COMPARISON OF EVALUATION PROCESSES IN DESIGN ACTIVITIES AND CRITIQUING SYSTEMS: A WAY TO IMPROVE DESIGN SUPPORT SYSTEMS

Nathalie Bonnardel

CU-CS-681-93 November 1993

University of Colorado at Boulder
DEPARTMENT OF COMPUTER SCIENCE

Comparison of Evaluation Processes in Design Activities and Critiquing Systems: A Way to Improve Design Support Systems

Nathalie Bonnardel [1,2]

[1]Department of Computer Science and Institute of Cognitive Science
University of Colorado, Boulder, CO 80309-0430
Tel: (303) 492-1102 Fax: (303) 492-2844
nathie@cs.colorado.edu

[2]CNRS (Programme Cognisciences)
France

ABSTRACT

Certain critiquing systems are efficient for evaluating design solutions, but can still be improved to facilitate human-computer interaction. From this perspective, we argue that analyses of human activities in workplaces can inform system design. We illustrate this by analyzing, through observations, interviews, and experiments, the evaluation process used by designers and comparing its characteristics with those of critiquing mechanisms embedded in various systems. The results of this comparison allow us to highlight similarities and differences concerning the objects undergoing evaluation; the judgments, procedures, and knowledge involved in the evaluation; and the functions fulfilled by the evaluation process. Recommendations to improve design support systems are derived from this comparison.

KEYWORDS: design activity, evaluation process, support systems, critics.

1. INTRODUCTION

Assisting designers in the evaluation of design solutions seems particularly useful because they have to evaluate various options throughout their activities, and they may encounter difficulties in doing so due to the "open-ended" and "ill-defined" character of the problems they must solve. Design problems are considered *open-ended* because they allow a diversity of possible solutions (Fustier, 1989) among which designers have to choose. Thus, designers not only have to assess single solutions developed during problem-solving, but also to evaluate alternative solutions. Moreover, because of the *ill-defined* character of these problems (Reitman, 1965; Eastman, 1969; Simon, 1973), designers have only poorly defined goals, and they therefore have to define criteria for evaluating design solutions. However, it may be difficult for novice designers in a domain to define the relevant criteria, especially if they have not constructed some personalized criteria, identified by Guindon (1990) as a characteristic of designers.

Certain support systems, termed *"critiquing systems,"* assist users in evaluating their own solutions (see Fischer, Lemke, Mastaglio, and Morch, 1991). These systems evaluate or "critique" the user's solution, highlight its disadvantages and, sometimes, its advantages. Such systems have been developed to support the design activity but, although promising and efficient, they can be improved to support more parts of the evaluation process and facilitate human-computer interaction.

With this objective, we argue that the identification of improvements to systems may result not only from user tests, as is frequently the case, but also from *analyses of human activities related to the systems' functionalities*. The hypothesis underlying such analyses is that human activity may serve as a model for defining certain characteristics of support systems. More precisely, we argue that the analysis of the evaluation process used by designers can inform system design about the critiquing mechanisms that are well-grounded and those that still have to be embedded in design support systems. From this perspective, we analyzed and compared the evaluation process used by designers and the critiquing mechanisms embedded in various support systems with regard to five main characteristics:

- the objects undergoing evaluation,
- the nature of the evaluative judgments,
- the procedures used for evaluating solutions,
- the knowledge elements taken into account for evaluating solutions,
- the role of evaluation in the progress of the reasoning.

Before presenting the results of this comparison, we describe our domain of study of design activities together with the data collection methods we employed. Then, we highlight the similarities and differences of the evaluation processes used by designers and critiquing mechanisms. Based on these results, we finally suggest certain improvements that could be carried out on design critiquing systems.

2. ANALYSIS OF THE EVALUATION PROCESS IN DESIGN ACTIVITIES

The identification of characteristics of the evaluation process used by designers was based on several field studies that took place in an industrial design domain: the *design of aerospace products* (such as reflectors and structures intended for integration into the ARIANE rocket). For these studies, we employed various data collection methods.

2.1 Domain of Study

The design of aerospace products constituted an interesting domain of study because it allowed us to analyze the effect of *natural factors* on different aspects of the evaluation process.

A first factor was linked to the nature of the problems and solutions designers dealt with. Certain problems encountered by designers presented common characteristics with problems previously solved; we describe such problems as "traditional." Other problems were, on the other hand, entirely different from previous problems; we qualify these as "novel." The design solutions generated for these two classes of problems were initially situated at different levels of abstraction. The solution to traditional problems was usually inspired by solutions previously generated for similar problems, and was situated at a "physical level" (certain of its technical features were already defined at the beginning of the problem-solving). On the contrary, an innovative solution had to be elaborated for the novel problems, and this solution was generally situated at a "conceptual level" (its technical features were not defined at the beginning of the problem-solving).

In addition to the "type of solutions," two other factors were related to the designers themselves: their "type of expertise" and their "level of expertise." Indeed, certain designers were engineers and in charge of calculation tasks, whereas other designers were technicians and in charge of tasks of drawing and realization of plans. The designers first reflected together on the type of solution to develop for solving a given design problem, and then performed their specific tasks to define more precisely the characteristics of the solution adopted. The designers could also be distinguished according to their "level of expertise" in certain design domains: some of them were considered by their peers to be expert or experienced in a specific design domain if they had been working in this domain for tens of years; other designers were considered inexperienced when they had been working in a design domain for only a couple of years or less.

2.2 Methods of Data Collection

Our purpose was to study real situations of design activity, so we faced a conflict linked to the choice of data collection methods: observations in the workplace, which can be accused of yielding only primarily anecdotal data, or laboratory studies, which can provide results that have little significance for real situations (see Rieman, 1993). Moreover, it seems particularly difficult to construct experiments to study the design activity because even when several designers have to deal with the same problem, they construct different representations of it, and in fact solve different problems (see Bisseret, Figéac-Letang, Falzon, 1988). Thus, we decided to resort to several data collection methods – interviews, observations, and experiments – in order to collate the results obtained through these different methods.

We first performed *interviews and observations* in real time, combined with comments made by the expert, in order to determine the role of the evaluation process when it is applied to design solutions. The interviews and observations were conducted with two experts who had been working in the company for about 30 years. During the interviews[1], the first expert described the design of a structure (called SYLDA – SYstem of Launching Double Ariane) that was intended to carry satellites inside the ARIANE rocket. The purpose of the analysis of the gathered data was to identify regularities in the intervention of the evaluation process (for a full presentation of this analysis, see Bonnardel, 1992a). As the method of interviews is subject to a risk of "rationalization" of their own activity by the experts – they may describe their activity as more structured that it is in reality (see Visser, 1990) – we collated the results obtained through the interviews and results obtained through observations in real time. These observations were performed on the activity of a second expert (the first one being retired), all along the design of a prototype of a new type of reflector (a deployable reflector). Because of potential difficulties in interpreting observations, the designer was prompted to make comments on his activity (eventually, by asking him some questions). The analysis of these data allowed us to "reconstruct" the different steps of the design activity (see Bonnardel, 1992a).

In order to prompt designers to verbalize some of the reasoning behind the evaluation of design solutions and to express the knowledge elements they took into account for evaluating design solutions, we constructed two experimental situations.

The first one was based on a scenario of "project transfer" (see Bonnardel, in press) established among three designers: an engineer (E) and a drafter (D), both in charge of the design of a structure (called ARTEP – ARiane Technology Experiment Project) intended to carry equipment and small satellites inside the ARIANE rocket; and another designer (R, for "receiver" of information), expert in the design of similar products. During the situation of design project transfer, designers E and D had to transmit to designer R

---

[1] The interviews were conducted by Visser and Falzon as a part of a previous contract between INRIA and AÉROSPATIALE (see Visser & Falzon, 1988).

the information elements needed to continue the work. Because they were placed in a *collective situation of design*, the designers spontaneously verbalized a part of their reasoning in order to defend their proposals and show that they were well-founded.

In the second experimental situation, 14 designers were presented with two real design problems, each associated with a set of solutions. These designers were asked to *evaluate the proposed solutions* and select one of them while mentioning the "criteria" they were taking into account. One of the problems was novel and belonged to the reflector domain: "design a reconfigurable reflector" (i.e., a reflector for which the profiles of the reflecting structure had to be modifiable in orbit). The other problem was traditional and related to the structures intended for integration in the ARIANE rocket: "design an ACU" (i.e., a payload adaptor that serves as an interface between the launcher and a structure intended to carry satellites). Both the problems and the solutions had been defined by two department heads (specialized in one of the domains) who spontaneously produced conceptual solutions for the novel problem and physical solutions for the traditional problem. The designers who took part in this experiment could be distinguished according to their type and level of expertise in the design domains involved in the experiment.

The verbalizations of the designers who participated in these two experimental situations were recorded and transcribed, and these data were completed by notes of observations from the experimenter (for a full presentation of the analysis of the data from these situations, see Bonnardel, 1992a).

2.3 Validity of the Methods Used: Discussion

We already pointed out the main risk of analyses based on interviews – the rationalization by the experts of their own activity – but the results obtained through this method are collated with results obtained through observations in real time, which were performed in a "natural" design situation. Thus, our discussion is more focused on the situations of design project transfer and evaluation of sets of solutions, which were "artificially" constructed. These two situations present common characteristics with “natural” situations of design: the design problems considered in these situations were real, the subjects were professional designers, and the situations took place in field conditions. Moreover, the situation of project transfer was plausible because "transfers of orders" can occur in the field of aerospace design, and group collaboration is effective when designers encounter difficulties in their activity. The main interest of this situation was to prompt the designers to express some of their reasoning, without any intervention from an experimenter (when such interventions occur, designers know that they are addressing a person outside of the field, and may restrict their verbalization to information of moderate interest to other designers). The situation of evaluation of sets of solutions presented two main differences with natural situations: (1) the subjects' task was limited to the evaluation and selection of already defined solutions, and did not form part of a complete design activity; (2) each subject was presented with a set of possible solutions, whereas in a natural situation a designer generally focuses on only one solution (even if he or she has evoked several).

Nonetheless, it may happen that several solutions are considered at the same design stage when several designers cooperate in order to solve the same problem and they each suggest a different solution.

3. THE EVALUATION PROCESSES IN DESIGN ACTIVITIES AND CRITIQUING SYSTEMS

Our comparison of the evaluation processes used by designers and critiquing systems is focused on the main characteristics of these processes: the objects undergoing evaluation, the judgments, procedures, and knowledge elements involved in the evaluation, as well as the role of evaluation in design activities.

3.1 Objects Undergoing Evaluation

A prerequisite to providing computational critics to designers is to be sure that such critics will really support them (and not disturb them) in their activity. In order to show the well-groundedness of this fact, we first focused our attention on the intervention of judgments in design activities. Indeed, we consider that if designers assist each other in evaluating solutions, by providing their own judgments to other designers, critiquing systems can also be really to support designers in evaluating their solutions.

Our main argument results thus from the analysis of the activity of designers engaged in collective situations of design, such as the one developed during the dialogue of project transfer. Indeed, we observed that designers not only evaluate the solutions they themselves generate but also solutions envisioned by other designers, to extend or modulate the judgments expressed by the latter (see Figure 1). This confirms the well-groundedness of critics in design support systems.

- "From the point of view of cost, the metallic is the best"
- "Monolithic, in my opinion, would be the best ... You can make holes anywhere, whereas with the sandwich-carbon, you have to place inserts, tomake the structure more rigid there ..."
- "Yes, but from the point of view of development costs [it is different]: we have never developed monolithic structures whereas they [the clients] want something that is not expensive"

Figure 1. Example of solution evaluation by designers.

Several support systems developed in various domains embed computational critics (see Table 1). Most of the systems we consider are related to the domain of design activities, but we also mention certain critiquing systems relating to medical domains (in which certain of the first critiquing systems have been developed).

Table 1. Systems being compared with human activities.

Domains	*Systems*	*Solutions*
Architectural design (of kitchens)	• JANUS (Fischer, McCall & Morch, 1989) • KID – Knowing In Design – (Nakakoji, 1993) • AGENTSHEETS[2] (Repenning & Citrin, 1993)	kitchen floor plans kitchen floor plans kitchen floor plans
Design of programs • program frameworks (for user interfaces) • programs in Lisp	• FRAMER (Lemke & Fischer, 1990) • LISP-CRITIC (Fischer, 1987; Mastaglio, 1990)	program frameworks Lisp codes
Design of voice dialogs	• VDDE – Voice Dialog Design Environment – (Repenning & Sumner, 1992)	phone-based interfaces
Medicine • anesthesiology • oncology	• ATTENDING (Miller, 1984) • ONCOCIN (Langlotz & Shortliffe, 1983)	therapeutic plans therapeutic plans

According to our hypothesis, critiquing systems should reflect specificities of the reasoning that designers develop while using the evaluation process. However, we identified two *differences* in the objects evaluated by designers and critiquing systems:

1. The designers can evaluate design solutions situated at various *levels of abstraction* (i.e., either conceptual or more or less physical), whereas the functioning of the critiquing systems we analyzed is limited to the evaluation of physical solutions that present concrete characteristics. Indeed, the users of critiquing systems have to specify the characteristics of their solutions, for example, by selecting options in menus, which implies a certain level of precision (see the ATTENDING system), or by directly constructing their solutions on the system (see the KID and VDDE systems).

2. The designers, contrary to critiquing systems, evaluate both *design solutions* and the *criteria and constraints* they take into account during the evaluation and/or the generation of design solutions. As we will focus on the criteria and constraints designers refer to for evaluating solutions, we group them under the generic term "evaluative referents." The assessment of evaluative referents allows the designers to determine an order of priority among the referents they take into account (e.g., "first, there is a criterion of cost, then of mass"). The results of such an evaluation were expressed explicitly during the experimental situations, or manifested implicitly during the observations (through the choice of solutions that satisfied

---

2 AGENTSHEETS has a particular status: when applied to the domain of kitchen floor plan design, it is based on evaluative knowledge similar to the one taken into account in JANUS and KID, but it cannot really be considered as a critiquing system because this knowledge is used for defining constraints that forbid certain choices during the construction of design solutions (and not for providing messages to the users).

certain evaluative referents but not other ones). Despite the fact that critiquing systems do not support users in assessing evaluative referents, some of them, especially the KID system, take into account the criteria and constraints the users specify.

3.2 Nature of the Evaluative Judgments

3.2.1 Judgments About Design Solutions

We identified two main *similarities* in designers' judgments and critics when they are applied to design solutions:

1. Both designers and critics can validate and invalidate features of solutions (see Figures 1 and 2).
2. In spite of this fact, both designers and critics highlight more critical or invalid features than positive or valid features.

For example, when designers have to deal with several concurrent solutions, they try to select one of them and, in doing so, take into account primarily the solutions' drawbacks (advantages are most frequently considered only when this strategy is deemed insufficient for selecting a solution). Such an asymmetry appears also in certain critiquing systems, at the level of the intervention strategies of the critics. In the JANUS and KID systems, the critics that intervene to invalidate features of solutions are based on an active intervention strategy: they fire as soon as one or several feature(s) of the current design solution are incompatible with certain evaluative referents. On the contrary, the critics that validate features of solutions are based on a passive intervention strategy: they appear on the screen only if the user inquires (e.g., by selecting the option "praise all" in a menu). Such an asymmetry between judgments of validation and invalidation may have a functional explanation: negative judgments lead to decisive changes in the progress of the activity (such as the modification or rejection of the solution), whereas positive judgments only provide confirmation of the fact that the chosen options were appropriated to the current step of problem-solving.

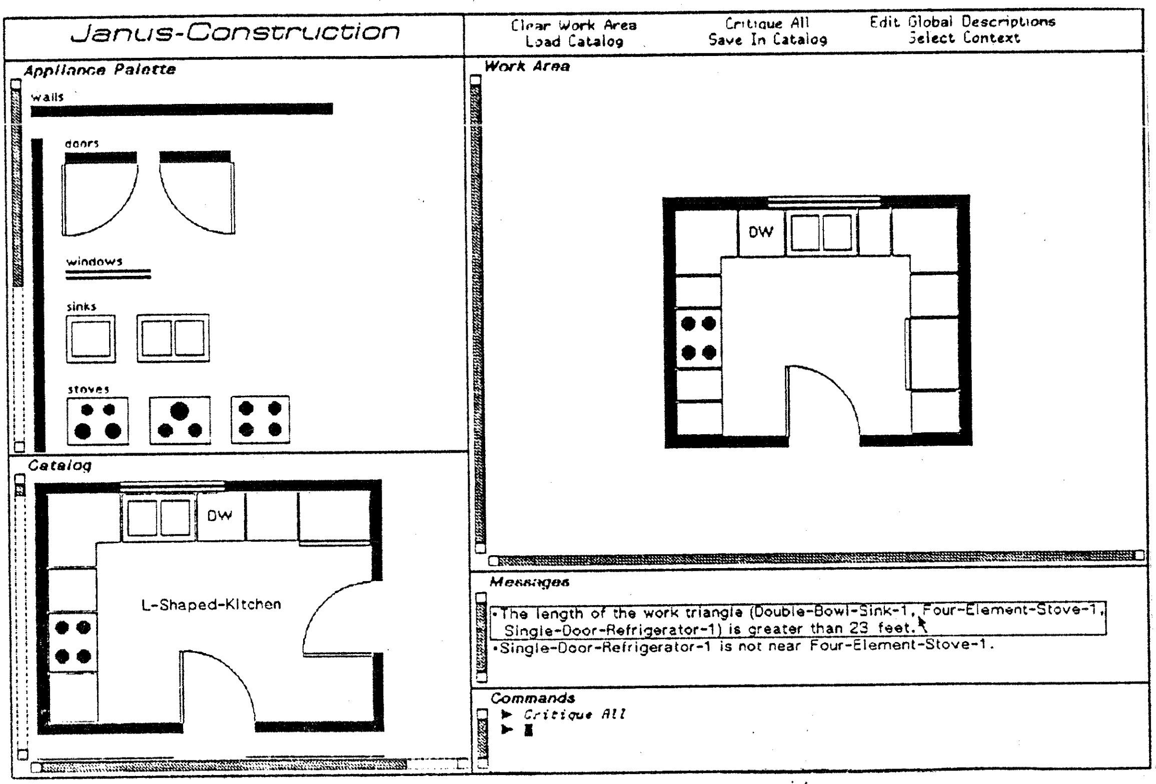

Figure 2: Example of critic provided by the KID system.

The expression of the goodness of a solution is explicit when a fact is stated and attached to a desirable value (e.g., "this structure is much too heavy"), whereas the judgment of value remains implicit when only a fact is stated (e.g., "this structure is heavy"). Both types of expression of judgments can be observed in natural situations of design. However, the statement of only a fact requires that the users infer the desirable value that is not explicitly expressed. Such an inference can be easy for experienced designers who have already dealt with similar judgments but requires access to more information for inexperienced designers.

The objective of critiquing systems is to support various users in evaluating solutions, which includes both experienced and inexperienced designers in a domain. Thus, we argue that the critics provided to the users should first explicitly specify the goodness of the solutions' features in order to fit the needs of inexperienced designers. In addition, the expression of the goodness value highlights more the advantages or disadvantages of solutions and, therefore, may increase the designers' motivation to seek better design solutions. Nevertheless, the users should also have the opportunity to adapt certain critiquing mechanisms themselves, in order to allow experienced designers to be presented only with stated facts if they wish it (e.g., when they already know the goodness value attached to these facts).

3.2.2 Judgments About Criteria and Constraints

As we previously pointed out, the judgments about the evaluative referents allow the designers to determine an order of priority among the referents. The analysis of the interviews and the dialogue of project transfer enable us to characterize the nature of the judgments that underlie such an order of priority. This order first depends on the *status* assigned by designers to evaluative referents. Indeed, the designers distinguish between evaluative referents that must be respected by the solutions – which they call "destructive" – and those that are desirable, though not obligatory, to be respected by the solutions – called "of choice." Destructive referents have thus to be considered in priority by designers. Then, the designers evaluate the relative importance of evaluative referents of choice, which allows them to complete the definition of an order of priority among the evaluative referents.

Usually, evaluative referents resulting from clients' specifications are considered destructive whereas those that depend on designers' decisions are considered of choice. Nonetheless, certain discussions among designers may occur about the status of evaluative referents they themselves have defined (see Figure 3).

- "It [the structure] cannot exceed 40 kg."
- "But we were the ones that defined it."
- "Yes, but now that it is defined, we have to keep it."

Figure 3. Discussion about the status of a constraint.

Judgments about both the importance and the status of evaluative referents can be taken into account in certain critiquing systems, although these systems do not support users in assessing evaluative referents. For example, the KID system allows the users both to specify their own criteria and constraints and to assign a weight to them. Certain judgments can also have been expressed before the intervention of users and built into the critiquing system. Thus, the FRAMER system includes two types of critics that have a distinguishing status (see Fischer et al., 1991):
- "mandatory critics," which must be satisfied for program frameworks to function properly;
- "optional critics," which inform the user of issues that could be dealt with in another way.

3.3 Procedures for Evaluating Solutions

Certain similarities between the procedures used by the designers and those used by critiquing systems for evaluating solutions can be pointed out. The critiquing mechanisms of these systems appear similar to two evaluative procedures that we identified through the observations and the experimental situations: the "analytical" and "comparative" procedures. We also identified a third evaluative procedure (called "global") that is used by experienced designers to determine immediately (without conducting a deep analysis) if a solution can fit a problem; as it is based on their own experience, this procedure may be difficult to embed in a system.

The *analytical* procedure of evaluation is used by designers when they have to deal with a single solution: it allows them to identify disadvantages and advantages of solutions with regard to evaluative referents. For example, designers can assess that a solution is "too heavy" (with regard to a criterion of lightness) or "not rigid enough" (with regard to a criterion of rigidity). Similarly, certain systems evaluate the users' solutions with regard to evaluative referents. For example, the JANUS and KID systems, as well as the VDDE system, identify undesirable features of design solutions with regard to various evaluative referents (either defined from general design principles or from specifications mentioned by the users).

The *comparative* procedure of evaluation (that may also be called differential) is used by designers when they have to deal with several alternative solutions: it is also based on evaluative referents but allows moreover a "term to term" comparison of the alternative solutions. For example, the designers consider the criteria of cost and ease of realization for assessing two solutions: "The advantage of such a solution is that, even if it may be expensive, when it's finished, it's finished; whereas with this technique, you realize a frame and then you have to do the machine-tooling again and again." Certain critiquing systems embed such an evaluative procedure when they identify the solutions' disadvantages or advantages by comparing the user's solution with one or several solutions that the system itself prioritizes. For example, the ONCOCIN system uses problem data specified by the user (e.g., information about the health state of the patient) to generate a solution (a therapeutic plan), which is not presented to the user but used for performing a comparison with the solution the user suggests. Another example is provided by the ATTENDING system, which compares the risks of the user's solution with those of alternative solutions selected by the system.

The functioning of critiquing systems is thus basically different according to the procedure of evaluation that is adopted: only certain features of the solutions are taken into account when an analytical procedure is used, whereas all the differences between the user's solution and the system's solution(s) are taken into account when a comparative procedure is used. Each of these procedures presents certain weaknesses (see Fischer et al., 1991): the analytical one is submitted to risks of omission of certain drawbacks of the solutions, and the comparative (or differential) procedure cannot be applied if the user's solution and the system's solution(s) are fundamentally different (the system can only affirm that the solution it generates itself is appropriate for solving the problem). However, the choice of one or the other of these procedures may be dependent on the type of problem to be solved (e.g., it is difficult for systems to generate solutions to design problems) and the step of problem-solving (i.e., facing one single solution, or several alternative solutions one of which has to be selected).

3.4 Evaluative Knowledge Elements

Similarities can also be pointed out concerning the number and the nature of the evaluative referents taken into account by the designers and those taken into account by critiquing systems (when they embed an analytical procedure).

3.4.1 Number of Evaluative Referents

The analysis of the data gathered during the situation of evaluation of sets of solutions allowed us to determine the number of evaluative referents the designers were considering. It appeared that they took into account a relatively small number of evaluative referents: an average of 8 for the traditional problem, and an average of 6 for the novel problem (out of a total number, respectively, of 18 and 19 different types of evaluative referents for the traditional and the novel problems). This result therefore supports the idea of the use by designers of a *limited set of evaluative referents* (Guindon, 1990), and suggests that even a small number of evaluative referents could underlie the definition of critics that are important in a given domain.

Another result was related to the designers' level of expertise in the domain: in the case of the novel problem (but not of the traditional one), a greater number of evaluative referents were expressed by the designers who were experienced in the domain of the problem than by those who were inexperienced (average of 8.4 evaluative referents vs 4.75). This result may be explained by the fact that even when designers use the analytical procedure of evaluation, they may refer to a *solution of reference*, which could consist of the standard solution used in the domain and would be characterized by several evaluative referents. These referents would be those the designers take into account for evaluating solutions that present similarities with the standard solution. Therefore, when the problem is traditional, even designers inexperienced in the domain have already seen products developed consecutively to the adoption of the standard design solution and have constructed a mental representation of this solution of reference. On the contrary, when the problem is novel, the designers who are inexperienced cannot refer to such a standard solution, and do not have enough knowledge in the domain to construct a solution of reference by themselves. Such an interpretation thus decreases the distinction between the analytical and comparative procedures.

3.4.2 Nature of the Evaluative Referents

The gathered data also enabled us to distinguish between two types of evaluative referents:

- evaluative referents that are *directly applicable* to design solutions because they are related to concrete characteristics of these solutions; e.g., "limit the number of actuators" (the actuators being small motors used, in this case, for modifying the profile of the surface of the reflector);

- evaluative referents that need to be *further defined* before the designers can apply them on solutions. For example, the "interest" of a solution can be assessed only with regard to more specific criteria (such as its cost, versatility, etc.), and these criteria can in turn be translated into more concrete evaluative referents (such as "use already known technologies" or "limit the use of machine-tooling" in order to limit the cost).

We call *constraints* the evaluative referents that can be directly applicable to evaluate design solutions because they both limit and orient the choices of the designers (in accordance to the definition proposed by Mostow, 1985), and *criteria* the evaluative referents that need to be further defined before being applicable to evaluate design solutions. An argument in favor of this distinction is that the criteria seem to be preferentially used by designers during the evaluation of solution, whereas the constraints seem to be considered also during the generation of solutions (see Bonnardel, 1992a). Thus, during the first experiment, it appeared that only 8 constraints from a total of 18 evaluative referents (or 66%) were mentioned by designers while evaluating the solutions proposed to solve the traditional problem, and only 7 constraints from a total of 19 evaluative referents (or 37%) while evaluating the solutions proposed to solve the novel problem.

However, certain associations of criteria and constraints show the existence of relationships between them, such as in this comment from a designer: "for issues of <u>cost</u> and <u>reliability</u>, it is always worth *limiting the number of pieces*" (we <u>underscore</u> the criteria and show in *italic* the constraint). Such observations, together with the analysis of various data gathered during the interviews, allowed us to distinguish among three types of constraints (see Bonnardel, 1989):

- constraints that result from an operationalization of criteria (as shown in the above example), which we call "*constructed constraints*" because they are derived from criteria acquired by the designers as their experience grows. For example, the criterion of "reliability" can be operationalized by the constraint "limit the risks."
- constraints that result from an analysis of the problem's specifications (and a selection among them), called "*prescribed constraints*." For example, a higher department (or the client directly) can specify that "the height of the structure should not exceed 2.8 meters".
- constraints that result from an analysis of the consequences of the current state of problem-solving and/or consequences of constraints that have already been defined, called "*inferred constraints*[3]." For example, the constraint "homogenize the mesh configuration" is inferred from the current state of problem-solving, i.e., the adoption of a "mesh" solution for the novel problem; and the constraint "use technologies already developed" is inferred from the constructed constraint "limit the risks".

---

3 We can also call these constraints "deduced constraints" but the term "deduced" should not have the connotation of "logical deduction" (which would suggest that the conclusion reached by the designers is always exact), as we do not wish to discuss the exactness of the designers' steps of reasoning.

Although such a distinction between the evaluative referents embedded in the KID system has not been established, the different types of evaluative referents we have defined correspond to different critics that can be provided by this system (see Fischer et al., 1993). Certain critics, called "specific critics," are related to users' specifications, and others, called "generic critics," are related to general design principles and thus to criteria. In addition, the activation of such critics depends on features of the current state of problem-solving.

3.5 The Role of Evaluation in the Progress of the Reasoning

The various data we analyzed showed that the evaluation process plays a decisive role in the design activity (see Bonnardel, 1992b): this process allows (1) the determination of the design step that follows the one of solution evaluation, (2) the selection of a solution among several alternatives, and (3) the definition of an order of priority among criteria and constraints. In accordance with the hypothesis we previously expressed – i.e., the fact that human activity can serve as a model for specifying characteristics of support systems – design critiquing systems should support designers in these three stages of the design activity.

3.5.1 Determination of the Following Step of Design

The various data we gathered showed that, according to the results of the evaluation of the current solution, the designers develop or modify this solution (see Bonnardel, 1989). The evaluation process thus contributes to the control of the activity and, more precisely, guides the progress of the designers' reasoning. In accordance with this result, the role of evaluation in design critiquing systems should not be limited to judgments of validation about solutions (e.g., "local" judgments) but should also support the user in determining the following step of design.

Certain design critiquing systems play only a local role. Thus, the LISP-CRITIC system is efficient for identifying code errors but its critic messages remain at this "surface" level. Other systems facilitate more the progress of the user's reasoning. Thus, the KID system takes into account the objectives and preferences of the user in order to suggest certain solution elements. For example, if the users specify that they frequently entertain, the system suggests embedding supplementary elements (e.g., a double sink and a dishwasher) in the kitchen. This system can also provide a more indirect contribution to the progress of the reasoning through examples of completed kitchen floor plans (see Figure 2, window "catalog") that can be partially reused or serve as basis for analogical reasoning. In KID, the suggestions of solution elements as well as the design examples are provided to the users only when they inquire. A more active contribution is provided by the AGENTSHEETS system, which directly modifies the design solutions when they violate design constraints. This system can also establish compromises between conflictual constraints and, subsequently, modify the location of design elements. The functioning of this system probably increases the rapidity of problem-solving, but we can ask ourselves whether the users appreciate the direct modification of their solutions and if it does not limit their learning. The type of contribution to the progress of the reasoning could, in fact, depend both on the context in which the system is integrated and on the objective of the use of the system. For example, a system that forbids certain design options or modifies invalidated solutions may be desirable in situations with risks and constraints of rapidity, whereas a system that prompts users to reflect about their own solutions may be more desirable in situations of learning.

3.5.2 Selection of Design Solutions

The selection of one solution among several alternatives can be observed both in the activity of designers and in certain support systems. In *design activities*, designers can generate several alternative solutions both in collective and individual situations of design, and they have to select one of these solutions because their development in parallel would be too expensive (on cognitive and financial levels).

1. In *collective situations* of design problem-solving, each designer can suggest a different solution, as we observed in the situation of project transfer. In this case, the designers defend their own proposals by highlighting the advantages of their own solution and the disadvantages of alternative solutions (see Figure 1).

2. Designers in *individual situations* can also generate several solutions (most frequently, two solutions). The analysis of the observations and the situation of evaluation and selection of sets of solutions shows that designers in individual situations focus on solutions' disadvantages to select the solution that has the lowest number of critical points and/or the least important critical points. Critical points result from the incompatibility of certain solutions' features with evaluative referents, so the importance of critical points depends in fact on judgments about the status and the importance of the evaluative referents that are not satisfied by the solutions.

In certain cases, the selection of a solution seems to be easy. This is the case, for example, when the evaluative referents not satisfied by one of the solutions are judged to be either lower in number or less important than those not satisfied by the other solution(s). In the other cases, the selection of a solution can be complex. For instance, problems of choice appear when alternative solutions present the same number of critical points and/or when their critical points have equal importance. For example, during the design of the prototype of a deployable reflector (studied through observations and comments made by the expert), the designer envisioned two alternative solutions for supporting the reflecting surface: "using wires" or "using foils." In order to select one of them, he evaluated these solutions and compared their critical points (he thus found the "wire" solution not reliable because of risks of tangling wires, and the "foils" solution difficult to realize because it had never been used in the field of reflectors). The criteria "reliability" and "ease of realization" being of similar importance, the designer adopted a new "point of view" to evaluate the solutions: he took into account a criterion of '"spatializability," which allowed him to determine that the "wire" solution, as opposed to the "foil" solution, was not spatializable and, thus, incompatible with another evaluative referent. The "foil" solution therefore appeared preferable to the "wire" solution and was selected.

Problems of choice may also appear when the solutions are associated with factors of a different nature; for example, when one of the alternative solutions presents several critical points of moderate importance, and another one a single but more important critical point. The way to choose between such solutions remains to be determined, but it may depend on both the state of problem-solving and characteristics specific to the designers (such as the experience they have acquired during previous design situations).

Similar selections of solutions may occur in certain *support systems*, and may be based only on the system's decisions or depend on the user's preferences. For example, the selection of a solution in the ATTENDING system is based on a comparison of the risks associated with the solution suggested by the user and those associated with the alternative solutions considered by the system. The solutions associated with the most important risks are thus progressively eliminated (except the user's solution, which is kept in order to provide explanations to the user) independently of the users' preferences. The users decide only whether to adopt the solution suggested by the system. The "multicriteria decision" system, developed by Boissier and Al-Hajjar (1990) in the domain of the design of infrastructure of buildings, is, on the contrary, based on the user's preferences: three typical solutions (being more or less compatible with certain criteria) are presented to the user who has to express his or her opinions on them. The system takes into account these opinions for identifying an "aggregation operation" that is in accordance with the user's preferences. Then, the system uses it to evaluate the possible solutions and eliminate those that do not fit the user's preferences. Thus, it appears that, for both designers and the systems we analyzed, the selection of solutions results from a *progressive elimination of solutions presenting disadvantages*.

3.5.3 Definition of an Order of Priority Among Criteria and Constraints

The analysis of the design activities tends to show that designers may define an order of priority among criteria and constraints at the main stages of their activity:

- during the generation and development of solutions, first in order to consider the most important constraint(s) for delimiting the set of possible solution elements, then to determine the next most important constraint(s) for specifying once again the researched solution elements and further defining the set of possible solution elements;
- during the evaluation of solutions in order to define which evaluative referents to consider in succession to evaluate the solution(s) and, thus, respect a principle of economy: if the considered solution does not satisfy the predominant referents, it is directly rejected or modified, and it is not necessary to evaluate it according to other evaluative referents;

- during the selection process in order to identify and select the solution that presents the least important critical points and, in certain cases, that satisfies the predominant evaluative referents.

As previously mentioned, the designers seem to take into account first "destructive" referents, then referents "of choice" according to a sliding scale of predominance. However, the order of priority may depend on numerous factors, linked not only to the nature of the problem and the state of problem-solving, but also to the experience and "know-how" specific to each designer.

Certain systems, such as KID, can take into account priorities specified by the users about the criteria and constraints (see paragraph 3.2.2), but the specific advantage of the "multicriteria decision" system of Boissier and Al-Hajjar (1990) is that it does not ask the users to assign explicitly a weight to the evaluative referents but infers this weight from the users' opinions about typical solutions. This functioning seems to us particularly desirable because, as we observed in the aerospace domain, designers can encounter difficulties in expressing weights (especially if they consist of numerical values), whereas they more easily prioritize evaluative referents.

4. CONTRIBUTION OF THIS ANALYSIS

The contribution of the comparison of the evaluation processes used by designers and the critiquing mechanisms embedded in various systems is twofold:

- results of this comparison reveal a general framework of evaluative knowledge useful for analyzing design activities;
- these results also inform system design by indicating the components of critiquing systems that are well-grounded and those that could still be developed.

4.1 A Framework for Analyzing Design Activities

Certain characteristics of the evaluation process used by designers of aerospace products appeared to have much in common with characteristics of critiquing mechanisms embedded in design support systems related to very different domains, such as kitchen design, voice dialog design, and design of program frameworks for user interfaces. Thus, we argue that several results obtained in the domain of aerospace product design can be generalized to other design domains, especially the distinctions between criteria and different types of constraints, together with the distinction linked to the status ("destructive" and "of choice") of these referents. Indeed, as design problems are ill-defined, designers have to define constraints for "framing" their problems and evaluative referents for assessing solutions, as well as to determine orders of priority to manage constraints and criteria that may be both numerous and conflicting. This characterization of evaluative referents could thus form part of a *general framework for evaluative* *knowledge* that could be used by cognitive psychologists or knowledge engineers for exploring various design activities, and for analyzing the knowledge that is elicited by experts. To prompt elicitation of evaluative knowledge, experts could be asked to point out the advantages and disadvantages of different types of solutions. This would allow the collection of evaluative referents that characterize solutions in the domain, in order to reuse them when defining critics to embed in support systems.

4.2 Information for System Designers

The different results that have been presented illustrate the fact that *analyses of human activities can inform system design*. Indeed, these results point out which components of critiquing systems are well-grounded, based on their similarities with characteristics of the designers' reasoning. These analyses also indicate functionalities that could still be added to design critiquing systems.

Several components of the critiquing mechanisms appear *well-grounded* (see Table 2):
- the evaluation of physical solutions in order to express mainly judgments of invalidation;
- the use of either analytical or comparative (differential) procedures, conducted with regard to evaluative referents or solutions of reference;
- taking into account critics that reflect the users' specifications (prescribed constraints) or general design principles (criteria and constructed constraints), and that depend on the state of problem-solving (similarly to the inferred constraints);
- the tendancy of certain systems not only to "locally" assess solutions but also to guide designers in the progress of their reasoning.

These components are specifically related to design solutions, whereas judgments expressed by designers on criteria and constraints also play an important role in the design activity. This suggests that design critiquing systems should assist users not only in evaluating design solutions but also in assessing criteria and constraints, especially to guide novice designers in the choice of relevant evaluative referents and in the process of prioritizing them. Moreover, when the users have preferences about the evaluative referents that have to be prioritized, the systems should be able to define these preferences both from the users' specifications – when the users are able to express explicitly their preferences – (as the KID system does) and from the users' opinions about typical solutions – when the users cannot explicitly express their preferences – (as the "multicriteria decision" system does).

Table 2: Main results of the comparison of the evaluation process and critiquing mechanisms

	Designers	*Critiquing Systems*	*Specific Systems*
Objects undergoing evaluation	• solutions: conceptual & physical • criteria & constraints	• solutions: only physical -	found in all systems analyzed
Evaluative judgments on solutions: *on evaluative referents:*	• validation & invalidation • focus on invalidated aspects • discrete & gradual • status & predominance	• validation & invalidation • focus on invalidated aspects • only discrete • *not supported but used*	Janus, KID, VDDE Framer, KID
Evaluative procedures	• analytical • comparative (differential) • global	• analytical • comparative (differential) -	Janus, KID, VDDE Attending, Oncocin
Evaluative knowledge	• limited set of evaluative referents • criteria, constructed constraints • prescribed constraints • deduced constraints • solutions of reference	• limited set of evaluative referents • generic critics* • specific critics* * *dependent on the current step of problem-solving* • solutions of reference	Janus, KID, VDDE Janus, KID, VDDE KID Janus, KID, VDDE Attending, Oncocin
Role of evaluation	• determination of the following step of design • selection of design solutions • definition of an order of priority among criteria and constraints	• determination of the following step of design: - *indirect contribution* - *more direct* • selection of design solutions based on: - *the system's choice* - *the user's preferences* • definition of an order of priority among criteria and constraints: - *by the user or the system designer* - *by the system and in accordance to the user's preferences*	 Janus, KID AgentSheets Attending Multicriteria Decision system KID Framer Multicriteria Decision system

In addition, we specifically identified certain aspects of the evaluation process in the design activities: the evaluation of conceptual solutions, and the use of a global mode of evaluation in order to determine immediately if a solution can fit a problem. These aspects seem to depend on the designers' own experience and, thus, may be difficult to embed in current design critiquing systems. Nevertheless, the development of the various functionalities we suggested could increase the compatibility of the functioning of critiquing systems and the cognitive "functioning" of the users, and therefore contribute to facilitate human-computer cooperation.

ACKNOWLEDGMENTS

I would like to thank Tamara Sumner and Ray McCall for their helpful comments and discussions on a previous version of this paper. Thanks also to Kumiyo Nakakoji for her demonstration and explanation of the KID system.

This research was supported by CNRS (Programme Cognisciences), and the analyses of human activities were performed while the author was working at INRIA (Projet de Psychologie Ergonomique) under a contract with AEROSPATIALE, and was a member of the CREPCO-University of Provence (France).

REFERENCES

Bisseret, A., Figéac-Letang, C., and Falzon, P. (1988). *Modeling opportunistic reasonings: The cognitive activity of traffic signal setting technicians*. (Research report no. 893). Rocquencourt: INRIA.

Boissier, D. & Al-Hajjar, J. (1990). Logique floue et aide à la décision: Application à l'infrastructure des bâtiments. *Actes du Congrès O1.DESIGN* (Cabourg, France, October 21-23).

Bonnardel, N. (in press). Expertise transfer, knowledge elicitation, and delayed recall in a design context. *Behaviour & Information Technology.*

Bonnardel, N. (1992a) Le rôle de l'évaluation dans les activités de conception. Ph.D. Dissertation, University of Provence, July.

Bonnardel, N. (1992b). The functions of evaluation in design activities. *Abstracts of the Fitfth Conference of the European Society for Cognitive Psychology* (Paris, France, September 12-16).

Bonnardel, N. (1989). *L'évaluation de solutions dans la résolution de problèmes de conception.* (Rapport de Recherche n°1072). Rocquencourt: INRIA.

Eastman, C.M. (1969). Cognitive processes and ill-defined problems: A case study from design. *Proeedings of the First Joint International Conference on I.A.* (Washington, D.C.), 669-690.

Fischer, G. (1987). A Critic for Lisp. *Proceedings of the 10th International Joint Conference on Artificial Intelligence IJCAI'87* (Milan, Italy, August), 177-184.

Fischer, G., Lemke, A.C., Mastaglio, T., and Morch, A. (1991). Critics: An emerging approach to knowledge-based human-computer interaction. *International Journal of Man-Machine Studies*, 35, 695-721.

Fischer, G., McCall, R., and Morch, A. (1989). Design environments for constructive and argumentative design. *Human Factors in Computing Systems, Proceedings of CHI'89* (Austin, TX, April 30-May 4), 269-275.

Fischer, G., Nakakoji, K., Ostwald, J., Stahl, G. and Sumner, T. (1993). Embedding computer-based critics in the context of design. *Proceedings of INTERCHI'93* (Amsterdam, The Netherlands, April 24-29), 157-164.

Fustier, M. (1989). *La résolution de problèmes: méthodologie de l'action.* Paris: Editions ESF & Librairies Techniques.

Guindon, R. (1990). Knowledge exploited by experts during software system design. *International Journal of Man-Machine Studies*, Special Issue: What programmers know, 33 (3), 279-304.

Guindon, R. (1990). Knowledge exploited by experts during software system design. *International Journal of Man-Machine Studies*, Special Issue: What programmers know, 33 (3), 279-304.

Langlotz, C.P., and Shortliffe, E.H. (1983). Adapting a consultation system to critique user plans. *International Journal of Man-Machine Studies*, 19, 479-496.

Lemke, A.C., and Fischer, G. (1990). A cooperative problem-solving system for user interface design. *Proceedings of AAAI-90* (Cambridge, MA, August), 479-484.

Mastaglio, T. (1990). User modelling in computer-based critics. *Proceedings of the 23rd annual Hawai International Conference on Systems Sciences* (Kailua-Kona, January), 403-412.

Miller, P.L. (1984). *"A Critiquing Approach to Expert Computer Advice: ATTENDING"*, Boston: Pitman Advanced Publishing Program.

Mostow, J. (1985) Towards better models of the design process. *AI Magazine*, Spring, 44-57.

Nakakoji, K. (1993). *Increasing Shared Knowledge Between Humans and Design Environments: The Role of a Specification Component.* Ph.D. Dissertation, University of Colorado, April.

Reitman, W.R. (1965). *Cognition and Thought : An information processing approach.*, New-York: Wiley.

Repenning, A., and Citrin, W. (1993). Agentsheets: applying grid-based spatial reasoning to human-computer interaction, *Proceedings of the 1993 IEEE Workshop on Visual Languages* (Bergen, Norway: IEEE Computer Society Press).

Repenning, A., and Sumner, T. (1992). Using agentsheets to create a voice dialog design environment. *Proceedings of the 1992 ACM/SIGAPP Symposium on Applied Computing* (Kansas City, MO, March 1-3), 1199-1207.

Rieman, J. (1993). The diary study: a workplace-oriented research tool to guide laboratory efforts. *Proceedings of INTERCHI'93* (Amsterdam, The Netherlands, April 24-29), 321-326.

Simon, H.A. (1973). The Structure of Ill Structured Problems. *Artificial Intelligence*, 4, 181-201.

Visser, W. (1990). More or less following a plan during design: Opportunistic deviations in specification. *International Journal of Man-Machine Studies*, Special Issue: What programmers know, 33 (3), 247-278.

Visser, W., and Falzon, P. (1988). Recueil et analyse de l'expertise dans une activité de conception: questions de méthode. *Psychologie Française*, numéro spécial "Psychologie de l'expertise", 33 (3), 133-138.