phases can be grouped in a structuring, a weighting and an evaluation cluster (Chapter 4). The most significant phases, in terms of effort devoted to them, are the attribute processing phase, in which attributes are identified and formulated, and the absolute weighting phase, where attributes are weighted in isolation (so, not against each other). These findings pertain to sub-problem 1A of the problem statement.

**1B What characterizes laymen's performance of importance assessment processes?**

Based on the literature on expertise, we identified the characteristics of laymen and the way these laymen perform importance assessment processes (Chapter 3). We did not, however, compare laymen with experts and hence cannot establish that the way our subjects performed importance assessment processes are unique to laymen. We could, based on the literature, identify several reasons why experts may perform importance

assessments differently from laymen. We propose that experts are likely to use more formal definitions of attributes, to decompose attributes more systematically and to devote more attention to integrating sub-attributes, possibly by defining common denominators such as money. These propositions were not tested. Being laymen does not prevent actors from having high confidence in their importance judgment. We cannot tell whether they differ from experts in this respect, but one would expect laymen to have relatively low confidence. In sum, we laid a basis for addressing differences between laymen and experts in future research by identifying variables that may describe differences between laymen and experts, but did not research the differences ourselves.

**2: What quantitative and/or qualitative relationships exist between these activities?**

Actors perform importance assessment processes in a largely unsystematic fashion, but certain working rules can be observed with significant numbers of actors (Chapters 3, 4 and 5). The quantitative and qualitative relationships between the various phases remain largely unclear, although possible functions of phases taken in isolation could be identified (Chapter 4). These findings pertain to sub-problem 2 of the problem statement. This behaviour of actors is in accordance with an incremental approach.

**3: To what extent, and how, do actors limit the cognitive workload performing the activities of an importance assessment?**

Actors reduce he complexity of the importance assessment process by taking only a limited number of possible weight values per attribute into consideration, in an identifiable pattern. However, while they obviously do not take all possible attributes into consideration, we found no pattern in the attributes that were considered (Chapter 5). These findings pertain to sub-problem 3 of the problem statement.

Altogether this research has practical implications for support of importance assessment processes. For example, there are several possibilities for developing instruments aimed at helping actors to improve the quality of their importance assessments (Chapters 3, 4 and 5). Examples are: encouraging actors to make mental models of the relationships between attributes and decomposing attributes systematically.

Furthermore, the research methodology as developed has good potential for further applications. The research method we developed is suitable for addressing problems without having at the start of the research a theoretical framework or a clear image of variables that may be relevant, and yet obtaining results sufficiently accurate for quantitative analysis (Chapter 2).

Our model, on the one hand, showed that individual importance assessment processes can be described with a limited number of concepts like the phases in the model (thus identifying common features in these processes) while on the other hand providing sufficient detail to identify differences between importance assessments of individual actors (for example, with regard to the ways actors reduce the number of weight values

taken into consideration).

But we merely scratched the surface of the importance assessment process. *The research generated many new questions*. Some of these questions pertain to the explanation of the results obtained. Which, if any, of our possible explanations for the high confidence that out subjects had in their work is correct? Why is so much more effort devoted to absolute weighting than to relative weighting? Why do subjects never try to find a common denominator and do they not utilize the problem-solving tools they possess to the full extent? Why do they, during the weighting phases of the importance assessment process, make so little use of work done in previous structuring and weighting phases? Other questions pertain to the applicability of the results, obtained in a laboratory context, to the world outside the confines of the university. Hoe do actors in real decision situations behave? Can importance assessment processes be improved on the bases of research utilizing the Weight Assessment Model? Furthermore, we studied a specific group of subjects, as discussed in Chapters 1 and 2. How are importance assessments conducted in practice, by experts, who at the same time may evaluate alternatives?

One fool can ask more questions than ten wise men can answer. Where should the priorities lay in future research on importance assessment processes? We propose the following options.

A: Studying variables addressed in this research with a group of subjects sufficiently large to ensure statistical validity. Now that we have identified variables relevant to importance assessment processes at least under the circumstances prevailing in our research, we can now study some of them in a far less labour-intensive way with many more subjects. Fore example, the coding scheme we developed makes a much more efficient quantitative analysis of think-aloud protocols possible than the analysis based on the Grounded Theory that yielded the majority of the results in our research.

B: Repeating our study with other populations. Experts are an obvious choice (see Chapter 3). The question is not only in which respects experts differ from laymen, but also whether the domain-specific nature of expertise is confined to the area of expertise or extends to importance assessment processes. An interesting question is whether experts use other ways of reducing the complexity of importance assessment processes than laymen (see Chapter 5). Experts may have less need for complexity reduction, but it is equally possible that they are better at it. Do experts perhaps work more systematically than laymen (Chapter 4)? Other groups are also of interest, like actors who are familiar with formalized decision processes in which importance assessments play a role. How does this effect the way they go about such processes?

   Subjects who have to make decisions with a strong ethical or value component may behave differently from our subjects who were involved in business decisions in which the value component was relatively small. Research comparing various groups could provide much more in-depth knowledge about the nature of importance assessment processes than this study was able to provide.

C: Investigating the usefulness of the WAM in decision situations where importance assessment processes are not isolated from assessing the scores of alternatives on the attributes. In our study, importance assessment was deliberately isolated from the judging of the attractiveness of alternatives, for reasons explained in the introduction and Chapter 2. In real life this will often not be the case. A possible role of the WAM in this respect is to provide a structure for the study of consensus-building in the various phases of a decision process. Does the way in which importance assessments are made (perhaps varying between actors and phases of the process) influence the nature of the consensus-building process? And if so, in what way? A (highly speculative) notion would be that the role of relative weighting increases in importance during the process, relative to that of absolute weighting.

D: Developing instruments capable of measuring more exactly the variables discussed in this study, under real-life conditions. For example, the degree in which the subjects worked in a systematic fashion was measured in a very course manner, with a number of dichotomous indicators, the relative importance of each of which was not established (see Chapters 4 and 5). Thus, while subjects could be labelled as working systematically or not, it was impossible to measure how systematically a particular subject worked. New measuring instruments are required in order to study importance assessment processes more in-depth.

E: Developing instruments for improving the quality of importance assessment processes. Suggestions for this are done in several of the chapters . Possibilities are: encouraging actors to make mental models, to look for common denominators and to consciously utilize a stopping criterion when generating attributes or decomposing an attribute into sub-attributes. Possibly, increasing the effort devoted to relative weighting as a means to check the results of absolute weighting (to which considerable effort is devoted, as noted in Chapters 4 and 5), may also improve the quality of importance assessment. However, we do not know whether the behavior we observed also manifests itself in a real-life situation. Applying instruments without first studying real-life situations has the risk that the behavior the instruments are designed to encourage or discourage is not relevant in practice.

F: In relation to points 3 and 4, it should be noted that a method for measuring the quality of importance assessment processes has yet to be found. The resulting importance judgment is not in itself a measure of quality. The satisfaction of actors with the importance assessment processes they performed seems a logical indicator of quality, but its drawbacks are obvious. Expert judgments are a possibility, but the validity of such a method would need to be established. A possible pitfall is that experts may not be able to appraise the value component in importance assessments, which is by nature difficult to judge in terms of ‘good’ or ‘ bad’. Also, in non-routine decisions, the value of expert judgment will often be limited. Still, if research on importance assessment processes is to fulfil its potential regarding practical value, indicators for the quality of these processes have to be developed.

G: Studying the role of importance assessment in the integration of multidisciplinary information. This touches the core of research on management, as explained in the introduction.

We also would like to draw attention to a conclusion that pertains not just to research on decision processes but to research on management in general. Our research shows that behaviour directly relevant for research on management can be studied in a laboratory context, at the level of individual actors. This is an area that has, in our view, long been neglected. Management scientists often use research methods like case studies, surveys, simulations, business games and the like, generally with the aim of collecting great numbers of variables or indicators in complex organizational contexts, where the input of individual actors are seen as the result of processes which are regarded as 'black boxes' or as determined by global factors such as the social context. However valuable this kind of research is, the input of individual actors can be analysed, as we have shown, and instruments aimed at improving individual cognitive processes and therefore the inputs that shape organizational processes can be developed. This area is, of course, partly covered by psychology, but as this thesis shows, psychological issues that are specific for research on management offer opportunities for management scientists. Interestingly some psychologists approach the problem of connecting individual inputs with organizational processes from the opposite direction. Instead of extending their research from real-life to a laboratory environment they want to use the knowledge gained in observing behaviour in controlled conditions (mostly concerning simple tasks), to study the fulfilment of complex tasks in organizational contexts. This is a tendency in research on expertise (Van der Heijden, 1998). We believe we have shown in this thesis that a meeting in the middle is possible and productive.

It is clear that researchers on importance assessment processes need not be afraid of running out of topics. This thesis is a mere starting point, which it was always meant to be. We feel the concept of importance assessment has a lot to contribute to research on management and can help to further develop its identity *vis a vis* other areas of science.

# Summary / samenvatting

## Hoofdstuk 1: Introductie

### De probleemstelling

Een beslissing bestaat uit het maken van een keuze tussen twee of meer alternatieven waarbij idealiter het aantrekkelijkste alternatief wordt gekozen. Volgens de utiliteitstheorie dient de aantrekkelijkheid van de alternatieven te worden beoordeeld op een aantal kenmerken ofwel attributen. Het belang van elk attribuut wordt uitgedrukt in een gewicht. De probleemstelling van dit proefschrift luidt als volgt:

*1: Wat is de structuur van het denkproces door middel waarvan actoren die zijn betrokken bij niet-routinematige beslissingen binnen een organisatorische context het belang bepalen van attributen van de verschillende alternatieven?*

*1A: In termen van welke mentale activiteiten kan het afwegen van het belang van attributen (hierna te noemen afwegingsprocessen) worden beschreven?*

*1B: Hoe kan de wijze waarop leken (actoren met geringe expertise met betrekking tot een taak) afwegingsprocessen uitvoeren worden gekarakteriseerd?*

*2: Welke kwalitatieve en/of kwantitatieve relaties zijn aanwezig tussen de verschillende mentale activiteiten bij afwegingsprocessen?*

*3: In welke mate, en op welke wijzen, beperken actoren de cognitieve werklast bij afwegingsprocessen?*

De probleemstelling wordt op de volgende punten ingeperkt:

1: Het onderzoek is descriptief, niet normatief. Wij bestuderen hoe afwegers te werk gaan, niet hoe ze te werk zouden moeten gaan. Wel zijn de te toetsen verwachtingen soms gebaseerd op normatieve theorieën;

2: Het onderzoek betreft individuele actoren (afwegers);

3: Het onderzoek betreft niet-routinematige beslissingen.

4: Het onderzoek betreft situaties waarbij sprake is van het bepalen van gewichten. Er zijn beslissingsstrategieën waarbij dit niet aan de orde is, maar die worden buiten beschouwing gelaten.

5: Het onderzoek betreft situaties waarbij het vaststellen van gewichten gebeurt voordat alternatieven worden geëvalueerd. In veel praktijksituaties zullen de twee samengaan.

Het afwegingsproces moet worden onderscheiden van de uiteindelijke afweging (het belangrijkheidsoordeel ofwel de aan attributen toegekende gewichten). Het afwegingsproces betreft de denkprocessen die leiden tot de gewichten. Zulke processen zijn uiterst relevant voor de bedrijfskunde en voor inzicht in de dagelijkse praktijk van management. Een manager moet bij beslissingen verschillende soorten informatie (technische, financiële) integreren tot een keuze uit alternatieven. Ondanks de multidisciplinaire pretenties van de bedrijfskunde is niet bekend hoe dit gebeurt. De input en output van multidisciplinaire beslissingsprocessen zijn uitgebreid bestudeerd, wat zich afspeelt in het hoofd van de manager bij het integreren van informatie is voor bedrijfskundigen, maar ook voor psychologen, nog grotendeels onbekend. Er is veel bekend over het genereren van alternatieven, het scoren ervan, en over instrumenten om gewichten te achterhalen, maar niet over denkprocessen met betrekking tot het vaststellen van gewichten. Hierover handelt dit proefschrift.

Er is slechts een klein aantal proefpersonen onderzocht. Dit gaf de gelegenheid om een groot aantal aspecten van afwegingsprocessen diepgaand te bestuderen (nodig omdat vooraf niet duidelijk was welke aspecten aan de orde zouden komen) maar verlaagt de statistische validiteit. Verder betrof het onderzoek studenten in een laboratoriumsituatie, hetgeen de externe validiteit beperkt. De interne validiteit is echter relatief hoog, en dat werd belangrijker geacht dan een hoge externe validiteit. Immers, door het bestuderen van afwegingsprocessen onder gecontroleerde omstandigheden kon een model worden geconstrueerd dat analyse van (veelal complexe) praktijksituaties mogelijk maakt, ook al moet het model op grond van dergelijke analyses wellicht worden aangepast.

**Hoofdstuk 2: Methode van onderzoek**

De methode van onderzoek wordt in dit proefschrift zodanig beschreven dat hij kan worden toegepast voor onderzoek dat op de volgende punten overeenkomt met het onze:

- het betreft onderzoek naar cognitieve processen;
- er is zo weinig specifieke kennis beschikbaar dat niet vooraf hypothesen kunnen worden opgesteld omtrent de structuur van de te onderzoeken processen;
- het doel is te komen tot niet alleen een kwalitatieve beschrijving van cognitieve processen, maar ook tot een kwantitatieve analyse.

De onderzoeksmethode kent zeven stappen:

***1: Het opstellen van de probleemstelling (zie hoofdstuk 1) en bepalen van de waarnemingsmethode***

Wij kozen voor de hardopdenkmethode waarbij proefpersonen een opdracht hardop denkend moeten uitvoeren. De belangrijkste voordelen van deze methode zijn dat gedetailleerd inzicht wordt verkregen in denkprocessen, dat geen reconstructie achteraf nodig is en dat de proefpersonen hun gedrag niet behoeven uit te leggen.

***2: Het ontwerpen van het experiment***

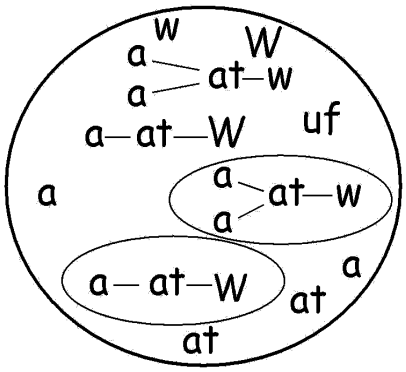
Het experiment werd als volgt vormgegeven:
Er is gekozen voor een laboratoriumexperiment teneinde onbeheersbare invloeden van omgevingsfactoren te vermijden. Achttien proefpersonen kregen de opdracht om voor een fictief transportbedrijf het belang van de veiligheid van aan te schaffen minibusjes af te wegen tegen het belang van comfort. De proefpersonen waren volkomen vrij in de wijze van afwegen hadden gemaakt.

***3: Het ontwerpen van het dataverzamelingsinstrument***

Het dataverzamelingsinstrument werd als volgt ingezet:
Alle uitspraken van de proefpersonen werden op band opgenomen. Daarnaast legden de proefleiders hun observaties vast.
Na uitvoering van de opdracht werd een kort interview gehouden waarin enkele algemene vragen werden gesteld, zoals naar de keuzes die de proefpersonen tijdens het afwegingsproces.

***4: Het ontwikkelen van een voorlopig, globaal analyseschema***

Er was een voorlopig analyseschema nodig dat voldoende richting gaf om de protocollen op een enigszins gestructureerde wijze te bestuderen (nodig voor conceptualisatie van de inhoud van de protocollen). Maar anderzijds moest het schema niet vooraf de analyse onnodig beperken in de zin dat belangrijke maar niet meteen in het oog springende zaken over het hoofd zouden worden gezien. Dit bleek de grootste uitdaging tijdens het onderzoek. Het voorlopige analyseschema werd opgebouwd op basis van het onderstaande statische model van afwegingsprocessen waarvan de elementen afkomstig zijn uit de utiliteitstheorie.




uf = vorm van de nutsfunctie, at = attribuut, w = gewicht, W = gewichtsinterval,
a = argument

*Figuur 1: Globaal model van het afwegingsproces*

Aan de hand van dit model werden enkele algemene typen activiteiten geïdentificeerd. Deze dienden als basis voor een eerste analyse van de protocollen. De activiteiten waren:

1: Structurering van het afwegingsprobleem:

- het kiezen van de vorm van de nutsfunctie;
- bewerking van attributen.

2: Oplossen van het afwegingsprobleem:

- absolute versus relatieve weging van attributen;
- holistische versus dimensionele weging (weging van ‘veiligheid’ en ‘comfort’ versus onderdelen oftewel subattributen daarvan);
- het verbinden van attributen met gewichten en ondersteunende argumenten.

***5: Het uitvoeren van de kwalitatieve analyse***

De kwalitatieve analyse werd uitgevoerd op basis van het ontwikkelde globale analyseschema, met behulp van een procedure ontleend aan de ‘Grounded Theory’. De procedure bestaat uit de volgende stappen:

1: Het identificeren van verschijnselen. Soms wezen de hardopdenkprotocollen op mentale activiteiten. Die werden, indien mogelijk, getypeerd als werkregels volgens welke de proefpersonen het afwegingsproces doorliepen.

2: Het groeperen van verschijnselen in categorieën. Bijvoorbeeld: afzonderlijke werkregels die leidden tot verandering van gewichten werden samengevoegd in de categorie ‘gewichtsveranderende werkregels’. Hierdoor kon de grote hoeveelheid data van de protocollen worden teruggebracht.

3: Het benoemen van de categorieën. Dit betekende dat de voorlopige namen van de categorieën (op basis van de gemeenschappelijkheid van de verschijnselen binnen elk van de categorieën) moesten worden vervangen door namen die de betekenissen van de categorieën ten opzichte van elkaar aanduidden. In feite ging het om meer dan naamgeving: het ging om het bouwen van de globale structuur van het model van het afwegingsprocessen. De modellering van een proces impliceert een fase-indeling van categorieën.

4: Het ontwikkelen van de categorieën in termen van eigenschappen en dimensies. Na de inductieve stappen 1 t/m 3 was dit de eerste van twee deductieve stappen. De omschrijvingen van de categorieën werden verder verfijnd en de onderlinge grenzen werden beter aangegeven. De precisie was beperkt; het was een voorbereiding op het opstellen van het codeerschema voor de kwantitatieve analyse (zie hierna).
5: Het vaststellen van relaties tussen de verschillende categorieën en variabelen. De volgtijdelijke relatie tussen fasen is al genoemd. We keken bijvoorbeeld naar de het gezamenlijk aan- of afwezig zijn van bepaalde verschijnselen bij proefpersonen.

***6: Het uitvoeren van de kwantitatieve analyse.***

Sommige variabelen konden in de loop van de kwalitatieve analyse zo nauwkeurig worden gedefinieerd dat kwantitatieve analyse mogelijk was. Dit gold onder meer voor de hoeveelheid aandacht die de proefpersonen gaven aan elk van de fasen van het afwegingsproces en voor de mate waarin afwegingen waren gekoppeld aan argumenten.

***7: Het vaststellen van de validiteit***

Wij menen dat de interne validiteit van het onderzoek voldoende is, evenals de inhoudelijke validiteit. Dit is te danken aan de toegepaste methode die een gestructureerde benadering van een complex probleem mogelijk maakt. De begripsvaliditeit is tamelijk laag, gezien de beperkte beschikbare kennis vooraf.
De externe validiteit is laag; studenten werden onderzocht in een laboratoriumcontext. Afwegingsproblemen doen zich in te praktijk voor in organisaties en de actoren die de afwegingen maken hebben een zekere expertise op het betreffende gebied. Maar de mate van interne validiteit die is bereikt maakt dit nadeel zeker goed.

**Hoofdstuk 3: Een model van afwegingsprocessen. De rol van expertise**

Het model van afwegingsprocessen als resultaat van dit onderzoek waarvoor de basis is gelegd in het vorige hoofdstuk, ziet er als volgt uit:

| Nummer van de fase | Naam van de fase | **% segmenten gewijd aan de fase** | **% proefpersonen bij wie elementen van de fase werden waargenomen** |
|---|---|---|---|
| 1 | Probleemidentificatie | 6.74 % | 100 % |
| 2 | Bewerking van (sub-) attributen | 30.33 % | 100 % |
| 3 | Absolute weging van subattributen | 27.22 % | 100 % |
| 4 | Homogene weging van subattributen | 4.53 % | 66.7 % |
| 5 | Heterogene weging van subattributen | 1.50 % | 55.6 % |
| 6 | Weging van hoofdattributen | 12.54 % | 100% |
| 7 | Evaluatie | 17.14 % | 100% |

*Tabel 1. De fasen van het APM*

De fasen kunnen als volgt worden getypeerd:

- **Fase 1: Probleemidentificatie**

  Hierin formuleert de afweger de te maken afweging in zijn of haar eigen woorden. Randvoorwaarden, zoals de eis dat de afweging moet kunnen worden uitgelegd aan het management, kunnen aan de orde komen.

- **Fase 2: Bewerking van attributen**

  De af te wegen attributen worden bewerkt op één of meer van de volgende wijzen:

  - *Splitsing van een attribuut in twee of meer subattributen*. 'Veiligheid' kan bijvoorbeeld worden gesplitst in 'actieve veiligheid' en 'passieve veiligheid';

  - *Integratie*. Dit is het tegenovergestelde van splitsing: het samenvoegen van subattributen;

  - *Concretisering*. 'Veiligheid' kan bijvoorbeeld worden geconcretiseerd als 'het aantal doden en gewonden per miljoen passagierskilometer';

  - *Abstractie*. Dit is het tegenovergestelde van concretisering;

  - *Herformulering*. Hierbij blijft de betekenis van het attribuut gelijk terwijl de aanduiding ervan wordt veranderd. 'Comfort' kan bijvoorbeeld worden geherformuleerd als 'passagierscomfort' of 'reiscomfort'.

- **Fase 3: Absolute weging**

  Hierbij worden attributen afzonderlijk gewogen, zonder ze met elkaar te vergelijken. ‘Veiligheid is zeer belangrijk’ is een absolute weging.

- **Fase 4: Homogene weging van subattributen**

  Hierbij worden subattributen die behoren tot hetzelfde hoofdattribuut tegen elkaar afgewogen. Een voorbeeld van zo’n afweging is ‘de kwaliteit van de gordels is belangrijker dan de remweg’. Beide subattributen vallen onder het hoofdattribuut ‘veiligheid’.

- **Fase 5: Heterogene weging van subattributen**

  Hierbij gaat het om subattributen die behoren tot verschillende hoofdsattributen. Bijvoorbeeld: ‘de kwaliteit van de gordels (een subattribuut van ‘veiligheid’) is belangrijker dan de aanwezigheid van airconditioning (een subattribuut van ‘comfort’)’.

- **Fase 6: Weging van hoofdattributen**

  In de opdracht die de proefpersonen uitvoerden was dit het doel: het afwegen van ‘veiligheid’ tegen ‘comfort’. 13 van de 18 proefpersonen maakten deze afweging, de vijf overige stopten bij fase 5.

- **Fase 7: Evaluatie**

  De evaluatie kan de gekozen gewichten betreffen, de gekozen argumenten, de werkwijze enzovoort.

Naast deze fasen viel een aantal ondersteunende activiteiten te onderscheiden die bij elke fase een rol konden spelen. Op deze activiteiten is niet nader ingegaan.

Het model, verder te noemen het afwegingsprocesmodel (APM) biedt een algemeen kader voor de bestudering van afwegingsprocessen. Daarnaast is een aantal verwachtingen getoetst met betrekking tot het gedrag van leken bij afwegingsprocessen, bekeken vanuit de theorie over expertise.

De basis van onze verwachtingen waren vier dimensies van expertise:

1: De kennisdimensie (‘weten’);

2: De vaardigheidsdimensie (‘kunnen’);

3: De metacognitieve dimensie (‘weten wat je weet en kunt’);

4: De dimensie van sociale erkenning (‘jezelf kunnen presenteren en gezien worden als expert’).

De belangrijkste resultaten zijn de volgende:

1: De proefpersonen (leken) gebruikten algemene probleemoplossingtechnieken zoals het structureren van het probleem (fasen 1 en 2 van het APM) voor met de oplossing ervan te beginnen (fasen 3 t/m 6). Overigens besteedden zij minder aandacht aan het structureren van het afwegingsprobleem dan wij verwachtten van leken voor wie het probleem totaal nieuw is;

2: Overeenkomstig onze verwachting benaderden de proefpersonen het afwegingsprobleem meer in de breedte dan in de diepte. Attributen werden gesplitst in een groot aantal subattributen maar over causale relaties tussen de attributen, nuttig bij het vinden van een gemeenschappelijke noemer, werd nauwelijks nagedacht. Integratie van subattributen kwam nauwelijks voor;

3: Conform onze verwachtingen besteedden de proefpersonen relatief veel aandacht aan evaluatie;

4: In tegenstelling tot wat wij verwachtten hadden de proefpersonen veel vertrouwen in de kwaliteit van hun werk. Het is mogelijk dat de proefpersonen erop vertrouwen dat hun projectmanagementvaardigheden zouden leiden tot een goed resultaat en dat zij de vaardigheden bezitten om dot resultaat overtuigend te verdedigen (dimensies 2 en 4). Ook een gering niveau van metacognitie (dimensie 2) kan een verklaring zijn.

Op basis van de resultaten en van de theorie op het gebied van expertise zijn enkele verwachtingen te formuleren over hoe experts in de praktijk afwegingsproblemen aanpakken. Wij verwachten dat ze:

- expliciete definities van subattributen gebruiken waarvoor afwegingsheuristieken kunnen worden gevonden;
- streven naar compleetheid van subattributen met vermijding van redundantie;
- subattributen integreren teneinde gemeenschappelijke noemers te vinden voor zoveel mogelijk subattributen;
- slechts beperkte aandacht besteden aan evaluatie.

**Hoofdstuk 4: Structurering en relaties tussen fasen bij afwegingsproblemen**

Het afwegingsprocesmodel (APM) dat is gepresenteerd in hoofdstuk 3 heeft een zekere ‘gezond verstand’ logica. Het lijkt bijvoorbeeld voor de hand te liggen om eerst subattributen te wegen en dan pas hoofdattributen. Immers, als de gewichten van de

hoofdattributen bekend zijn is de afweging voltooid en heeft het wegen van subattributen geen toegevoegde waarde meer. Hetzelfde geldt voor het bewerken van attributen; dit gebeurt ten behoeve van de weging en dus ervóór. De fasen bouwen dus op elkaar voort; de input van een fase wordt gevormd door de output van de voorgaande fasen. Maar hoe aantrekkelijk deze logica ook lijkt, de vraag is of hij daadwerkelijk wordt gevolgd.

We benaderen dit probleem door twee ideaaltypische soorten afwegingsgedrag te definiëren en te onderzoeken welk van de twee soorten gedrag onze proefpersonen vertonen. Deze twee soorten gedrag zijn in de literatuur over besluitvorming en probleemoplossen op vele wijzen omschreven. Wij duiden ze aan als 'rationeel' en 'incrementeel'. Rationeel gedrag kenmerkt zich door doelgerichtheid, een gestructureerde en planmatige aanpak, sequentiële uitvoering van deelactiviteiten (fasen) van een taak, en een streven de 'objectief' meest optimale oplossing te vinden. Incrementeel gedrag kenmerkt zich door het al zoekende komen tot een acceptabele oplossing, waarbij weinig vooraf wordt gepland, voortdurend gedrag wordt aangepast aan de omstandigheden, er dus niet vooraf een helder beeld is van het einddoel en de weg daarheen en de relaties tussen de verschillende activiteiten onduidelijk zijn. We formuleren verwachtingen voor beide soorten gedrag en bestuderen de proefpersonen volgens de methode die is beschreven in hoofdstuk 2. De zeven fasen van het APM worden verdeeld in een structureringscluster (fasen 1 en 2), een wegingcluster (fasen 3 t/m 6) en een evaluatiecluster (fase 7). Wij gaan in dit hoofdstuk niet in op dit laatste cluster.

**Het structureringscluster**

Het gedrag van de proefpersonen was vrijwel geheel incrementeel. Alle verwachtingen met betrekking tot incrementeel gedrag werden overtuigend bewaarheid.
Dit hield in:

1: De proefpersonen vertaalden de opdracht niet in eisen aan het eindresultaat;

2: De proefpersonen bewerkten de attributen niet systematisch. Attributen werden niet gedefinieerd. Meestal bleek hun betekenis uit de subattributen waarin ze werden gesplitst;

3. Het begrip 'gewicht' werd niet gedefinieerd.

Dit alles betekent dat de output het structureringscluster niet goed aansloot op de vereiste input van het wegingcluster.

**Het wegingcluster**

In het wegingcluster gaat het om de kwantitatieve en kwalitatieve relaties tussen de in de verschillende fasen toegekende gewichten.
Indien er kwantitatieve of kwalitatieve relaties bestaan tussen de fasen van het wegingcluster is het aannemelijk dat de fasen op elkaar voortbouwen. Maar er was geen enkele relatie, kwantitatief of kwalitatief, vast te stellen tussen de verschillende fasen van

het wegingcluster. Ook dit is in overeenstemming met incrementeel gedrag. De proefpersonen besteedden veel aandacht aan absolute weging (fase 3). Hieruit kwamen verzamelingen belangrijke en minder belangrijke subattributen voor beide hoofdattributen. Maar waarom het ene hoofdattribuut nu uiteindelijk belangrijker was dan het andere kon op basis van de resultaten van fase 3 niet worden vastgesteld.

Als de fasen van het APM onderling zo weinig samenhangen leidt dit tot de vraag wat de functies van de fasen voorafgaand aan fase 6 dan zijn. Ons vermoeden is dat het structureringscluster en de absolute weging dienen als 'framing': als middel voor de proefpersonen om voor zichzelf duidelijk te krijgen wat zij, al is het impliciet, verstaan onder de af te wegen attributen, en om argumenten te vinden waarom attributen al dan niet belangrijk zijn. Dit is een interessante vraag voor vervolgonderzoek.

## Hoofdstuk 5: Beperkte rationaliteit bij afwegingen

Het wegen van attributen is, in theorie, uiterst complex. Voor alternatieven waaruit in het kader van een beslissing moet worden gekozen kan een zeer groot aantal attributen worden geformuleerd waarop ze kunnen of moeten worden beoordeeld. Elk van die attributen kan weer een groot aantal subattributen omvatten. Ook is in principe een oneindig aantal gewichten mogelijk.
Het is duidelijk dat geen enkele afweger alle mogelijke combinaties van gewichten en attributen kan beoordelen op hun geschiktheid. Hij of zij zal dus slechts een deelverzameling bekijken. Het slechts een beperkt aantal alternatieve oplossingen (in ons geval attributen en gewichten) in overweging nemen, in plaats van alle mogelijke oplossingen, wordt 'beperkte rationaliteit' ('bounded rationality') genoemd. De probleemstelling van dit hoofdstuk is:

> *Welke specifieke (expliciete of impliciete, bewuste of onbewuste) vormen van beperkte rationaliteit kunnen worden onderscheiden bij niet-routinematige afwegingsprocessen ten behoeve van organisatiebeslissingen, en hoe vaak komen deze vormen voor?*

### Systematieken bij de keuze van attributen

Bij het kiezen van attributen die moeten worden afgewogen moet worden gelet op de volgende aspecten:

1: Compleetheid moet woeden nagestreefd; alle attributen die van belang kunnen zijn bij de keuze van een alternatief moeten worden meegenomen;

2: Redundantie van attributen moet worden vermeden. Dat wil zeggen: attributen die de keuze van het alternatief niet beïnvloeden moeten niet worden meegenomen, onder meer omdat de complexiteit van de afweging daardoor onnodig wordt vergroot;

3: Attributen moeten onderling onafhankelijk zijn; er mag geen sprake zijn van interdependentie. Twee attributen A en B zijn onderling afhankelijk als de score van een alternatief op attribuut A de score op attribuut B beïnvloedt. Dit is bijvoorbeeld het geval als A gedeeltelijk samenvalt met B. De complexiteit van de afweging wordt dan vergroot omdat de onderling onafhankelijke componenten moeten worden gescheiden van de afhankelijke component.

**Systematieken bij de keuze van gewichten**

Gewichten zijn waarden op een schaal van belangrijkheid. In de gekozen gewichten kan sprake zijn van een systematiek. Het aantal mogelijke gewichten kan ook worden beperkt door gebruik van één of meer positioneringmethoden, waardoor een deel van het gewichtsinterval buiten beschouwing wordt gelaten. Bijvoorbeeld: een afweger kan vaststellen dat weliswaar niet duidelijk is hoe belangrijk de veiligheid van een minibusje is in verhouding tot het comfort, maar dat veiligheid in ieder geval de belangrijkste is van de twee.
Als het aantal in overweging te nemen gewichten wordt beperkt is het mogelijk dat een afweger nagaat of de overblijvende gewichten compleet zijn (of alle nodige gewichten beschikbaar zijn). Deze compleetheid betreft zowel het bereik van de gewichten als de precisie (de afstand tussen twee opeenvolgende gewichten). Verder is de vraag aan de orde of gewichten worden weggelaten omdat ze redundant zijn dan wel zonder duidelijke reden Een laatste vraag is of, als proefpersonen meerdere schalen en/of positioneringmethoden gebruiken, zij aandacht besteden aan mogelijke interdependenties (die kunnen leiden tot strijdigheid van schalen of positioneringmethoden).

**Resultaten: de attributen die in overweging worden genomen**

Geen van de proefpersonen besteedde meer dan incidentele aandacht aan compleetheid, redundantie en interdependentie of gaf aan zich er zorgen over te maken. De proefpersonen genereerden weliswaar op basis van eigen kennis en ervaring en op basis van het uitgereikte informatiemateriaal grote hoeveelheden attributen, maar er leek geen duidelijk criterium te bestaan voor het gewenste aantal of soort attributen.

**Resultaten: de gewichten die in overweging werden genomen**

Alle proefpersonen op één na gebruikten een duidelijke systematiek bij het beperkte aantal in overweging genomen gewichten. De helft gebruikte één of meer positioneringmethoden. De gebruikte systematieken worden weergegeven in de onderstaande tabel.

| Systematiek | % proefpersonen dat de systematiek gebruikte |
|---|---|
| **Schalen** | |
| 1: Beperkt aantal ordinale waarden (bijvoorbeeld A, B en C) | 22% |
| 2: Beperkt aantal waarden op een schaal van 0 tot 1 of 1 tot 10 (bijvoorbeeld intervallen van 0,1) | 50% |
| 3: Beperkt aantal verhoudingsgetallen (A is 1, 2, 3 maal zo belangrijk als B) | 22% |
| **Positioneringsmethoden** | |
| 4: Beginnen met een extreme waarde (A is veel belangrijker dan B), dan naar een middenwaarde zoals A is even belangrijk als B) | 17% |
| 5: Beginnen met een middenwaarde, dan naar een extreme waarde | 33% |
| 6: Ankering (in dit geval de 20-80 regel) | 5% |

*Tabel 2: Systematieken van in overweging genomen gewichten en positioneringmethoden*

Sommige respondenten gebruikten combinaties van schalen en positioneringmethoden waardoor het aantal in overweging te nemen gewichten aanzienlijk werd beperkt.

Geen van de proefpersonen maakte zich druk over compleetheid of redundantie van de gebruikte gewichten. Een definitieve verklaring kon niet worden gevonden, maar de gebruikte schalen waren zo algemeen bekend dat de compleetheid ervan als vaststaand kon worden aangenomen. De positioneringmethoden waren een uiting van ervaren redundantie, maar zij werden toegepast zonder argumentatie op dit punt.

De conclusie is dat binnen de beperkte rationaliteit bij afwegingen er geen systematiek is te herkennen in de keuze van af te wegen attributen, maar wel in de gebruikte gewichten. Dit opent perspectieven voor verder onderzoek, mede gericht op het ontwikkelen van instrumenten om de keuze van af te wegen attributen systematischer te laten verlopen. Aanname hierbij is dat zo de kwaliteit van het afwegingsproces kan worden verbeterd, maar deze aanname zal moeten worden getoetst.

## Hoofdstuk 6: Afsluiting

Dit onderzoek heeft een eerste inzicht verschaft in de structuur van afwegingsprocessen. De gebruikte onderzoeksmethode heeft afdoende interne validiteit opgeleverd en lijkt te kunnen worden gebruikt voor onderzoek naar andere cognitieve processen. Het APM bleek kwantitatief toetsbaar en kan worden gebruikt om afwegingsprocessen verder te bestuderen.
Maar het onderzoek heeft veel nieuwe vragen opgeroepen, die inspiratie kunnen zijn voor verder onderzoek.

Enkele mogelijkheden zijn:

1: Het onderzoek herhalen bij andere populaties, zoals experts in de bedrijfskundige praktijk;

2: Het ontwikkelen van instrumenten waarmee de variabelen die van belang zijn bij afwegingsprocessen nauwkeuriger kunnen worden gemeten dan in dit onderzoek is gedaan;

3: Instrumenten ontwikkelen om de kwaliteit van afwegingsprocessen te verbeteren, en de effectiviteit van deze instrumenten vaststellen. Met name het structureringscluster lijkt mogelijkheden te bieden, bijvoorbeeld met betrekking tot het stimuleren van systematische splitsing, het vaststellen van causale relaties tussen (sub)attributen en het op basis hiervan integreren van deze (sub)attributen;

4: Het onderzoeken van de wijze waarop managers multidisciplinaire informatie integreren, gebaseerd op de onderzoeksmethode die is gebruikt in dit proefschrift, en op het ontwikkelde model van afwegingsprocessen. Zo kan inzicht worden verkregen in een kerngebied van bedrijfskunde als multidisciplinaire wetenschap.

# Curriculum Vitae

Hans Heerkens (1958) is assistant professor at the department of Business, Public Administration and Technology of the University of Twente, the Netherlands. He teaches on research and problem-solving methodology and on management in the aerospace industry. Over the years, he published several articles in scientific journals on these subjects. For the Open University, he participated in the writing of courses on public administration and management. He is a member of the American Institute for Aeronautics and Astronautics and a freelance writer for aerospace trade magazines. At the post-academic education institute TSM Business School he teaches in-company courses on problem solving and for the Dutch Association for Research on Management (NOBEM) he participates in methodology courses for PhD-students. Hans is a member of the Scientific Council of the VNL (the association of Dutch aviation and aerospace organizations) and of the Dutch Advisory Board of the Knowledge Center for Aviation. He acts as coach for methodological aspects of research on aerospace policy, for example for the DGL (Dutch Civil Aviation Authority).