To appear in the *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, Montreal, August 1995.

Adaptive Provision of Evaluation-Oriented Information: Tasks and Techniques

Anthony Jameson, Ralph Schäfer, Joep Simons, and **Thomas Weis***
Department of Computer Science, University of Saarbrücken
P.O. Box 151150, D-66041 Saarbrücken, Germany
{jameson, ralph, jsimons, tweis}@cs.uni-sb.de

Abstract

Evaluation-oriented information provision is a function performed by many systems that serve as personal assistants, advisors, or sales assistants. Five general tasks are distinguished which need to be addressed by such systems. For each task, techniques employed in a sample of systems are discussed, and it is shown how the lessons learned from these systems can be taken into account with a set of unified techniques that make use of well-understood concepts and principles from Multi-Attribute Utility Theory and Bayesian networks. These techniques are illustrated as realized in the dialog system PRACMA.

During the past two decades, a number of AI systems have been developed whose overall task can be characterized as *evaluation-oriented information provision*: The user (to be called the *evaluator*, or $\mathcal{E}$) has the goal of making evaluative judgments about one or more *objects*; the system (or *information-provider*, $\mathcal{I}$) supplies $\mathcal{E}$ with information to help $\mathcal{E}$ make these judgments. Table 1 lists a representative sample of five such systems, which will be referred to as *EOIPs*.[1] The number of such systems seems likely to grow in the near future, especially given the recent interest in personal assistants —some of which advise their users on evaluative judgments —and teleshopping, which should increase the demand for automated sales assistants.

EOIP systems differ considerably in the techniques they employ for interaction with the user and for internal processing. For example, the communication with the INFORMATION FILTERING SYSTEM and the SALES ASSISTANT is realized with direct manipulation and hypertext techniques, whereas the other three systems use some form of natural language. There are also large differences in the theoretical frameworks and terminology in which EOIPs are presented. These differences impede exchange and consolidation of results. The present paper aims (a) to remedy this state of affairs by providing a unified framework for analyzing the techniques used in EOIPs; and (b) to advance the state of the art by presenting some new techniques which should be generally applicable within EOIPs.

Table 1: Overview of Five Representative Systems for Evaluation-Oriented Information Provision

System	Reference	Example Evaluator	Example Objects
GRUNDY	[Rich, 1979]	Library user	Library books
INFORMATION FILTERING SYSTEM[a]	[Sheth and Maes, 1993]	Reader of network news	Individual news articles
CONSULT[b]	[Elzer *et al.*, 1994]	University student	University courses
SALES ASSISTANT	[Popp and Lödel, 1994]	Potential buyer (e.g. of personal computer)	Products offered for sale
PRACMA	[Jameson *et al.*, 1994]	Potential used-car buyer	Available cars

[a] No system name was given in the cited paper.
[b] This name was introduced after the appearance of the cited paper.

Table 2 gives an overview of five general tasks which are at least potentially relevant to any EOIP. These will be discussed in turn in the five sections to follow. The new techniques will be presented in the context of the fifth of the reference systems, PRACMA. The excerpt from an example dialog in Table 3 both gives a sense of the nature of PRACMA's dialogs and provides initial examples of the five tasks.

1 Task 1: Predict Overall Evaluations

It is almost inevitable for an EOIP to try to predict how the user $\mathcal{E}$ would evaluate individual objects in the domain if he[2] had complete information about them. For example, though it is clear in the used-car domain that the buying decision will ultimately be made by $\mathcal{E}$, $\mathcal{I}$ needs to predict $\mathcal{E}$'s overall evaluations in order to narrow the discussion to one or more

*This research is being supported by the German Science Foundation (DFG) in its Special Collaborative Research Program on Artificial Intelligence and Knowledge-Based Systems (SFB 314), project N1, PRACMA.

[1] Not included are expert systems that perform evaluation tasks using evaluation criteria that have no necessary relationship to the criteria of the user (see, e.g., [Klein and Shortliffe, 1994]).

[2] For clarity, masculine and neuter pronouns will be used to refer to $\mathcal{E}$ and $\mathcal{I}$, respectively.

Table 2: Overview of Five General Tasks for an Evaluation-Oriented Information Provider

1. **Predict Overall Evaluations:** Anticipate how $\mathcal{E}$ would evaluate one or more domain objects, perhaps relative to one another, given complete knowledge about them.
2. **Predict Partial Evaluations:** Anticipate the impact that information about an attribute of an object would have on $\mathcal{E}$'s evaluation of that object.
3. **Interpret Evidence:** Update the model of $\mathcal{E}$'s evaluation criteria on the basis of evidence in $\mathcal{E}$'s actions.
4. **Elicit Evidence:** Induce $\mathcal{E}$ to perform actions that will constitute evidence for the task "Interpret Evidence".
5. **Select Dialog Moves:** Determine what type of dialog move to make (e.g., formulate recommendation; ask question about $\mathcal{E}$'s criteria; allow $\mathcal{E}$ to act next).

objects.

Almost all EOIPs appear to be based on some particular conceptualization of how $\mathcal{E}$ would evaluate an object given complete information about it.[3] In most cases the conceptualization can be seen as a variant of a conceptualization known by the name *Multi-Attribute Utility Theory* (*MAUT*) and similar names (see, e.g., [von Winterfeldt and Edwards, 1986]).

1.1 The MAUT Conceptualization

Some basic concepts of this conceptualization are illustrated in Figure 1, which shows part of a *value tree* that a used-car customer consulting PRACMA might use to evaluate a particular car. Each leaf corresponds to an *attribute*, which for each particular object has a *level* within a given range. For each attribute, $\mathcal{E}$ has a *value function* which assigns to each possible level a *value* between 0 and 10 (for example, for the attribute "Mileage", the values might be 10, 3, and 0 for the levels "0 – 10,000", "40,000 – 50,000", and "80,000 – 90,000", respectively). To take into account differences in the importance of attributes, each branch in the value tree has an *importance weight* between 0.0 and 1.0. For example, the importance weights on the branches leading down to "Mileage" and "Time to Inspection" specify their relative weights with respect to the *value dimension* of "Reliability"; the weights on the branches leading down to "Reliability" and "Safety" in turn specify the relative weights of these dimensions with respect to the overall evaluation of the car. The absolute importance weight of each individual attribute (shown below the corresponding leaf in Figure 1) is the product of the weights on the branches leading to it. If $\mathcal{E}$ had complete information about an object, $\mathcal{E}$ could arrive at an overall evaluation by determining the object's value with respect to each attribute, multiplying each such value by its absolute importance weight, and adding up the weighted values.[4]

[3]Strictly speaking, $\mathcal{I}$ need not presuppose that $\mathcal{E}$ really evaluates objects in accordance with the conceptualization used, only that it is useful for $\mathcal{I}$ to act as if $\mathcal{E}$ did so. Given the constructive, task-dependent, and situation-dependent nature of human evaluation processes (see, e.g., [Payne *et al.*, 1992]), for descriptive purposes any single conceptualization is best viewed as a rough approximation.

Table 3: Part of an Example Dialog with PRACMA Illustrating Five Tasks of Evaluation-Oriented Information Provision

Dialog contribution	Task(s) performed by $\mathcal{I}^a$
$\mathcal{I}$: What kind of car are you looking for?	4. Elicit evidence
$\mathcal{E}$: It shouldn't cost more than about 10,000. It shouldn't be too old.	3. Interpret evidence
$\mathcal{I}$: What kind of work do you do?	4. Elicit evidence
$\mathcal{E}$: I am a teacher.	3. Interpret evidence
$\mathcal{I}$: I have something you might like.	1. Predict overall evaluations
It's a Rabbit. Its year of construction is 1990.	2. Predict partial evaluations
$\mathcal{E}$: How long is the time to the next official inspection?	3. Interpret evidence
$\mathcal{I}$: The time to the next inspection is 2 years.	2. Predict partial evaluations
$\mathcal{E}$: <No reaction>	3. Interpret evidence
$\mathcal{I}$: Its mileage is 40,000.	2. Predict partial evaluations
$\mathcal{E}$: That's good.	3. Interpret evidence

[a]Task 5, "Select dialog moves", is performed each time $\mathcal{I}$ produces an utterance or gives $\mathcal{E}$ a chance to do so.

Variants of MAUT Used in EOIPs

This basic conceptualization takes different forms in different EOIP systems and sometimes remains implicit.

- ▷ The SALES ASSISTANT treats the value function for an attribute as a fuzzy membership function representing a concept like "Has at least 32 Mb RAM", and the weight of an attribute is represented by a membership function corresponding to a natural language formulation like "quite important".
- ▷ CONSULT may, for example, ascribe to $\mathcal{E}$ a "negative preference" for courses beginning after 6 p.m., in effect ascribing a particular value function mapping times of day onto values; the characterization of this negative preference as "strong" in effect assigns to this attribute a high importance weight.
- ▷ The INFORMATION FILTERING SYSTEM, which uses techniques from Artificial Life, provides an evolving population of *agents*. Each agent in effect ascribes to $\mathcal{E}$ a simple value tree that it uses to evaluate and recommend news articles. Each attribute corresponds to the presence or absence of a particular keyword (or other feature) in the article being evaluated, and each attribute has at any

[4]A single attribute can affect the evaluation of an object with respect to more than one value dimension (e.g., "Horsepower" has implications for both "Sportiness" and "Environmental Friendliness"). Although cases involving these and additional complications are handled by PRACMA (cf. [Schäfer, 1994]), a discussion of their proper treatment would exceed the scope of this paper.

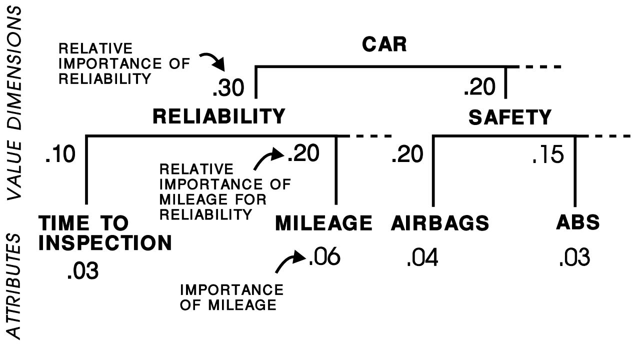

Figure 1: Part of a value tree that might be used by an Evaluator in PRACMA's example domain.

time an importance weight, which can change as the agent adapts to $\mathcal{E}$.

Throughout this paper, terminology from the MAUT literature will be used instead of the terminology used in the original publications on the reference systems.

1.2 Handling Uncertainty About the Evaluator's Criteria

Even assuming that $\mathcal{E}$'s evaluation processes can be described perfectly in terms of a value tree, an EOIP can rarely have a complete and accurate view of $\mathcal{E}$'s evaluation criteria. For example, with respect to the value tree in Figure 1, $\mathcal{E}$s can differ widely in the importance weights they attach to value dimensions like "Reliability", and independently of this they may attach idiosyncratic relative weights to individual attributes like "Mileage".

Treatment in Other EOIPs

Some systems, such as the INFORMATION FILTERING SYSTEM and the SALES ASSISTANT, in effect make use of their best specific estimate as to the content of $\mathcal{E}$'s value tree. They therefore do not distinguish between predictions in which they are confident and those which represent mere guesses. This distinction may in fact be of minor importance if the EOIP evaluates a large number of objects on $\mathcal{E}$'s behalf and if the consequences of an incorrect prediction are not serious.

Other EOIPs represent uncertainty about $\mathcal{E}$'s evaluation criteria explicitly.

▷ When ascribing to $\mathcal{E}$ a particular value function, CONSULT associates with this ascription (a) a confidence rating and (b) a list of endorsements for the ascription. When predicting how $\mathcal{E}$ would evaluate various courses, the system takes into account only attributes about whose value functions it has at least moderate confidence.

Note that if an EOIP restricts its attention to attributes about which it is confident, it still cannot be confident that its overall predictions are accurate. For example, an object that rates highly with respect to one attribute may be extremely attractive to $\mathcal{E}$ even though $\mathcal{I}$ as yet has no evidence that $\mathcal{E}$ assigns high importance to that attribute. It is therefore desirable for $\mathcal{I}$ to be able to derive some sort of confidence interval for its predictions of $\mathcal{E}$'s overall evaluations.

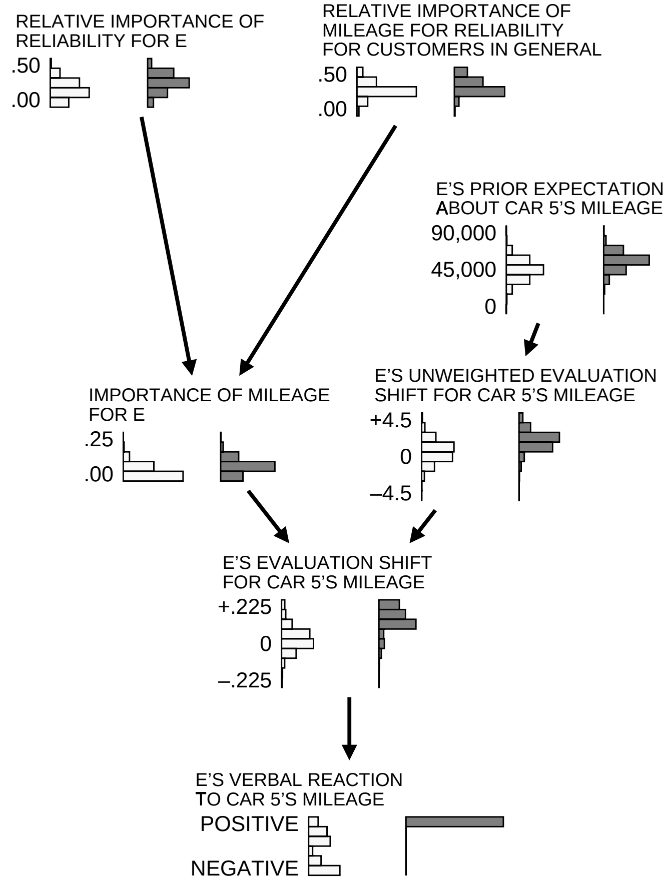

Figure 2: Part of a Bayesian network constructed by PRACMA to predict and interpret evaluative reactions to a statement.

(Arrows point from parent to child nodes. The darker histograms represent the beliefs the system derives through upward propagation on the basis of $\mathcal{E}$'s positive reaction to the statement "Its mileage is 40,000"—cf. section 3.)

Managing Uncertainty with Bayesian Networks

This problem (and others to be discussed below) can be handled effectively with the help of Bayesian networks.[5] This approach will be discussed in most detail in connection with PRACMA's handling of Task 2, "Predict partial evaluations"; but some of the basic concepts can be illustrated in terms of the three network nodes depicted in the upper left-hand part of Figure 2. These nodes show how $\mathcal{I}$'s uncertainty concerning the importance weights relevant to the attribute "Mileage" can be handled. (Until section 3 we will refer only to the first of the two histograms shown for each node.)

In the node RELATIVE IMPORTANCE OF RELIABILITY FOR E, the first histogram depicts a probability distribution representing $\mathcal{I}$'s initial belief about a variable X, namely the relative weight that the current $\mathcal{E}$ attaches to "Reliability". Whereas in Figure 1 X simply had the value .30, here $\mathcal{I}$'s belief about X is a probability distribution over the possible values that X can assume. For reasons of computational tractability, X is approximated as a discrete variable with five possible

[5] For theoretical and technical background on Bayesian networks see, e.g., [Pearl, 1988], whose notation and concepts are adopted in the present paper, or [Neapolitan, 1990].

values, corresponding to the midpoints of the intervals .00–.10, .10–.20, .20–.30, .30–.40, and .40–.50. $\mathcal{I}$'s probability distribution for X can therefore be viewed as a five-element vector $BEL(x)$.[6] In the example, $\mathcal{I}$ considers it most probable that $\mathcal{E}$'s weight is around .15 but that the weight might also be as high as about .35. The node RELATIVE IMPORTANCE OF MILEAGE FOR RELIABILITY FOR CUSTOMERS IN GENERAL represents $\mathcal{I}$'s belief about the average weight of the mileage attribute, relative to the value dimension of reliability, in the population of customers that $\mathcal{I}$ deals with.

On the basis of its beliefs about these two variables, $\mathcal{I}$ can form a belief about the IMPORTANCE OF MILEAGE FOR E, as indicated by the arrows showing that this third node is a *child* of the former two *parent* nodes. But the relationship between the parent nodes and the child node is probabilistic: Even if $\mathcal{I}$ knew the exact values of the two variables in the parent nodes, it could not be sure that the value of the variable in the child node was simply their product, because the relative weight that this particular $\mathcal{E}$ attaches to "Mileage" may deviate from the relative weight for customers in general.

In a Bayesian network, a probabilistic relation between two parent nodes corresponding to variables X and Y and a child node corresponding to a variable Z is represented by a matrix of conditional probabilities $P(z|x,y)$ which contains one probability for each possible combination of values of Z, X, and Y. For these particular three nodes, the matrix is generated using a function which specifies that the probability of Z taking a given value z is highest when that value is close to the product of the values x and y of the two parent variables.[7]

$\mathcal{I}$ arrives at a belief concerning the child variable through the standard top-down propagation procedure for singly connected Bayesian networks.[8] In this example, we see that $\mathcal{I}$ has a moderate amount of uncertainty about the weight that information about a car's mileage will have in determining $\mathcal{E}$'s evaluation of it; the probability distribution for IMPORTANCE OF MILEAGE FOR E would be still narrower, for example, if $\mathcal{I}$ somehow had acquired a definite belief about the RELATIVE IMPORTANCE OF RELIABILITY FOR E.

This type of uncertainty ultimately affects $\mathcal{I}$'s prediction of $\mathcal{E}$'s overall evaluation of a given car. This prediction task involves many nodes not depicted in Figure 2, but its treatment is largely analogous to that of the prediction of partial evaluations, which is discussed in the following section.

[6]For each possible value x of X,

$$BEL(x) \triangleq P(X = x|\boldsymbol{e}),$$

where $\boldsymbol{e}$ represents all of the evidence received so far.

[7]The basic function is given by

$$P(z|x,y) = F(z - xy + .05) - F(z - xy - .05),$$

where F is the cumulative distribution function for a normal distribution with mean 0 and standard deviation 0.10.

[8]See, e.g., [Pearl, 1988, chap. 4]. In the present relatively simple case, where there is initially no other evidence concerning the child variable, the resulting belief vector is related to those for the parent nodes as follows:

$$BEL(z) = \sum_{x,y} P(z|x,y) BEL(x) BEL(y).$$

2 Task 2: Predict Partial Evaluations

In addition to predicting the evaluation of entire objects, an EOIP should be able to anticipate the impact that information about a particular attribute of an object will have on $\mathcal{E}$'s evaluation of that object. First, even if $\mathcal{I}$ is already certain that a given object should be recommended to $\mathcal{E}$, $\mathcal{I}$ may have to explain its recommendation.[9] Second, $\mathcal{I}$ may in some cases not attempt to evaluate entire objects for $\mathcal{E}$ at all, pursuing instead the more modest goal of efficiently supplying information that allows $\mathcal{E}$ to arrive at evaluations of his own.

2.1 Treatment in Other EOIPs

Some EOIPs supplement their recommendation of an object with a description of those attributes of the object that the system expects to have the greatest impact on $\mathcal{E}$'s evaluation.

- ▹ In particular, CONSULT illustrates that this sort of selection of attributes must sometimes concern *relative* evaluations: When recommending an alternative to a university course selected by the user, the system describes only those attributes with respect to which the alternative course is likely to be evaluated substantially higher by $\mathcal{E}$ than the original choice.

This capability for selective description of objects is not required only in natural language systems with a narrow communication bandwidth. For example, systems like the INFORMATION FILTERING SYSTEM often present an overview of a fairly large set of objects, each of which has to be characterized briefly. So it may be worthwhile, for example, for $\mathcal{I}$ to select an especially evaluation-relevant subset of each object's attributes to be displayed in graphical or tabular form.

2.2 Predicting Partial Evaluations with Bayesian Networks

The way in which Bayesian networks can be applied to this task will be illustrated for a particular type of relative partial evaluation: An *evaluation shift* is the change in $\mathcal{E}$'s evaluation of an object with respect to a given attribute after $\mathcal{E}$ has received information about the object's level with respect to that attribute (cf. [Jameson, 1989]). An evaluation shift is actually more relevant than an absolute evaluation for determining which facts $\mathcal{I}$ should mention. For example, even if $\mathcal{I}$ knows that $\mathcal{E}$ assigns a high weight to the attribute "Mileage", there is little point in mentioning that a given car has low mileage if $\mathcal{E}$ has already been told that it is only a few weeks old: The statement could in this case hardly produce a substantial shift in $\mathcal{E}$'s evaluation.

A straightforward way of using Bayesian networks to predict an evaluation shift would be to make separate predictions of $\mathcal{E}$'s evaluations of an object *before* and *after* $\mathcal{I}$'s statement with respect to an attribute (e.g., that Car 5's mileage is 40,000). Each of these predictions would make use of $\mathcal{I}$'s prediction of IMPORTANCE OF MILEAGE FOR E (cf. Figure 2 and the discussion in the previous section); a comparison of the nodes representing the two resulting predictions would give some indication of $\mathcal{E}$'s likely evaluation shift. This method

[9]Klein and Shortliffe [1994] present sophisticated techniques for explaining evaluative decisions which have been arrived at using a particular, known value tree—which may have been acquired either from the user or from an independent expert.

would be invalid, however: $\mathcal{I}$'s uncertainty about $\mathcal{E}$'s importance weights would enter into both of the predictions, leading to much more uncertainty in the prediction of the evaluation shift than would be necessary or justified.

The remaining part of Figure 2 (except for the bottommost node, which will be discussed in the next two sections) shows how this problem can be avoided. PRACMA dynamically constructs a partial network like this whenever it considers making a statement about a particular attribute of a car—in this case, the statement "Its mileage is 40,000". The node E'S PRIOR EXPECTATION ABOUT CAR 5'S MILEAGE represents $\mathcal{I}$'s belief about what $\mathcal{E}$ would consider the most likely mileage for Car 5 before $\mathcal{E}$ obtained any information from $\mathcal{I}$; as the first histogram for the node illustrates, $\mathcal{I}$ can initially have only a rather indefinite belief about this expectation. On the basis of this variable, $\mathcal{I}$ can try to predict the extent to which $\mathcal{E}$'s unweighted evaluation of Car 5's mileage (on a scale from 0 to 10) will shift upward or downward after $\mathcal{I}$'s statement. $\mathcal{I}$'s belief about this shift is shown in the first histogram for E'S UNWEIGHTED EVALUATION SHIFT FOR CAR 5'S MILEAGE: $\mathcal{I}$ considers it slightly more probable that the shift will be positive than that it will be negative (because 40,000 is a bit more likely to be a lower mileage than $\mathcal{E}$ expects than it is to be a higher one).[10]

The node E'S EVALUATION SHIFT FOR CAR 5'S MILEAGE is of most relevance for $\mathcal{I}$ in deciding what to say, as it reflects the impact that $\mathcal{I}$'s statement is likely to have on $\mathcal{E}$'s overall evaluation of Car 5, which of course depends in part on IMPORTANCE OF MILEAGE FOR E. The prediction of this variable on the basis of its two parent variables proceeds in a way similar to that described in section 1 for the prediction of IMPORTANCE OF MILEAGE FOR E itself (in both cases the basic underlying relationship is multiplicative). The first histogram for E'S EVALUATION SHIFT FOR CAR 5'S MILEAGE shows that $\mathcal{I}$ considers it unlikely that its statement will influence $\mathcal{E}$'s overall evaluation (which will be on a scale from 0 to 10) by more than about .1 in either direction. This example illustrates that it is often possible to make a fairly definite prediction about a *change* in a variable even if one has only indefinite beliefs about the initial and later levels of the variable—if the uncertainty that is common to the two beliefs is handled appropriately.

3 Task 3: Interpret Evidence

Most EOIPs refine their models of $\mathcal{E}$ on the basis of evidence supplied by $\mathcal{E}$ during the interaction. The ways in which $\mathcal{E}$ can give useful clues include: explicitly characterizing his evaluation criteria ("I'm interested in politics"), making requests for particular types of information ("What books/courses/articles do you have that involve politics?"), expressing evaluative judgments he has arrived at ("This object [which involves politics] looks good"), and reporting personal characteristics that have implications for his evaluation criteria ("I'm a law student").

3.1 Treatment in Other EOIPs

The most common approach to processing this type of evidence is to adjust one or more parameters of $\mathcal{I}$'s model of $\mathcal{E}$ (e.g., $\mathcal{I}$'s representation of the importance of politics for $\mathcal{E}$) in the direction suggested by the evidence, with the magnitude of the adjustment depending on the nature of the evidence.

- ▷ When an article suggested by one of the INFORMATION FILTERING SYSTEM's agents is evaluated positively by the user, the importance weight associated with each of the article's keywords is increased.
- ▷ Whenever GRUNDY processes a self-description or an evaluative reaction to a library book from the user, the system adjusts a number of quantitative assessments it has made—for example, concerning the specific $\mathcal{E}$'s interest or concerning the long-term content of the general stereotypes that the system has associated with the user.

In these systems, the direction and relative magnitudes of the adjustments in $\mathcal{I}$'s model of $\mathcal{E}$ can be justified fairly plausibly, but there is a good deal of arbitrariness in the details. This limitation may be of minor importance if $\mathcal{I}$ will have the opportunity to process a large amount of evidence concerning a given aspect of its model; in such cases the model can ultimately converge on realistic values even if the individual adjustments are not optimal. Where evidence is much more limited—for example, when it concerns a specific $\mathcal{E}$ and is acquired during a single interaction—it is desirable for adjustments to $\mathcal{I}$'s model to be justifiable more specifically.

3.2 Probabilistic Evidence Interpretation

This goal can be achieved within the Bayesian network framework used by PRACMA, if the relevant aspects of $\mathcal{E}$'s behavior are represented by nodes which have precisely defined probabilistic links to the nodes that represent unobservable states of $\mathcal{E}$. Although the conditional probabilities defining these links may be based on intuitively plausible assumptions made by the designer rather than on empirical data, at least the details of the system's inferences can be understood and justified in terms of these assumptions.

This way of handling evidence in $\mathcal{E}$'s actions is illustrated by the way PRACMA interprets an explicit evaluative reaction like "That's good" following a statement that it has made (cf. Table 3). The type of reaction (including possibly "<no reaction>") that $\mathcal{E}$ produces is represented by a node in the Bayesian network—E'S VERBAL REACTION TO CAR 5'S MILEAGE in Figure 2. This node distinguishes several categories of evaluative verbal reactions that were observed in an unpublished empirical study. The matrix of conditional probabilities linking this node with its parent E'S EVALUATION SHIFT FOR CAR 5'S MILEAGE was derived indirectly from the data of this study.

Before $\mathcal{I}$ observes $\mathcal{E}$'s reaction, $\mathcal{I}$ has only an indefinite belief as to what E'S VERBAL REACTION TO CAR 5'S MILEAGE will be, as shown in the first histogram for the node. But after $\mathcal{E}$ has responded with "That's good", $\mathcal{I}$ has a completely definite belief, shown in the second histogram. Now a process of *upward propagation* can begin, in which the beliefs associated with the ancestor nodes of E'S VERBAL REACTION TO CAR 5'S MILEAGE

[10] The matrix of conditional probabilities linking these two nodes presupposes that $\mathcal{E}$'s value function for "Mileage" is similar to the one that $\mathcal{I}$ assumes for customers in general, but it takes into account possible idiosyncratic variation.

are updated in the light of the new evidence.[11] The second histogram for each ancestor node shows the updated beliefs. $\mathcal{I}$'s belief about E'S EVALUATION SHIFT FOR CAR 5'S MILEAGE is most directly affected: It is now almost certain that $\mathcal{E}$'s evaluation shift was in fact positive. Less directly, $\mathcal{I}$ confirms that $\mathcal{E}$ expected Car 5 a priori to have a higher mileage than 40,000 (E'S PRIOR EXPECTATION ABOUT CAR 5'S MILEAGE), and $\mathcal{I}$ also increases the extent to which it believes that $\mathcal{E}$ assigns high importance to "Reliability" in general and to "Mileage" in particular. Note also the slight positive shift in RELATIVE IMPORTANCE OF MILEAGE FOR RELIABILITY FOR CUSTOMERS IN GENERAL, which shows that $\mathcal{I}$ is gradually learning, on the basis of $\mathcal{E}$'s actions, about the evaluation criteria of customers in general.

An entirely analogous approach is used in PRACMA to interpret the fact that $\mathcal{E}$ has asked a question about a specific attribute. When, on the other hand, evidence becomes available that is directly related to $\mathcal{E}$'s evaluation criteria or to a relevant personal characteristic (cf. the examples in Table 3), less complex processing is required. For example, when $\mathcal{E}$ says "I'm especially interested in reliability", a corresponding node is attached directly as a child under the node RELATIVE IMPORTANCE OF RELIABILITY FOR E. The conditional probabilities linking the child node to the parent node reflect the fact that the likelihood of $\mathcal{E}$'s making a statement like this is a positive function of the actual importance of reliability for $\mathcal{E}$ but that the utterance does not uniquely determine any particular degree of importance.

4 Task 4: Elicit Evidence

Given $\mathcal{I}$'s need for evidence from $\mathcal{E}$ in order to update its model of $\mathcal{E}$, one natural task for $\mathcal{I}$ is to take steps so as to increase the likelihood that useful evidence of particular types will become available. For example, in the car sales domain, professional salespersons emphasize that they actively acquire a model of the customer by asking questions about personal characteristics and evaluation criteria and by encouraging the customer to express evaluative responses ([Simons, 1994]).

4.1 Treatment in Other EOIPs

EOIPs that exploit the broad band-width of modern human-computer interfaces can make it easy for $\mathcal{E}$ to enter information about himself optionally and with a minimum of distraction from his primary task.

- ▹ The INFORMATION FILTERING SYSTEM allows $\mathcal{E}$, after reading an article, to express his evaluation by clicking on a thumbs-up or thumbs-down icon displayed above the article; and to express interest in particular attributes by highlighting words in the text of the article.

In cases where techniques such as these are not applicable and/or where the consequences of $\mathcal{I}$'s use of an inaccurate model can be serious, some more obtrusive elicitation of information from $\mathcal{E}$ may prove inevitable. One issue that then arises is how $\mathcal{I}$ can selectively elicit the information that will be of the greatest value. In EOIPs to date, this kind of selection decision has typically been made by the designer, not by the system itself on-line.

- ▹ GRUNDY always asks a new user to supply some self-descriptive words, and when it has described a potentially interesting book, it asks $\mathcal{E}$ "Does that sound good?". It is only when $\mathcal{E}$ has given a negative response to a question like this that GRUNDY asks questions chosen for their expected information value: It asks about $\mathcal{E}$'s evaluation of individual attributes of the book, starting with attributes for which $\mathcal{I}$ is most uncertain about $\mathcal{E}$'s evaluation.

4.2 Systematic Assessment of Information Value

If $\mathcal{I}$'s model of $\mathcal{E}$'s evaluation processes is cast in the form of a Bayesian network, general techniques for predicting the value of new information within this framework (cf. [Pearl, 1988, 6.3 – 6.4]) can be applied. A well-known approach within decision theory involves comparing the expected value of an outcome if a decision is made on the basis of some new information with the expected value if it is made without that information. For example, how much more is a used-car buyer's purchase likely to be worth if he performs particular tests on a candidate car before making his choice (see, e.g., [Qi *et al.*, 1994])? In the context of evaluation-oriented information provision, this approach would require quantitative evaluation of the ultimate consequences of a decision made by the informant $\mathcal{I}$ to elicit (or not to elicit) a given piece of information from $\mathcal{E}$. But such consequences are in general hard to anticipate and to quantify. For example, $\mathcal{I}$'s failure to elicit a relevant fact about $\mathcal{E}$ might ultimately lead to a less satisfactory decision by $\mathcal{E}$, or it might just cause a lengthening of the interaction between $\mathcal{E}$ and $\mathcal{I}$.

In such cases a useful criterion is often the extent to which new information will reduce the system's *uncertainty* about particular *target variables* whose values importantly influence the system's behavior.[12] In an EOIP that uses Bayesian networks in the way PRACMA does, interesting target variables include the importance weights that $\mathcal{E}$ assigns to the various value dimensions.

For example, suppose that in the example dialog (Table 3) $\mathcal{E}$ did *not* spontaneously express any evaluative reaction to $\mathcal{I}$'s statement about Car 5's mileage. Then $\mathcal{I}$ would have had to decide whether to elicit such a reaction (e.g., by asking "What do you think of that?"). One of the main benefits of doing so would be the kind of reduction in $\mathcal{I}$'s uncertainty about the node RELATIVE IMPORTANCE OF RELIABILITY FOR E that

[11] Upward propagation essentially uses Bayes' Rule to adjust the probability associated with each possible value of a variable in an ancestor node in accordance with the conditional probability of the observed evidence given that value. Although the computations are in general more complex, in the simple case of the two nodes at the bottom of Figure 2, the updated belief vector $BEL'(x)$ for the parent variable X after the observation $Y = y'$ is related to the prior belief vector $BEL(x)$ as follows:

$$BEL'(x) = \frac{BEL(x)P(y'|x)}{\sum_x P(y'|x)}.$$

[12] The cost $C(X)$ of $\mathcal{I}$'s uncertainty about a quantitative variable X can be defined in terms of the variance of the probability distribution representing $\mathcal{I}$'s belief about X, i.e.

$$C(X) = \sum_x BEL(x)(x - \mu_x)^2,$$

where μ_x is the mean of the distribution for X, i.e.

$$\mu_x = \sum_x BEL(x)x.$$

is illustrated by the change from the first to the second histogram for that node in Figure 2. Since $\mathcal{I}$ doesn't know in advance what type of reaction $\mathcal{E}$ will express, $\mathcal{I}$ must in effect perform the updating shown in Figure 2 for each possible reaction type, weighting the resulting uncertainty reductions by the prior probabilities of the reaction types[13] (these are shown in the first histogram for E'S VERBAL REACTION TO CAR 5'S MILEAGE). Generally speaking, eliciting evaluative reactions is especially worthwhile when a fact has contrary implications for two different relevant dimensions (e.g., high horsepower is positive for "Sportiness" and negative for "Environmental Friendliness"); in such cases, a single reaction by $\mathcal{E}$ often yields considerable information about the importance he assigns to the two value dimensions.

$\mathcal{I}$ can use the same general technique when deciding whether to elicit other types of reaction by $\mathcal{E}$, such as statements about personal characteristics and evaluation criteria (cf. the remarks at the end of section 3 on how such statements are interpreted by PRACMA).

5 Task 5: Select Dialog Moves

The five reference systems discussed have illustrated a number of types of dialog move that $\mathcal{I}$ can make, for example: asking about $\mathcal{E}$'s personal characteristics, answering questions, and volunteering unsolicited information. Though some criteria have been discussed in the preceding sections for choosing a move of a particular type (e.g., deciding which object to recommend), the more general question remains of how $\mathcal{I}$ should decide *which type of move* to make at which time; and when to give $\mathcal{E}$ the chance to make a move.

5.1 Treatment in Other EOIPs

A survey of other EOIPs suggests three general principles with respect to this question:

1. $\mathcal{I}$ should try to achieve an efficient and coherent sequence of dialog moves.

- ▹ The default ordering of the screens in the SALES ASSISTANT follows a sequence that is presumably efficient in most cases: obtaining information about various aspects of $\mathcal{E}$ and then making use of it to recommend products.

2. The criteria for selecting dialog moves should take into account evaluation-relevant information.

- ▹ CONSULT's decision as to whether to suggest a different university course than the one chosen by $\mathcal{E}$ depends not only on the dialog state but also on whether $\mathcal{I}$ has found a course which seems clearly superior.

This dependence of dialog moves on dynamically applied evaluation-relevant criteria makes sense in that the whole point of the interaction is to support $\mathcal{E}$'s evaluation process.

3. Both $\mathcal{E}$ and $\mathcal{I}$ should be able to influence the course of the interaction.

- ▹ The user of the SALES ASSISTANT has the option of ignoring the default screen sequence and navigating freely through the system.

Not only do users appreciate this sort of freedom, it also supports the goal of efficiency: Although $\mathcal{I}$ knows more about the domain objects than $\mathcal{E}$ does, $\mathcal{E}$ in general knows more about his own evaluation criteria. Because of this distribution of relevant knowledge over the two participants, each participant may at any point be in a better position to determine the direction the dialog should take.

5.2 Dialog Control Through Flexible Planning

PRACMA models the process of participating in a dialog in terms of the generation and execution of *dialog plans*. It uses a planner ([Weis, 1994]) which implements a basically hierarchical planning approach (cf. [Moore and Paris, 1989]) extended by special plan operators for modeling iterations on subgoals.

1. Efficiency and coherence. On the highest level, PRACMA's hierarchy of plan operators divides the dialog into phases corresponding to those found in sales dialogs (e.g., the phase in which $\mathcal{I}$ actively tries to acquire information about $\mathcal{E}$ precedes the phase for presenting information about relevant objects). On a lower level in the hierarchy, for each dialog phase there are several optional *strategies* that specify sequences which are efficient and correspond to dialog conventions (e.g., for the active acquisition phase, the strategies include "Ask about personal characteristics" and "Ask about requirements").

2. Evaluation-dependence. The applicability conditions of the plan operators refer not only to the nature of the preceding dialog moves but also to aspects of the model of $\mathcal{E}$'s evaluation process. In other words, the considerations mentioned in the previous sections that determine *which particular move of a given type* PRACMA makes—e.g., what kind of evaluation shift a given statement would produce in $\mathcal{E}$—are also used to determine *which type of move* the system makes at a given moment. This evaluation-dependent determination of what to do next often requires iteration: repeatedly achieving a given subgoal until it no longer appears worthwhile to do so.

3. Mixed initiative. After each of $\mathcal{I}$'s dialog moves, $\mathcal{I}$ gives $\mathcal{E}$ a chance to make the next move, even when $\mathcal{I}$ has already planned an appropriate next move of its own.[14] To enable PRACMA to accommodate a variety of dialog moves by $\mathcal{E}$, including those which don't fit well into $\mathcal{I}$'s plan, the dialog strategies include lower-level *tactics*, whose selection is influenced by $\mathcal{E}$'s actions. For example, the strategy "Ask about personal characteristics" includes a tactic which is applicable when $\mathcal{E}$ (unexpectedly) asks a specific question: $\mathcal{I}$ answers the question minimally and, after executing this tactic, continues to pursue the same strategy.

6 Conclusions

Table 4 summarizes the advances achieved by the techniques discussed in this paper relative to the overall state of the art in evaluation-oriented information provision.

A more general conclusion is that research in this area can benefit from increased use of relevant theoretical frameworks and techniques that are not specific to this topic. This strategy is analogous, for example, to the strategy underlying recent work that applies numerical uncertainty management techniques to the problem of plan recognition (see, e.g., [Charniak and Goldman, 1993; Bauer, 1995]). This type of research has

[13] Efficient techniques for performing the relevant computations are discussed in [Pearl, 1988, 6.4.2].

[14] Simulated facial expressions are currently being integrated through which $\mathcal{I}$ will be able to signal, among other things, the extent to which it considers it desirable for $\mathcal{E}$ to make the next move.

Table 4: Benefits of the Described Techniques for Systems for Evaluation-Oriented Information Provision

Tasks 1 and 2: Predict Overall and Partial Evaluations

- Prediction in terms of a probability distribution, yielding differentiated information to support $\mathcal{I}$'s dialog decisions.
- Simultaneous management of uncertainty with respect to a broad range of variables, from importance weights for $\mathcal{E}$s in general to the prior expectation of a particular $\mathcal{E}$ concerning an attribute of a particular object.
- Appropriate treatment of uncertainty in the prediction of evaluation shifts and other relative evaluations.

Task 3: Interpret Evidence

- Principled adjustment of $\mathcal{I}$'s beliefs concerning a variety of possible causes of $\mathcal{E}$'s observed behavior.
- Explicit representation of the probabilistic relationships between the observable behavior of $\mathcal{E}$ and unobservable variables.

Task 4: Elicit Evidence

- Dynamic selection of information-eliciting moves on the basis of context-dependent estimation of the value of the resulting information.

Task 5: Select Dialog Moves

- Integration of dialog planning with quantitative user modeling, permitting flexible support of $\mathcal{E}$'s evaluation processes.

shown that there is often a good fit between the tasks recurrently performed by a particular type of system and existing more general techniques; but that it is nonetheless a challenging research goal to work out an appropriate conceptualization of a task in terms of these techniques.

References

[Bauer, 1995] Mathias Bauer. A Dempster-Shafer approach to modeling agent preferences for plan recognition. *User Modeling and User-Adapted Interaction*, 1995. To appear in a special issue on Numerical Uncertainty Management in User and Student Modeling.

[Charniak and Goldman, 1993] Eugene Charniak and Robert P. Goldman. A Bayesian model of plan recognition. *Artificial Intelligence*, 64:53–79, 1993.

[Elzer *et al.*, 1994] Stephanie Elzer, Jennifer Chu-Carroll, and Sandra Carberry. Recognizing and utilizing user preferences in collaborative consultation dialogues. In *Proceedings of the Fourth International Conference on User Modeling*, pages 19–24, Hyannis, MA, 1994.

[Jameson *et al.*, 1994] Anthony Jameson, Bernhard Kipper, Alassane Ndiaye, Ralph Schäfer, Joep Simons, Thomas Weis, and Detlev Zimmermann. Cooperating to be noncooperative: The dialog system PRACMA. In Bernhard Nebel and Leonie Dreschler-Fischer, editors, *Proceedings of the Eighteenth Annual German Conference on Artificial Intelligence*, pages 106–117, Saarbrücken, 1994. Berlin: Springer.

[Jameson, 1989] Anthony Jameson. But what will the listener think? Belief ascription and image maintenance in dialog. In Alfred Kobsa and Wolfgang Wahlster, editors, *User Models in Dialog Systems*, pages 255–312. Springer, Berlin, 1989.

[Klein and Shortliffe, 1994] David A. Klein and Edward H. Shortliffe. A framework for explaining decision-theoretic advice. *Artificial Intelligence*, 67:201–243, 1994.

[Moore and Paris, 1989] Johanna D. Moore and Cecile L. Paris. Planning text for advisory dialogues. In *Proceedings of the Twenty-Seventh Annual Meeting of the Association for Computational Linguistics*, pages 203–211, Vancouver, 1989.

[Neapolitan, 1990] Richard E. Neapolitan. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*. Wiley, New York, 1990.

[Payne *et al.*, 1992] John W. Payne, James R. Bettman, and Eric J. Johnson. Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, 43:87–131, 1992.

[Pearl, 1988] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988. Revised second printing: 1991.

[Popp and Lödel, 1994] Heribert Popp and Dieter Lödel. Fuzzy techniques and user modeling in sales assistants. Manuscript submitted for publication, 1994.

[Qi *et al.*, 1994] Runping Qi, (Nevin) Lianwen Zhang, and David Poole. Solving asymmetric decision problems with influence diagrams. In Ramon Lopez de Mantaras and David Poole, editors, *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pages 491–497, Seattle, 1994. San Francisco: Morgan Kaufmann.

[Rich, 1979] Elaine Rich. User modeling via stereotypes. *Cognitive Science*, 3:329–354, 1979.

[Schäfer, 1994] Ralph Schäfer. Multidimensional probabilistic assessment of interest and knowledge in a noncooperative dialog situation. In Christoph G. Thomas, editor, *Proceedings of ABIS-94: GI Workshop on Adaptivity and User Modeling in Interactive Software Systems*, pages 46–62, Sankt Augustin, Germany, 1994.

[Sheth and Maes, 1993] Beerud Sheth and Pattie Maes. Evolving agents for personalized information filtering. In *Proceedings of the Ninth Conference on Artificial Intelligence for Applications, CAIA-93*, pages 345–352, Orlando, FL, 1993.

[Simons, 1994] Joep Simons. The elicitation of selling strategies and their formalization in a natural language system. Master's thesis, Department of Cognitive Science, University of Nijmegen, The Netherlands, 1994.

[von Winterfeldt and Edwards, 1986] Detlof von Winterfeldt and Ward Edwards. *Decision Analysis and Behavioral Research*. Cambridge University Press, Cambridge, England, 1986.

[Weis, 1994] Thomas Weis. VIPER: Ein verteilter inkrementeller Dialogplaner für eine Multi-Agenten-Umgebung [A distributed incremental dialog planner for a multi-agent environment]. Master's thesis, Department of Computer Science, University of Saarbrücken, 1994.