Decision making in conceptual engineering design: an empirical investigation

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: GIROD, M. ... et al., 2003. Decision making in conceptual engineering design: an empirical investigation. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 217 (9), pp. 1215-1228

Additional Information:

- This article was published in the journal, Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture [© Professional Engineering Publishing] and is available at: http://journals.pepublishing.com/content/119784/?sortorder=asc&p_o=52

Metadata Record: https://dspace.lboro.ac.uk/2134/4521

Version: Published

Publisher: © Professional Engineering Publishing

Please cite the published version.
This item was submitted to Loughborough’s Institutional Repository (https://dspace.lboro.ac.uk/) by the author and is made available under the following Creative Commons Licence conditions.

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/
Decision making in conceptual engineering design: an empirical investigation

M Girod1, A C Elliott2, N D Burns3 and I C Wright4*
1Airbus Deutschland GmbH, Hamburg, Germany
2IMPACT Research Group, Department of Computer Science, Loughborough University, Loughborough, Leicestershire, UK
3Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, Leicestershire, UK
4School of Engineering, Coventry University, Coventry, UK

Abstract: This paper describes the aims, methodology and results of an empirical research project. The aim was to gather new insight about how collaborative decision-making processes in the conceptual engineering design phase of the product development process takes place. The insight contributes to bridging the gap between the theoretical development of new decision-making methods in academia and the needs of practitioners. The researchers observed, recorded and transcribed three workshops involving groups that were engaged in the collaborative decision-making processes. Through analysing the transcripts, a set of decision-making related activities was identified. Each activity’s time consumption was quantified, and a number of decision-making process models at different levels of detail were developed. During this analysis, various observations on particular process characteristics were made. It was observed that generating formal structures and documentation was beneficial to the decision-making process. This was particularly true with respect to developing problem understanding and consistency. The outcomes should be valuable for the design and development of improved decision-support systems.

Keywords: engineering design, decision-making process, empirical research, activities, process model

1 INTRODUCTION

The engineering design process can be seen as a series of interrelated operations that is driven by decisions[1]. Thus, an effective design process relies heavily upon effective decision making. As a consequence, supporting decision making can play a significant role in achieving design process improvements. The research project described in this paper aimed to improve the product development process by contributing towards the development of context-independent decision-making support. Of particular interest was decision making in selection problem situations during the concept design phase.

The authors reviewed[2] a number of methods from different research disciplines that may be beneficial in supporting the resolution of selection problems during conceptual design. However, although a variety of relevant methods have been developed in academia, they are not used widely in industrial practice[3]. On the other hand, it has been realized that the selection of engineering design concepts causes problems in many practical cases[4] and that much product development time is wasted by making poor decisions[5]. This indicates that practitioners could indeed benefit from effective decision-making support. It was not the authors’ intention to develop a new method. Instead, the overall project, of which this paper is a part, took a step back and sought to identify practitioners’ needs empirically, i.e. the requirements for decision-making support. This would help to establish a sound foundation upon which available or new methods might be critically assessed.

The approach taken to identify decision-making support requirements was analysis of empirically established decision-making process models. The overall project therefore comprised two main stages: firstly, establishing the models and, secondly, analysing these models. Within this paper the authors describe their decision-making process models and their elements (activities), in addition to discussing observations made during analysis of the models. A full discussion on the decision-making
support requirements that were identified as the result of the project’s second stage is beyond the scope of this paper and will be the subject of another publication.

To set the background for this paper some discussion of currently available decision-making support methods is required. This discussion concentrates on methods that specifically address the needs of selection problems in product development applications. This is followed by an outline of the authors’ own research methodology. The remainder of the paper introduces the resulting models and discusses the observations made.

2 DECISION-MAKING METHODS

The highly axiomatic multi-attribute utility theory (MAUT) [6] was developed in 1976. MAUT was not developed in the context of engineering design, but emerged from operational research (OR). Similar principles are used by rating methods based on quantitative evaluation matrices. These rating methods are widely known in engineering design, at least in design education, as they are more straightforward to use, which is an advantage. This, in particular, is because they usually do not apply MAUT’s utility functions, which associate utility scores with alternative performances. Instead, the utility scores are directly assigned. Because no explicit utility functions are used, the axiomatic basis of rating methods is weaker than in MAUT, which may be considered as a drawback. Both of these are examples of purely quantitative methods, all of which have a common disadvantage, i.e. all qualitative/subjective information has somehow to be quantified, which is not always easy and accurate.

Also based on matrices are the method of controlled convergence [7] and decision-making with QFD (quality function deployment) [8], which are both qualitative. In Pugh’s method [7] all alternatives are qualitatively compared with a datum. The merits are: (a) the method is straightforward and easy to use and (b) it is not necessary to quantify qualitative information. The drawbacks are: (a) all criteria are assumed to be equally important and (b) all available quantitative information has to be transformed into qualitative statements, which implies a loss of information. Decision making with QFD also involves a qualitative comparison of all alternatives to a datum. Additional advantages of this method over Pugh’s work are that it introduces weights for the criteria and also there is a mechanism that can be supportive in building a strong relation between decision criteria and customer requirements.

The methods described above assume that the available information is known with certainty. However, this assumption is not always valid. There are some very sophisticated methods for the support of design decision-making problems, which can model different forms of uncertainty in information. For the engineering decision support system, Herling [9] uses a method that applies the Bayesian uncertainty representation. Yang and Sen [10] developed a method based on ‘evidential reasoning in design’ using the Dempster–Shafer theory of evidence. The advantage of these two approaches is their capability of modelling subjective belief regarding the expected performance of alternatives. However, they are not able to model possible uncertainty produced by ‘vagueness’, e.g. linguistic imprecision.

Vagueness can be modelled by methods based on fuzzy sets. An example is discussed by Thurston and Carnahan [11]. A merit of fuzzy methods is that linguistic imprecision is recognized as an element of the decision model. This means that human decision makers can express themselves in a natural way, without being required to force their natural expressions into other types of information format. The drawback of methods that are capable of modelling uncertainty (subjective belief or vagueness) is the relatively high effort required for modelling the decision situation and processing the input information.

A number of other decision-making methods have emerged from multicriteria decision-making research. Prominent examples, apart from the MAUT discussed above, are the outranking methods, such as Promethee [12], Electre [13] and the analytic hierarchy process [14]. Detailed discussions on these and other methods are available [2, 15].

3 RESEARCH APPROACH

Establishing a descriptive theory that explains the behaviour of decision-making processes in conceptual design required the authors to develop an understanding of the process. This was done by generation of a model that describes the decision-making activities as they take place. The model is the outcome of analysing appropriate processes that were observed under controlled experimental conditions.

One way of ensuring a thorough analysis of a particular process is by using a researcher’s personal observations resulting from their presence in the process being investigated [16]. Research approaches that place the investigator into the field or the natural setting of the phenomena under investigation are usually labelled ‘ethnographic’. This type of research is in contrast to approaches where an investigator identifies a sample of subjects and brings them into the study through the use of questionnaires or interviews. A disadvantage of ethnographic investigations is their large time consumption [17].

Despite this, the authors favoured such an approach for this study because ethnography allows for the gathering of highly detailed data [18].

Initial attempts to set up an industry-based ethnographic study encountered a serious problem: potential partners were reluctant to provide access to their organizations for research purposes. The reason was that the
organizations considered concept phase decision-making meetings to be too commercially confidential to allow observation and recording. Therefore, a decision was made to base the research on observations from case studies that were conducted under laboratory conditions in organized workshops. Apart from practicality, this had the advantage of reducing the complexity of the study and increasing the level of control, i.e., to keep certain variables constant across a number of workshops. These variables related to the environment in which the decisions were taken and the context, which are the available alternatives and criteria. Keeping these variables constant allowed for comparison of results across a number of workshops and between different groups. Similar projects in engineering design, also using university-based observational studies, have previously been undertaken [19, 20].

3.1 Data collection

For the data collection, five workshops were organized that involved groups of between three and five members. Four of these groups were final year engineering design students and one group consisted of professional engineers. Therefore, there were two values regarding the variable ‘professional status of group members’. This gave the opportunity to study effects that may be caused by differences in the participants’ professional experience.

The first two of these five workshops were run as pilot studies. These aimed at testing the workshops’ general set-up and the equipment used. The three workshops that were to be transcribed and analysed fully then followed. These contained two student groups and the group of professional engineers.

All groups were given an assignment that they had to complete in a manner of their choice. They were asked to evaluate a set of conceptual design alternatives and to select, as a group, the most effective one. The groups were provided with the alternatives and some evaluation criteria. Two groups (one student and the ‘professional’ group) chose to use a formal approach, involving evaluation matrices and the use of numerical scales to evaluate the importance of various assessment criteria and solution suitability, and one group (students) chose to use a completely informal subjective approach to resolve their assignment. Therefore, there were two values regarding the variable ‘chosen approach’. This provided the opportunity to study effects that may be caused by differences in the general decision-making approach.

The alternatives to be evaluated had been generated through brainstorming sessions in a previous and unrelated industrial study. A major British company had approached the University for help regarding the resolution of a design problem. The company had provided criteria for an evaluation of alternative solutions. Additional evaluation criteria were identified while the solutions were generated through brainstorming. For this study’s workshops, a subset of the generated solutions and the company’s criteria was reused, as well as the additionally identified criteria. All workshop participants had also been involved in these previous brainstorming sessions.

Since the methodology involved the study of groups who used discussions to solve their assignment, the workshops could be recorded without the need to ask them to ‘think aloud’. Recordings were made using video tape as well as audio tape and mini disc. A researcher was present for the entire duration of all workshops to observe the conversations as well as to oversee the recording equipment. Each workshop took between 45 and 80 minutes. The recordings were used to prepare full transcripts for further analysis.

3.2 Data analysis

The aim of analysing the transcripts was to generate a descriptive model of the observed decision-making processes. Such a process model was seen to represent a structured representation of relationships between the process elements. These process elements are the activities carried out by the group members to fulfil the given assignment and make a selection of the best design alternative from a number of alternatives. To identify the activities, content analysis of the transcripts was undertaken. The purpose of content analysis was to define categories of interest and then assign the syntactic data to them, the many words in the gathered data (the transcripts) being transferred into many fewer categories of meaning (the activities) [21]. To gain an initial orientation for the content analysis, use was made of a number of previous studies [19, 20, 22, 23] that generated a variety of activity categories for the engineering design process. However, the authors could not find any study that had the same focus as the one described here. Therefore, it was not possible to simply reuse any of the sets of categories established in previous studies.

Categories were defined in an iterative manner; i.e., during the analysis of transcripts, the set of categories was repetitively refined by adding new categories and by further detailing. This stopped when it became possible to apply one coherent set of categories to all three transcripts. By using this so-called ‘replication strategy’ [24], it was possible to identify a set of activity categories that was generally applicable for all three workshops.

The activity’s time consumption for each group was also quantified. This was used as a metric for comparing the groups’ decision-making processes. Having identified the activities, a descriptive decision-making process model was developed using pattern coding [25]. Pattern
coding was used to identify constructs that comprised particular patterns of activities; i.e. emergent themes or configurations were identified by pulling together material into units of meaning. In this research, a two-level pattern coding approach was used. At the upper level, relationships between constructs were identified. It was realized that the constructs represented process steps and that the relationships between the constructs provided a decision-making process model. This model showed that particular process steps reappeared in the process as repeated steps or loops. The lower level of pattern coding concentrated on these reappearing steps and the activity patterns within them. An analysis was undertaken of whether the pattern of activities for these steps was exactly the same every time it appeared in the process. The activities identified, the whole decision-making process model and the activity patterns within the reappearing process steps are discussed in the following sections.

4 ACTIVITY CATEGORIES

This section introduces the set of activities, including subactivities that were identified. Basically, these activities represent particular parts of the multifaceted

<table>
<thead>
<tr>
<th>Activities</th>
<th>Subactivities</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discussing the process approach</td>
<td></td>
<td>The group determines how to work through the assignment of evaluating a number of design alternatives and selecting one</td>
</tr>
<tr>
<td>Discussing the general process approach</td>
<td></td>
<td>Context independent approach: could be applied to any set of design alternatives and any set of criteria</td>
</tr>
<tr>
<td>Discussing the specific process approach</td>
<td></td>
<td>Context dependent approach: supplements general process approach and is only relevant for the specific activities and criteria at hand</td>
</tr>
<tr>
<td>Identifying criteria</td>
<td></td>
<td>The group identifies additional evaluation criteria</td>
</tr>
<tr>
<td>Defining criteria</td>
<td></td>
<td>The group tries to find a consistent understanding of the meaning and relevance of evaluation criteria</td>
</tr>
<tr>
<td>Weighting criteria</td>
<td></td>
<td>The group states the level of importance for evaluation criteria</td>
</tr>
<tr>
<td>Weighting criteria informally</td>
<td></td>
<td>Using linguistic quantifiers, such as ‘quite important’</td>
</tr>
<tr>
<td>Weighting criteria formally</td>
<td></td>
<td>Using a formal, pre-defined scale</td>
</tr>
<tr>
<td>Clarifying concept working principles</td>
<td></td>
<td>The group makes sure that they understand the working principles of the design alternatives being evaluated</td>
</tr>
<tr>
<td>Clarifying the product environment</td>
<td></td>
<td>The group makes sure that they understand what the environmental conditions are that the final product has to operate in</td>
</tr>
<tr>
<td>Determining the product environment</td>
<td></td>
<td>Identifying what the environmental conditions are</td>
</tr>
<tr>
<td>Making assumptions on the product environment</td>
<td></td>
<td>If the environmental conditions cannot be identified, assumptions are made</td>
</tr>
<tr>
<td>Deliberating subissues</td>
<td></td>
<td>The group identifies how an alternative behaves for a particular subissue (subproblem)</td>
</tr>
<tr>
<td>Discussing subissues</td>
<td></td>
<td>Generate and analyse solutions for subissue</td>
</tr>
<tr>
<td>Accepting assumptions about subissue solutions</td>
<td></td>
<td>Explicitly accept a generated solution for a subissue and assume it to be part of one of the design alternatives</td>
</tr>
<tr>
<td>Gaining external information</td>
<td></td>
<td>The group gathers information that cannot be generated by them</td>
</tr>
<tr>
<td>Determining or evaluating performances</td>
<td></td>
<td>The group expresses the expected performance of an alternative and/or evaluates the performance</td>
</tr>
<tr>
<td>Determining or evaluating restricted performances informally</td>
<td></td>
<td>Using informal statements to address an alternative’s performance with respect to a particular evaluation criterion (restricted performance); a preference may or may not be communicated</td>
</tr>
<tr>
<td>Evaluating comprehensive performances informally</td>
<td></td>
<td>Using informal statements to address an alternative’s performance with respect to the entire set of evaluation criteria (comprehensive performance); a preference is communicated</td>
</tr>
<tr>
<td>Evaluating restricted performances formally</td>
<td></td>
<td>Using a formal scale for a statement on an alternative’s restricted performance; a preference is communicated</td>
</tr>
<tr>
<td>Evaluating comprehensive performances formally</td>
<td></td>
<td>Using a formal scale for a statement on an alternative’s comprehensive performance; a preference is communicated</td>
</tr>
<tr>
<td>Raising evidence</td>
<td></td>
<td>The group attempts to justify different types of statements</td>
</tr>
<tr>
<td>Raising evidence on restricted performance of concepts</td>
<td></td>
<td>Gathering evidence for an alternative’s expected performance with respect to a particular evaluation criterion (restricted performance)</td>
</tr>
<tr>
<td>Raising evidence on comprehensive performance of concepts</td>
<td></td>
<td>Gathering evidence for an alternative’s expected performance with respect to the entire set of criteria (comprehensive performance)</td>
</tr>
<tr>
<td>Raising evidence on criteria weights</td>
<td></td>
<td>Gathering evidence for the importance weight of an evaluation criterion</td>
</tr>
<tr>
<td>Mapping intuition on to ranking</td>
<td></td>
<td>Group members express whether the established group ranking matches their individual intuitive rankings</td>
</tr>
<tr>
<td>Controlling the process</td>
<td></td>
<td>The group manages (steers) the process without adding content</td>
</tr>
</tbody>
</table>
discussions by the groups during the workshops. Each of these parts addresses some aspect of the decision-making process in a particular way. For example, whenever a group member made a remark like ‘… it’s absolutely crucial that this thing is going to be light-weight’, they informally weighted the importance of an evaluation criterion. Such a remark was assigned the activity category ‘weighting criteria informally’. Table 1 lists the entire set of activity categories and delivers some explanations on their meaning. A more detailed discussion on the meaning of these activities is available [26]. As described earlier, ‘informal’ is defined as the use of linguistic quantifiers like ‘quite large’ or ‘fairly heavy’, while ‘formal’ implies the use of an agreed pre-defined scale to assess criteria and performance (see Table 1).

4.1 Reliability of activity categories

The set of activity categories had only fully evolved when the content analysis of the third transcript was completed. To check the reliability of the categories with respect to stability (i.e. their ability to be reproduced) [21], the content analyses of the first and the second transcript was repeated with the fully evolved set of categories. The categories were found to be applicable to each transcript, which indicated their stability. A further reliability check of categories referred to accuracy. Accuracy is assessed by comparing the results of one coder with an established norm [21]. As an ‘established norm’, the results of a study by Dwarkanath and Wallace [27] were used. This study was used to assess accuracy because their research aim of ‘understanding and supporting design decision-making processes’ was very similar to the research aims of this study.

Dwarkanath and Wallace’s [27] observations support several activity categories identified in this study. In particular, these are ‘discussing the process approach’ (their ‘observation 5’), ‘identifying criteria’ (their ‘observation 4’), ‘deliberating sub-issues’ (their ‘observation 1’), ‘gaining external information’ (their ‘observation 5’), ‘raising evidence’ (their ‘observation 5’) and ‘determining or evaluating performances’ (their ‘observation 6’). Not all of the activity categories identified in this research are supported by the findings of Dwarkanath and Wallace. This may be explained by the different research foci and the different assignments given to their workshop participants. In their study, participants were given the assignment to generate a design solution, which is different from the participants’ assignment in this research. Thus, Dwarkanath and Wallace observed procedures in the early design stage. However, this research has a more detailed focus, concentrating on a specific procedure, which is the design process aspect of selecting a conceptual alternative. Even so, it is important to note that none of the activity categories identified by Dwarkanath and Wallace are opposed by the findings of this research.

4.2 Time consumption of activity categories

When the activity categories were identified it was noticed that they had very different time consumptions across the three groups’ decision-making processes. This was seen as an interesting process characteristic that quantified the time that the different categories consumed. For each category, the duration of all its occurrences throughout each group’s decision-making process was summed. The results for all three groups are shown in Fig. 1. The time consumed by the sub-activities was also quantified. This is discussed in detail by Girod et al. [26].

In general, it was decided to concentrate on the relative, rather than the absolute, time consumptions. This allows for comparisons across the three groups’ decision-making processes. In particular, it facilitated the study of whether the variables ‘professional status of group members’ and ‘decision-making approach’ had any effect on the different categories’ time consumptions.

5 DECISION-MAKING PROCESS MODELS

As described in section 3, two levels of pattern coding were applied to the transcripts. The first level revealed a general decision-making process model. This model shows the steps of the three observed decision-making processes including variants. The steps were labelled with names for the tasks that the groups tackled as part of their assignment. The second level revealed more detailed models for specific steps in the general decision-making process. The detailed models show which activities may be contained within particular steps.

5.1 General decision-making process model

The general decision-making process model, shown in Fig. 2, indicates that the groups went through three main process steps, which are: step 1, ‘prepare the process’; step 2, ‘explore criteria and alternatives’ and step 3, ‘conclude the process’. Step 1 consisted of a discussion on the process approach. The groups identified subsequent tasks and prepared to resolve them. Step 2 consisted of investigating relevant evaluation criteria (involving the exploration of their structure and meaning) and evaluating the alternative design concepts (involving the exploration of the alternatives’ performances). Step 3 consisted of ranking the alternatives’ overall performances and analysing the ranking. This involved aggregating the information that was gathered during the previous steps and making a final decision to conclude the process.

Two variants of how step 2 was approached were defined. It was decided to call variant 2A ‘comprehensive approach’ and variant 2B ‘criteria-based approach’.
The comprehensive approach (2A) had an informal character. It was adopted by one of the student groups. This approach comprised step 2A.1 ‘evaluate alternative comprehensively’ and step 2A.2 ‘support comprehensive evaluation’. In step 2A.1, an alternative was first evaluated with respect to its comprehensive (overall) performance. Then, in step 2A.2, the preceding comprehensive evaluation was supported by gathering particular advantages and disadvantages. This sequence was repeated for each alternative.

The formal criteria-based approach (2B) was adopted by the other student group and the group of professional engineers. It comprised step 2B.1 ‘investigate criteria’ and step 2B.2 ‘evaluate alternatives restrictedly’. The evaluation of alternatives in step 2B.2 was based on a criteria structure developed in step 2B.1. The groups
evaluated how the alternatives performed with respect to each criterion individually. Such evaluations were called ‘restricted evaluations’.

In any case, whether the alternatives were evaluated using the comprehensive approach or the criteria-based approach, all three groups established a comprehensive ranking order across all alternatives in step 3.1. If the criteria-based approach had been applied for each alternative a comprehensive performance needed to be determined before a ranking could be established. This was done by numerically aggregating the results of the restricted evaluations as generated in step 2B.2. Once a ranking order was established the groups proceeded by analysing the ranking in step 3.2. This analysis either concentrated on: (a) deliberating various subissues with reference to the highest ranked alternative or (b) on mapping group members’ individual, intuitive rankings on to the one established by the entire group. The former type of analysis seemed to have the purpose of thoroughly checking the highest ranked alternative’s feasibility. The latter type of analysis seemed to have the purpose of checking whether all individual group members were satisfied with the ranking as established by the entire group.

As can be seen in Fig. 2, the two steps comprising the criteria-based approach were broken down into further substeps.

Step 2B.1 ‘investigate criteria’ consisted of the substeps 2B.1.1 ‘structure criteria’, 2B.1.2 ‘pick a criterion’ and 2B.1.3 ‘determine importance of criterion’. Within the first substep the decomposition of the given evaluation criteria into subcriteria was discussed and evolved. Also, additional relevant evaluation criteria were identified, and the groups tried to clarify the meaning of all criteria. Within the following two substeps the importance of the criteria, as perceived by the entire group, was determined for each criterion in turn.

Step 2B.2 ‘evaluate alternatives restrictedly’ consisted of three substeps, ‘pick an alternative’, ‘pick a criterion’ and ‘evaluate alternative–criterion pair’. These substeps were sequenced by two different variants. In variant 2B.2a, which was called ‘alternative–criteria’, one alternative is picked and then evaluated with respect to all criteria before the next alternative is picked. At variant 2B.2b, which was called ‘criterion–alternatives’, one criterion is picked and then used for the evaluation of all alternatives before the next criterion is picked. Substeps 2B.2a.3 and 2B.2b.3 are basically the same tasks, to evaluate an alternative with respect to a criterion.

5.2 Detailed models

There were three substeps that were most interesting because they involved the largest variety of activity categories: substeps 2A.2 ‘support comprehensive evaluation’, 2B.1.3 ‘determine importance of criterion’ and 2B.2a/b.3 ‘evaluate alternative–criterion pair’. For these three substeps detailed models were established. The models for substep 2A.2 ‘support comprehensive
evaluation’ and for substep 2B.2a/b.3 ‘evaluate alternative–criterion pair’ were found to be very similar [28]. This paper only discusses the more complex of the two, which is substep 2B.2a/b.3 ‘evaluate alternative–criterion pair’, along with the other substep of interest, 2B.1.3 ‘determine importance of criterion’.

5.2.1 11.1 Step 2B.1.3: determine importance of criterion

The pattern of activities that were carried out within the step ‘determine importance of criterion’ is modelled in Fig. 3. The overall aim of this step was to express the importance of an evaluation criterion by establishing a formal weighting. Whenever the importance of a criterion was expressed, at least some of the activities in Fig. 3 were carried out. There was no obvious time sequence of activities. An exception is ‘controlling the process’, which most often initiated the substep ‘determining importance of criterion’. This was done by a control remark, e.g. ‘Let’s look at the importance of …’. The activity ‘weighting criteria formally’ was always carried out when a criterion’s importance was formally determined. The other activities in Fig. 3 were carried out sometimes.

Although there was no obvious time sequence, an apparent functional sequence of activities was perceived. These were ‘defining criteria’ and ‘clarifying the product environment’, which informed the activity ‘raising evidence on criteria weights’. This, in turn, generated a justification for the outcome of the activity ‘weighting criteria formally’. This outcome, i.e. an informal weighting, was then mapped on to a formal scale by the activity ‘weighting criteria formally’.

5.2.2 11.2 Step 2B.2a/b.3: evaluate alternative–criterion pair

The pattern of activities that were carried out within the step ‘evaluate alternative–criterion pair’ is shown in Fig. 4. The overall aim of this step was to evaluate an alternative with respect to an individual criterion (restricted evaluation). Whenever an alternative–criterion pair was evaluated, at least some of the activities in Fig. 4 were carried out.

As with determining the importance, evaluation had no obvious time sequence of activities apart from a control remark that usually commenced this substep. Again, it was possible to perceive an apparent functional sequence of activities: ‘gaining external information’ aided ‘determining the product environment’. If this environment could not be determined due to lack of information, the groups were ‘making assumptions on the product environment’. ‘Discussing subissues’ leads to ‘accepting assumptions on subissue solutions’. The activities ‘defining criteria’, ‘clarifying concept working principles’, and ‘discussing subissues’ were carried out most often, followed by ‘making assumptions on the product environment’ and ‘gaining external information’.

Fig. 3 Determine the importance of the criterion (step 2B.1.3)

Fig. 4 Evaluate the alternative–criterion pair (step 2B.2a/b.3)
principles’, ‘discussing subissues’, ‘accepting assumptions on subissue solutions’, ‘making assumptions on the product environment’ and ‘determining the product environment’ all informed the activity ‘raising evidence on restricted performance of concepts’. This activity, in turn, delivered a justification for the outcome of the activity ‘determining or evaluating restricted performances informally’. This outcome, i.e. an informal evaluation, was then mapped on to a formal scale by the activity ‘evaluating restricted performance formally’. ‘Identifying criteria’ had no particular position in the functional sequence. New criteria simply emerged at different times during the evaluations.

6 DISCUSSION OF RESULTS

While measuring and analysing the different activities’ time consumptions and while generating the decision-making process models a number of observations were made. Nine particularly interesting observations will be discussed within the remainder of this paper.

6.1 Observation 1

The groups spent approximately 90 per cent of their time discussing the process approach, deliberating subissues, raising evidence, determining or evaluating performances, controlling the process, defining criteria and gaining external information.

The quantity (number of categories)/value (time consumption) distributions in the three charts of Fig. 1 approximate to a Pareto distribution. A normal Pareto distribution consists of ‘A elements’, contributing 20 per cent in quantity and 70 per cent in value; ‘B elements’, contributing 30 per cent in quantity and 20 per cent in value; and ‘C elements’, contributing 50 per cent in quantity and 10 per cent in value. Because there are 12 categories, according to a normal Pareto distribution the set of ‘A + B elements’ (union of ‘A elements’ and ‘B elements’) would have six categories that together consume 90 per cent of the entire process time. It can be derived from Fig. 1 that this is approximately achieved for each group individually. Interestingly, it is observed that five of the six ‘A + B elements’ are the same categories for each group. These are:

(a) discussing the process approach,
(b) deliberating subissues,
(c) raising evidence,
(d) determining or evaluating performances,
(e) controlling the process.

These five categories seem to represent core activities in the observed decision-making processes, regardless of the chosen approach and the group members’ professional status.

The first observation confirms previous findings [27] with respect to groups. In the previous work, researchers observed individual designers and groups while they were generating conceptual designs, with the aim of understanding design decision-making processes. They found that considerable time was spent on ‘arguments process’ (the categories ‘discussing the process approach’ and ‘controlling the process’ defined in this research) and that deliberating subissues seems to play an important role. For the individual designers the previous work found very little time consumption for ‘arguments product’ (the categories ‘determining or evaluating performances’ and ‘raising evidence’ defined in this research). However, with respect to design groups, Dwarakanath [19] states that ‘alternatives tend to be continuously evaluated . . .’. This research also measured a very high time consumption for activities directly related to the evaluation of alternatives, i.e. ‘determining or evaluating performances’ and ‘raising evidence’.

It can be derived from Fig. 1 that the two formal groups also share the sixth category in their sets of ‘A + B elements’. This category is ‘defining criteria’. However, ‘defining criteria’ did not have much time consumption for the informal group. Instead, the sixth category in the set of ‘A + B elements’ for this group is ‘gaining external information’, which, in turn, did not have much time consumption for the two formal groups. Feldy [29], speaking from years of practical design experience, argues that efforts to address criteria (e.g. ‘defining criteria’) are a sign of effective decision making because all participants are helped to understand the goals of the process. Therefore, the author’s observation that ‘defining criteria’ was a core activity for the present formal decision-making groups, but not for the informal group, suggests that using formal methods may positively influence the groups’ effectiveness.

6.2 Observation 2

The relative times that the groups spent on specific activities were generally more dependent on the process approach (i.e. use or not of formal evaluation methods) rather than on the group members’ level of professional experience.

Even though there is agreement with respect to five core activity categories, being in all sets of ‘A + B elements’, the individual time consumptions of categories are notably different across the three groups’ processes. This can be seen in Fig. 1. Considering the relative times spent by each of the three groups for the activities classified by the categories in their sets of ‘A + B elements’, it was observed that the two formal groups had more similarities to each other than to the informal group. This was despite one of the formal groups being students and the other formal group being professional engineers. Similarities that are particularly distinctive from the informal group refer to the categories:
‘discussing the process approach’, ‘deliberating subissues’, ‘defining criteria’ and ‘gaining external information’. The informal group spent less time on ‘discussing the process approach’, more time on ‘deliberating subissues’ (without actually deliberating more individual subissues than the formal groups), less time on ‘defining criteria’ and more time on ‘gaining external information’. This observation suggests that the effects of using a formal method might override possible effects of different professional status.

As a result, an interesting exception was identified; i.e. with respect to ‘raising evidence’, the formal student group seemed to have more similarity to the informal student group rather than to the formal group of professional engineers. The engineers spent considerably more time in ‘raising evidence’ than either of the student groups. Students and professional engineers have also been compared in an observational study by Smith and Leong [30]. They found that professional engineers paid more attention to detail than the student groups. Conclusively, higher attention to detail may be an underlying reason for the authors’ observation that the professional engineers spent considerably more time on raising evidence than the students.

Further findings of Smith and Leong could not be confirmed by this study. They found that throughout the process the engineers engaged in more management activities than the students. Such behaviour was not noticed by the present authors. Management activities were categorized by the authors as ‘discussing the process approach’ and ‘controlling the process’. Figure 1 shows that the professional engineers did not spend more time on these activities than the students, at least not more than the formal student group. In Smith and Leong’s study no formal methods were apparent. Therefore, it is suggested that it may be the use of a formal method that resulted in extensive management activities in both formal groups.

6.3 Observation 3

The groups spent relatively little time on the aspect of criteria importance.

Expressing the importance of different criteria was the aim of the activity ‘weighting criteria’. The authors observed that all groups had the same tendency to spend very little or no time at all on this activity. This observation is particularly interesting in the light of Ehrlenspiel and Dylla’s [31] finding that ‘successful’ designers thoroughly analysed demands, which includes defining their importance. In this research, the authors found that none of the groups naturally showed this quality. However, it cannot be said to what extent this affected the groups’ effectiveness.

The only group that explicitly considered different levels of importance for the criteria was the formal group of students. This seems to suggest again that the application of formal approaches has an influence on the groups’ decision-making process. Yet, as the authors noticed that for some criteria weights little or no evidence at all was raised (see Fig. 3), they conclude that ‘weighting criteria’ was considered as little more than a formality, even by the formal groups.

6.4 Observation 4

Consistency regarding the application of evaluation criteria was a main difference between the formal groups and the informal group.

The informal group used a comprehensive approach for exploring the criteria and alternatives. As shown in Fig. 2, this approach did not involve much exploration of criteria. The group would pick an alternative and comprehensively evaluate it. ‘Comprehensively’ in this context means with respect to all criteria. This evaluation would then be supported by gathering advantages and disadvantages of the particular alternative. These advantages and disadvantages referred to evaluation criteria. In cases where a particular alternative had been comprehensively evaluated as ‘good’, only those criteria against which the alternative performed advantageous were used to support the evaluation, and vice versa. This meant that the comprehensive evaluation of a particular alternative was based on a subset of the evaluation criteria only. Also, the comprehensive evaluations of other alternatives were not based on the same subset of evaluation criteria, which suggests bias. It was interesting to note that the informal group dedicated more time to evaluating the alternative that was first on the list than for any other alternative. This may have biased their decision as they finally selected this first alternative as their preferred solution (see observation 9).

Ullman et al. [22] as well as Dwarkanath and Wallace [27] also empirically studied informal designers. Ullman et al. [22] found that designers usually focus on only a few criteria to reduce the evaluation’s complexity and that criteria were occasionally forgotten, and therefore not addressed during the alternatives’ evaluations. Dwarkanath and Wallace [27] noticed similar behaviour. Therefore, these two studies offer support for the present authors’ observation.

In this study, the two formal groups used criteria-based approaches involving formal evaluation matrices for exploring the criteria and alternatives. As indicated in Fig. 2, these approaches ensured that all criteria were consistently considered for the evaluation of all alternatives. This consistency may have removed the bias addressed above and may be the explanation for both formal groups having selected the same alternative (which was not the first alternative), independently from each other.

It seems that informal evaluations may lead to inconsistencies regarding the application of evaluation criteria
and to bias. The results of this study suggest that formal methods may help when handling complex evaluations because they appear to improve consistency and reduce bias and the risk of selecting a poor concept.

6.5 Observation 5

In particular, the formal groups showed a tendency towards evaluating the alternatives’ performances on a relative rather than an absolute basis.

As shown in Fig. 2, there were two variants for the formal criteria-based approach: variant a, ‘alternative–criteria’, and variant b, ‘criterion–alternatives’. Picking one alternative and then evaluating it with respect to all criteria before the next alternative is picked, as in variant a, seems to suggest evaluations on an absolute basis. This means the alternatives’ performances may be evaluated on an independent scale and not relative to each other by comparison. This was done consistently, using variant a, by one of the formal groups when they evaluated the alternatives’ performances with respect to those criteria that they defined as being constraints. These were binary evaluations: ‘satisfactory’ or ‘non-satisfactory’. The groups were quite clear in their assessment of what performance was considered to be ‘satisfactory’ and what was considered to be ‘non-satisfactory’. The same group used variant b to evaluate the alternatives with respect to those criteria that they defined as objectives. These were non-binary evaluations: different levels of performance were acknowledged. When groups acknowledged different levels of performance they made comparisons on a qualitative basis rather than evaluating them on an absolute scale.

The other formal group applied variant a with the stated intention of comparing each alternative to a datum only. Even so, they gradually involved more alternatives to compare performances across them. When they did this, they leaned towards variant b, picking a criterion and then used it for the evaluation of all alternatives before the next criterion was picked. The same group then applied variant b with the stated intention to evaluate the alternatives’ performances on an absolute rather than a relative scale. Yet again, despite their stated intention, a strong tendency towards performance comparisons across alternatives was observed. As with the previous formal group, it was not clear what performance the group considered ‘perfect’ and what performance they considered ‘useless’. No utility function was apparent.

Simply comparing alternatives with respect to criteria appears to be straightforward and does not seem to cause problems for the decision makers. This has also been found in other studies [4, 32]. However, such comparisons only relate the alternatives to each other rather than to an overall goal. This means that it is possible to establish which alternative is the best. The question of whether this alternative is sufficiently effective and should be further detailed or whether new, more effective alternatives should be produced cannot be directly answered. According to Feldy [29], decision-making processes often fail when it comes to judging the overall outcome. In fact, when the authors asked the groups whether they would recommend developing more alternatives, or going ahead with the selection, they struggled to find an answer. The only question that they could answer with confidence, based on the information produced by their decision-making method, was: ‘Which is the best alternative?’ Ullman et al. [33] claim that decision-making methods have traditionally only addressed this question rather than giving guidance for any further considerations.

In their empirical studies Ullman et al. [22] observed that when designers compared different alternatives, one was taken as a datum and the others were only compared to the datum rather than to each other. The authors did not observe this behaviour in their study. The informal group considered the alternatives individually and never explicitly used a datum. The formal groups did evaluate by comparing performances, but across a number of alternatives. The authors found that the use of a datum did not seem practical: one group intended and started to evaluate their alternatives by exclusively comparing each one to a datum, but gradually involved more alternatives in these comparisons. The authors see the formal groups’ tendency to compare performances across all alternatives as a result of applying an evaluation matrix, which clearly laid out all alternatives in front of them. In contrast, the participants in the study conducted by Ullman et al. [22] did not use a formal evaluation method.

6.6 Observation 6

The formal groups were consistent in their use of evaluation criteria, but they were not consistent regarding the explicit justification of evaluations.

The formal groups’ evaluations took place in the process step ‘evaluate alternative–criterion pair’ (see Figs 2 and 4). Justifying evidence for the evaluation of an alternative–criterion pair was produced by the activity ‘raising evidence on restricted performance of concepts’. It was observed that this activity was not carried out consistently for all alternative–criterion pairs by either group. This rendered a number of evaluations being unjustified, having a somewhat intuitive character, which makes it difficult to comprehend the evaluations retrospectively.

All groups were given the set of alternative concepts presented on paper. The formal groups chose to draw up evaluation matrices, clearly listing the criteria. It was interesting to note that the informal group was only consistent with respect to addressing all alternatives. The formal groups were only consistent with respect to addressing all alternatives and with respect
to applying all evaluation criteria. The implication is that the groups were only consistent in the aspects of the decision-making process that were documented. With respect to all other aspects, such as delivering justifying evidence, the groups were inconsistent. Such inconsistency can be interpreted as lack of accuracy and having a tendency towards being superficial. According to Ehrlenspiel and Dylla [31], these are clearly characteristics of less successful designers.

Fieldy [29] believes that a benefit of using decision-making methods is the provision of defensible reasons for a decision. The authors agree, but have seen that this could only be partly observed in this study despite some of the groups having used a methodological approach. The authors believe that it is not the use of a method per se but the need for explicit documentation that ensures the consistent provision of defensible reasons for a decision.

6.7 Observation 7

None of the groups produced thorough documentation of their decision-making processes.

This observation is related to the previous one. However, the authors would like to go further than the provision of justifying evidence for evaluations.

During all three groups' decision-making processes a lot of valuable information was gathered. Basically, all activities shown in Figs 3 and 4, apart from 'controlling the process', generated information. The authors believe that this information had the potential to be re-used in later design activities. However, as most of it was not documented by the groups in any form, it seems unlikely that it would have been available for later re-use.

The two groups that used formal evaluation methods produced some records by completing their evaluation matrices. However, these matrices only contained formal evaluation statements (scores), but no underlying justifications, assumptions, deliberated subissues, etc. The group that did not use any formal evaluation method did not produce any records at all. Ehrlenspiel and Lenk [32] made a similar observation in their studies. Court [34] reports on key findings from a number of empirical studies. Among these findings is the observation that there was a clear lack of formal records in design processes. Furthermore, Shah [35] claims this is a shortcoming observable in many companies.

Schluter [3], speaking from many years as head of different design departments in industrial companies, believes the problem with producing records is simply that it takes time. In fact, in one of the authors’ workshops, a group member suggested that they should take some notes, in particular on assumptions made (see activity 'making assumptions on the product environment'). However, nobody made any effort to actually do it.

6.8 Observation 8

The formal groups generated information in a format that could not be directly processed by their formal decision-making methods.

With respect to two aspects of the decision-making processes the authors noticed difficulties with capturing information: performance evaluations and expressions of confidence in evaluations. Regarding performance evaluations, it was a particular format (linguistically vague quantifications) that could not be captured. Confidence in evaluations could not be captured at all due to the particular evaluation matrices used.

Linguistically vague quantifications were statements that included linguistic quantifiers. A wide range of these quantifiers was used. For example, alternative A may be marginally/slightly/a little bit/a fair bit/much/a lot/a great deal better than alternative B. When vague, linguistic quantifications were mapped on to a qualitative scale, all inherent quantitative content was lost. When they were mapped on to a precise quantitative scale, a level of precision was pretended that had not actually been inherent in the original, vague information. Such mappings were made as part of the activities ‘weighting criteria formally’ (see Fig. 3) and ‘evaluating restricted performances formally’ (see Fig. 4).

Confidence in evaluations meant how likely it was that a comparative evaluation was actually true. Such likelihood was never expressed numerically as a per cent, but always by using linguistic terms. Most often, terms were used that could be associated with a likelihood of 100 per cent. For example, alternative A must be/is definitely better than alternative B. Yet, on a few occasions the authors also noticed terms that can be associated with a likelihood of ‘less than 100 per cent’. For example, alternative A may be/could be/is possibly/is probably better than alternative B. Whenever information that contained expressions of confidence was mapped on to the scales used by the groups, these expressions were lost because they could not be modelled at all.

Ullman et al. [33] also describe the occurrence of confidence expression. They describe expressions conveying various degrees of confidence. This was different in the present study because only the exceptional expression conveying a confidence ‘less than 100 per cent’ could be observed. However, there is a difference between the meaning of ‘confidence’ in the studies of Ullman et al. [33] and in this study. In Ullman’s research, ‘confidence’ meant how likely it was that an absolute evaluation was true, whereas in this study ‘confidence’ meant how likely it was that a comparative evaluation was true.

6.9 Observation 9

The two groups who used formal evaluation methods selected the same alternative independently of each other.
and both took considerably more time for their decision-making processes than the informal group.

As mentioned in observation 4, the authors noticed that the informal group spent a much higher proportion of time on evaluating the alternative that was first in the list, than on any other alternative. This first alternative was eventually also chosen, which may suggest bias. In contrast, the two formal groups spent similar proportions of time on the evaluation of each alternative and eventually selected an alternative that was not the first in the list.

Weiss and Hari [4] argue that, in most practical cases, the selection of design concepts is problematic, which is why they recommend the use of ‘methods’. They claim that industrial practitioners commonly select the first proposed concept without thoroughly checking the alternatives. This behaviour imposes the risk of selecting a poor concept, which may have dire consequences on the entire product development programme. Using formal methods seems to help to reduce bias, but it implies a considerably longer decision-making process, which was evident in the authors’ study. SchluÈter [36], arguing from an industrial practitioner’s perspective, believes that the time taken to use methods must be reduced as it slows down the overall design process. Moreover, he sees the methodical evaluation of different alternatives as usually unnecessary because the best alternative is apparent [3]. However, as the present authors’ three groups did not all select the same alternative there was obviously no apparent ‘best’ alternative in this study.

7 CONCLUSIONS

This study differs from previous studies with respect to the level of detail with which the conceptual engineering design decision-making process was studied. Through detailed modelling of the decision-making processes the authors were able to produce empirical evidence leading to a number of very specific observations. After a discussion of these observations the following conclusions are drawn:

1. There were five core activities that dominated all three observed decision-making processes, regardless of the chosen approach (formal or informal) and the group members’ professional status (professional engineers or students). These were: ‘discussing the process approach’, ‘deliberating subissues’, ‘raising evidence’, ‘determining or evaluating performances’ and ‘controlling the process’.

2. The groups were only consistent with respect to those aspects of the decision-making process that were documented. Documentation was generated by the formal groups, but not by the informal group. However, there were still aspects that were not documented by the formal groups and with respect to those aspects inconsistency occurred.

3. The formal groups spent much more effort on understanding the evaluation criteria than the informal group. However, neither of the approaches (formal or informal) lead to a thorough consideration of the relative importance of the evaluation criteria.

4. Activities for determining the ‘utility’ of performances, which would have enabled the decision makers to evaluate the alternatives on an absolute basis, were not used. Instead, the alternatives’ performances were compared with each other and then ranked. The tendency to compare, i.e. to rank rather than rate, was particularly evident when the groups applied formal processes.

5. Evaluations, as well as the confidence held in these evaluations, were often expressed in vague, linguistic terms. However, such terms were never recorded by the groups. This means that the information was lost.

6. Overall, the effects of using a formal method seemed to override the possible effects of professional status. In other words, student groups that used a formal approach in certain aspects of their evaluation and selection process performed better than the more experienced professional engineers when they (the professional engineers) applied informal methods. An exception to this was that the professional engineers spent more effort than the students in gathering evidence to justify their evaluation statements.

ACKNOWLEDGEMENTS

The authors would like to thank the Engineering and Physical Sciences Research Council (EPSRC) for their sponsorship as well as the individuals who participated in this research.

REFERENCES

5 Ullman, D. G. Toward the ideal mechanical engineering design support system, 2000 (submitted); http://www.engr.edu/~ullman.