



Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition

FLORIANA GRASSO

Department of Computer Science, University of Liverpool, UK.

email: floriana@csc.liv.ac.uk.

ALISON CAWSEY

Department of Computing and Electrical Engineering, Heriot Watt University,

Edinburgh, UK. email: alison@cee.hw.ac.uk.

RAY JONES

Department of Public Health, University of Glasgow, UK. email: r.b.jones@udcf.gla.ac.uk.

Conflict situations do not only arise from misunderstandings, erroneous perceptions, partial knowledge, false beliefs, etc., but also from differences in “opinions” and in the different agents’ value systems. It is not always possible, and maybe not even desirable, to “solve” this kind of conflict, as the sources are subjective. The communicating agents can, however, use knowledge of the opponent’s preferences, to try and convince the partner of a point of view which they wish to promote. To deal with these situations requires an argumentative capacity, able to handle not only “demonstrative” arguments but also “dialectic” ones, which may not necessarily be based on rationality and valid premises. This paper presents a formalization of a theory of informal argumentation, focused on techniques to change attitudes of the interlocutor, in the domain of health promotion.

© 2000 Academic Press

1. Introduction

Conflict is generally defined as a situation in which two agents (or the same agent in case of internal conflict) hold different and incompatible goals (Easterbrook, Beck, Goodlet, Plowman, Sharples & Wood, 1993; Castelfranchi, 1996). There is less agreement on how conflicts can be solved and what the actual reasons are for the differences between the agents’ goals. Acknowledging discordant beliefs seems to be the most common approach (Sycara, 1988; Galliers, 1992; Yang, 1992; Chu-Carroll & Carberry, 1996; Reed & Long, 1997a), on the grounds that misconceptions, misunderstandings and incomplete knowledge all contribute in creating contrasting goals.

Conflicts arising between cooperative agents can be solved, for example, by means of a conversation, enlightening and manifesting the erroneous beliefs and misconceptions,

with the aim of letting the two parties reach common knowledge, and an agreement, if possible, on a shared plan.

However, in many real-life situations, conflicts are not, or at least not completely, due to different beliefs about reality. Often they arise from a difference in the two parties' *opinions* about reality, or the way they differently *value* events and circumstances. Often the partners cannot give evidence for their opinions, but they are still claimed, and are at the basis of everyday argumentation.

It is not always possible, and maybe not even desirable, to solve these kinds of conflict, as they are subjective and linked to personal factors. Nevertheless, the communicating agents can take advantage of knowing each other's preferences, and use them in order to try to let the partner see the matter from their own point of view.

To deal with this type of *persuasive* dialogue requires an argumentative ability able to take into account not only *demonstrative* proofs but also a *dialectic* style of reasoning. We refer to Aristotle's meaning of dialectic argument as an argument using premises which might not be evidently true, and whose aim is to make the conclusion more *acceptable* to the addressee, as opposed to demonstrative reasoning which deals with certainty and valid inferences (Aristotle, 1928).

This paper presents the formalization of such a theory of *informal argumentation*, applied to advice giving concerning behavioural change. Advice giving is a very interesting style of conversation. Unlike tutoring or information-seeking dialogues, the relationship between the two dialoguing partners is more likely to be peer-to-peer (Forslund, 1995). They both have their own expertise, even though in different areas, and one of the two parties' aims is to help the other with respect to some issue. Advice giving is also different from decision support, where one partner has to help the other to choose among different alternatives, and from collaboration, where the two partners have to actually perform a task jointly.

An important property of advice giving is that, due to the nature of the relation with the advisee, the advice giver may have to face situations in which the advisee is not receptive to the advice, and possibly not even aware of any need for change. The advisor has then to overcome scepticism and similar barriers.

This paper explores these issues through the development of a new type of advice giving system based on dialectic argumentation. The particular domain chosen for our investigation is health promotion, and in particular the promotion of healthy nutrition. The system, as described in this paper, is fully implemented, but it still lacks a friendly user interface. However, the general approach has been evaluated by means of email studies, and through the development of a simpler web-based prototype. These studies assessed the appropriateness and usefulness of the approach for health promotion, and suggested directions for further research.

The paper, after giving more insight on the application domain and the theory involved, introduces the New Rhetoric, a theory of practical argumentation which has been formalized in this work. The implementation of a model of an arguing agent is then described, followed by a discussion on the current state of the system and the evaluation experiments made so far. Comments about related research and further planned work conclude the paper.

2. Promoting behaviour changes: an argumentative scenario

Health Promotion can be described as “the process of enabling people to increase control over, and to improve, their health” (WHO, 1986). It is about developing addressees’ personal skills, through both providing information and advocating healthier behaviours. To be effective, the promoter has to understand the mechanism of behaviour change.

Researchers in this field have proposed several models of how and why people change their habits with respect to health matters. Among them, the trans-theoretical model of change or *Stages of Change Model* (Prochaska & Clemente, 1992) suggests that individuals progress through very distinct stages on their way to change their behaviour.

- *Precontemplation* stage: people see no problem with their behaviour and have no intention to change.
- *Contemplation* stage: people come to understand their problem, and start thinking about solving it.
- *Preparation* stage: people are planning to take an action in the immediate future.
- *Action* stage: people are in the process of actively making behaviour changes.
- *Maintenance* stage: characterizing a period of healthy behaviour continued on a regular basis.

The process of change, however, is not linear and a *relapse* is always possible at each stage.

For each stage, the theory suggests some actions an advisor can take, in order to help the advisee pass to the following stage. For instance, precontemplation advisees should be presented with information about the problems concerned with their behaviour, as they will hardly accept any advice on how to change without knowing why they should change in the first place. This set of practical hints helped us to set up the advising strategy in our system.

Another well-established model which we use in our implementation is the *Health Belief Model* (Becker, 1974). Unlike the previous one, which is concerned with the actions to undertake to help people change, this model focuses on the kinds of beliefs that influence people’s behaviour, and has therefore been useful for us in order to model the advisee’s beliefs. The Health Belief Model assumes that for people to actually take an action to avoid a negative situation they need to believe that they are personally *susceptible* to it and that its occurrence would have at least moderate *severity* on some components of their life. Moreover, they need to be convinced that taking a particular action would in fact *reduce their susceptibility* to the negative state or, if it occurs, would *reduce its severity*, and that the action would not involve overcoming overwhelmingly important (for them) *barriers* (e.g. cost, convenience, pain, embarrassment and so forth).

These two models are successfully applied in many health promotion interventions, and in healthy nutrition promotion in particular (Roe, Hunt, Bradshaw & Rayner, 1997). Constructs of the Health Belief Model are very closely associated with the movement through the Stages of Change (Werch & Clemente, 1994), and the combination of the two models has proved to be even more effective (Roe *et al.*, 1997).

Progressing through stages, and changing health beliefs, however, often involves something more than a mere provision of information from the educator. Nutrition is an

example of such a situation. Several health organizations recently stressed that enhancing people's nutrition can have a crucial role in decreasing the incidence of illnesses such as cardiovascular disease or cancer, and encouraged the promotion of healthy dietary habits (US Department of Health and Human Services, 1991; Department of Health, 1992). However, changing people's diet has turned out to be a hard challenge, requiring several barriers to overcome. A number of studies tried to analyse this phenomenon. Among others, Fries and Croyle (1993) showed that stereotypes influence people's reaction to nutrition education. For example, a sample of college students interviewed described people eating low-fat food as *picky*, whereas people with a high-fat diet were pictured as *easygoing* and *more likely to attend parties*. As a consequence, students who held a negative image of people eating a low-fat diet, were more sceptical when presented with information about the benefit of such a diet. Similar stereotypes can be found in a survey by Sadalla and Burroughs (1981).

Another study by Ziebland, Thorogood, Yudkin, Jones and Coutter (1998) identified some of the barriers which seem to prevent people from changing their diet, covering a range of factors both internal, such as enjoyment or lack of willpower, and external, such as lack of money. Internal barriers (such as *I enjoy eating foods that are bad for me*) were most often selected by those interviewed, and people who only selected internal barriers were less likely to make positive changes.

Such a situation makes the task of a nutrition educator particularly difficult. Fries and Croyle (1993) claim that this kind of situation is similar to that faced in advertising. Similar argumentative skills may then be of crucial help. This suggests that, in addition to behaviour change theories, a theory is also needed to represent every day arguments and the way they are used to change opinions and values.

3. A theory of informal argumentation

In looking for a theory of argumentation which could help us build our nutrition advisor, we focused on *informal* argumentation theories (Walton, 1989) rather than those establishing mere logical relations among claims. We were especially interested in theories in which stress was given to the audience's perception of the argument, rather than to the soundness of the argument itself, and therefore took into account values and opinions, in addition to facts. This requirement excluded some models, such as Toulmin (1958), as inappropriate, being based on deductive rather than dialectic argument and concerning primarily the basic structure of arguments, and not how the detailed content of such arguments can be selected.

As we were primarily interested in the generation of arguments rather than their evaluation, we looked for a theory providing us with guidelines that, with not much effort, could be used in a planning process, by determining when a particular argument is useful and how it should be developed. Text generation systems which rely on a model of discourse coherence are very common, and perhaps the most widely used model in this respect is the Rhetorical Structure Theory (RST) (Mann & Thompson, 1988). While useful and successfully applied in the generation of explanatory texts and tutoring dialogues [among others, see Hovy (1993) and Moore and Paris (1993)], RST says, however, little about how to generate persuasive arguments, which was our primary interest.

We found a more suitable theory in the work by Chaim Perelman and Lucie Olbrechts-Tyteca. The two Belgian philosophers developed a theory, the *New Rhetoric*, aimed at identifying “discursive techniques allowing us to induce or to increase the mind’s adherence to the thesis presented for its assent” (Perelman & Olbrechts-Tyteca, 1969, p. 4). Their point of departure was the observation that, in every day life, people rarely produce just a perfect logical proof of their claims, as it is “illusory” that facts speak for themselves, even when presenting scientific theories. Moreover, people do not rely on what they know when they argue with an opponent, but rather they try to justify their views by appealing to the values and opinions of the people to whom they are addressing.

The notion of *audience* is clearly essential in this context: the same argument can produce very different results when addressed to different kinds of people. The aim of the orator will then be to identify the characteristics of the audience to which he or she can appeal in order to be more effective.

The two philosophers’ study, rather than establishing an *a priori* “logic of value judgements”, investigated how humans argue, by collecting and classifying arguments that are successful in practice.

While relatively unknown in the Artificial Intelligence and Natural Language community, a part from a few notable exceptions such as Sillince (1994), the *New Rhetoric* is considered a milestone in Argumentation Theory (van Eemeren *et al.*, 1996) and is a much cited work among philosophers (Maneli, 1994) and, as a natural application, in the field of law (Golden & Pilotta, 1986).

This section will briefly explain the basic concepts of the *New Rhetoric*.

3.1. ARGUMENTATION IS BASED ON PREMISES

The arguer’s discourse is commonly based on a set of premises, or points of departure for the argumentation. The more the premises are shared by the audience, the more successful the argument will tend to be. The *New Rhetoric*’s notion of premises takes into account not only facts, but also the importance the audience attaches to facts, providing the following classification:

Premises relating to the “real” consisting of the following statements.

- *Facts*: statements about reality which can be assumed to require no further justification (for instance *Apples are fruits*).
- *Truths*: more complex systems of connections between facts (e.g. scientific theories).
- *Presumptions*: opinions that the audience is likely to accept without a proof, but adherence to them may need to be reinforced (e.g. *Fruit is a healthy food*).

Premises relating to the “preferable” consisting of the following statements.

- *Values*: statements related to the preference of a particular audience for one thing as opposed to another (for instance, *a healthy body is very important* or *natural products have to be preferred*).
- *Hierarchies*: the ways in which an audience arranges its values.
- *Loci*: very general statements which serve to justify values or hierarchies, for instance, the *Locus of Quantity* (a thing is worth more than another for quantitative reasons)

justifies values such as *Duration* or *Stability*, in contrast with the *Locus of Quality* which justifies values such as *Beauty* or *Goodness*.

3.2. ARGUMENTATION CONSISTS OF SCHEMAS

The force of an argument lies not only on the acceptability of its premises, but also on how the arguer connects premises and conclusions. The New Rhetoric is a collection of a number of *schemas*, that is ways to arrange premises and claims that are successfully used by people in ordinary discourse, ordered with respect to two wide classes:

Argumentation by association is applied when two elements that are separate in the mind of the addressee are put together by the arguer by using a particular relation. Different links give rise to different styles of schema.

1. *Quasi-logical argumentation*, using links that “give the impression” that the argument presented has a logical connection between its sub-parts. This is made by using an exposition structure which resembles a logical or a mathematical proof, such as: *You said that you only buy cheap things, apples are cheap, so you should buy apples!* It should be stressed here that the logical connection may be only apparent, and that is the reason for the *quasi* connotation. In a perfectly logical argument, the consequence *must* follow from the premises (e.g. *All cats are mammals, lions are cats, therefore lions are mammals*), whereas in the example above the conclusion does not necessarily hold (the addressee may dislike apples!). According to Perelman and Olbrechts-Tyteca, using elements of ordinary language in a logical relation makes the argument seem stronger, as the soundness of logical and mathematical thinking is well acknowledged.
2. *Argumentation based on the structure of the reality*, using links about the reality that the audience perceives as real, such as cause/effect, object/attribute; group/constituents; means/end, etc. The arguer, by expressing the relation, hopes to pass the audiences’ acceptance of one statement to the other, for example, *Eating fruit helps slimming* encourages the audience to apply their (positive) view of slimming to that of eating fruit.
3. *Argumentation establishing the structure of the reality*, using *new* (for the audience) links between claims, such as by giving an example (*Fruit is very sweet: consider pineapples*), or by appealing to a model, especially to promote an action (*Healthy people have fruit for breakfast*).

Argumentation by dissociation is a more subtle way of presenting an argument: a division is introduced into a concept that the audience considered as a whole. It should not be confused with the opposition to association (the refusal to link two separate elements): in dissociation the audience is presented with just one element, which is then decomposed into two separate elements. For example: *You said that people who are concerned about diet are self-centred, but I prefer to consider them just responsible persons*. In this case a concept (*responsible*) is differentiated from another the audience considered as a whole (*self-centred*). Dissociation is usually done by introducing new terminology, so it is a “creative” process.

Argumentative schemas	Description	Example
<i>Quasi-logical</i>		
— Incompatibility	X and Y cannot coexist	You can either go out with your friends or see me tonight.
— Definition	Identity of description of X	To call this a <i>warm</i> day it would have to be at least 25 degrees!
— Tautology	$X = X$	A father is always a father.
— Reciprocity	X and Y should be given the same treatment	We all love those who love us.
— Transitivity	aRb bRc , therefore aRc	Our friends' friends are our friends.
— Part/whole	the properties of a set apply to its members	Tell me what sort of friends you have and I will tell you who you are.
— Whole/part	the properties of a set member apply to the set	His behaviour disgraced all his family.
— Comparison	X is smaller/taller, etc., than Y	You have won because you are more experienced.
— Probability	X is very likely to happen	Don't stop here: this car park is always full at this time.
<i>Based on the structure of reality</i>		
— Causality	X causes Y	He stole because he was hungry.
— Pragmatic	Y is a consequence of act X	Eating less fat helps slimming.
— Ends and means	X is a means to Y	More competition would let the prices fall.
— Waste	Opportunities and efforts should not be wasted	You have started this job, now you have to finish it.
— Direction	X will necessarily lead to Y	Once a cheater, always a cheater.
— Persons and acts	X is Y because he did Z	He is a miser: he didn't tip the waitress.
<i>Establishing the structure of reality</i>		
— Example	Establishing a rule	Women are great interviewers: think of Oriana Fallaci.
— Illustration	Reinforcing/explaining a rule	You should help your friends: when one of them is in trouble, stand with him.
— (Anti) Model	X is a person (not) to imitate	Italians know everything about healthy eating.
— Analogy	B is to A as C is to D	A life without love is like a year without spring.
<i>Dissociation</i>	Term I vs. Term II	You talk of law, I talk of justice.

FIGURE 1. Examples of New Rhetoric's schemas.

Examples of schemas are given in Figure 1. The classification, proposed in the Perelman and Olbrechts-Tyteca original essay as a collection of practical examples, together with philosophical considerations, was subsequently elaborated by Warnick and Kline (1992). Their aim was abstracting a collection of coding guidelines which could identify “as precisely as possible” the attributes of each schema, in order to make its

recognition (and reproduction) easier. This collection of attributes greatly helped us in our formalization.

4. Experimental evaluation of the theories

An experiment was carried out to test whether the combination of the three theories we assumed were needed in our formalization of an arguing advisor, the Stage of Change model, the Health Belief model and the New Rhetoric, was both useful and acceptable to health practitioners.

The experiment involved five researchers from the Department of Human Nutrition of the University of Glasgow. The theoretical aim of the experiment was to establish whether real nutritionists explicitly use any model of behaviour change and, more importantly, whether an argumentative style of discourse is used, and could be analysed in terms of the New Rhetoric constructs. Another, more practical, aim of this experiment was to assess whether an overall structure of a nutritionist's session could be identified, together with collecting a list of specific pieces of advice and information to arm our system with.

The experiment was carried out via email. After being told the aims of the project, the nutritionists were asked to engage in a conversation via email with a fictitious "user". Users' responses were constructed by the researcher based on five different "characters". Nutritionists were not constrained in the length or style of their replies, as the aim of the experiment was to keep the dialogues as natural as possible.

From time to time, the nutritionists were also asked to comment on stages of the dialogue, explaining why they posed a particular question or whether they had an overall plan and so on.

These dialogues were analysed in terms of both the behavioural models and the argumentative techniques. All of the nutritionists used a stages of change approach to the conversation (two of them explicitly stated it): for instance, "users" in contemplation or action stages were mainly given tips about how to carry on, whereas for the user in the precontemplation stage the nutritionist concentrated on underlining the possible hazards of bad nutrition behaviour, and the user in the relapse stage was encouraged to think of the benefit of the habit she lost. Also, from the argumentative perspective, the nutritionists tended to focus on the user's values rather than giving mere information, for instance by stressing the quickness of some ways of cooking (e.g. grilling) rather than their healthiness for users who claimed to have time constraints, and so on. Moreover, all nutritionists' messages could be reduced to one of the New Rhetoric's schemas, confirming that this theory provides useful tools to analyse argumentative dialogues in this domain.

5. Daphne: dialectical argumentation for providing healthy nutrition education

This section and the following will describe our model of a dialectic advisor, Daphne, in the domain of healthy nutrition. We first describe our agent's "mind", in terms of its knowledge representation language, and its beliefs and attitudes. In the next section, we then describe how this mind fits in a general "body", an architecture for producing dialectic advice.

5.1. REPRESENTATION OF DAPHNE’S KNOWLEDGE BASE

In order to reason about values and opinions, as well as facts of the domain, a knowledge representation language was defined. This language had to be able to handle the concepts in the New Rhetoric theory described in Section 3, the concepts specific to the health promotion model we use (see Section 2), and basic nutrition information. The language we defined consists of four classes.

1. The knowledge about the specific domain forms a *nutrition knowledge base* containing classifications and instances of foods and basic principles of human nutrition. It was generated by collecting and encoding excerpts from the Manual of Nutrition by the Ministry of Agriculture, Fisheries and Food (MAFF) (1995) plus several leaflets and booklets distributed by the Health Education Board for Scotland and the MAFF.
2. *A language describing behavioural changes* was defined, for representing the stages of change and the beliefs about health. It has as basic concepts *actions* (e.g. “eating fruit”), *states* (e.g. “having cancer”) and *individuals*. Several relations are defined among these concepts, inspired by the health promotion theories:

- $\left. \begin{matrix} \textit{Doing} \\ \textit{Not Doing} \end{matrix} \right\}$ an *action* can be $\left. \begin{matrix} \textit{effective} \\ \textit{useful} \\ \textit{essential} \\ \textit{irrelevant} \end{matrix} \right\}$ to $\left. \begin{matrix} \textit{achieve} \\ \textit{prevent} \\ \textit{maintain} \\ \textit{change} \\ \textit{weaken} \\ \textit{strengthen} \end{matrix} \right\}$ a *state*.

- An *individual* can be

$$\left. \begin{matrix} \textit{susceptible} \\ \textit{unsusceptible} \end{matrix} \right\}$$

to a *state*.

- An *individual* can be

$$\left. \begin{matrix} \textit{Precontemplator} \\ \textit{Contemplator} \\ \textit{InPreparation} \\ \textit{Active} \\ \textit{HasHabit} \end{matrix} \right\}$$

with respect to an *action*.

3. *A language for representing arguments* has also been defined, having as basic concepts *topics*, *values* (Good/Bad/Indifferent) and *perspectives* (e.g. QualityOfLife, Health). Again, relations hold among these concepts.
 - A *topic* can be considered from a certain *perspective*.
 - A *topic* can have a *value* from a certain *perspective*.
 - A triple $\{\textit{topic}, \textit{perspective}, \textit{value}\}$ can be more valuable than another triple $\{\textit{topic}, \textit{perspective}, \textit{value}\}$ (this encodes the New Rhetoric’s concept of *hierarchy of values*).

- A *perspective* can be more valuable than another *perspective* (this encodes the New Rhetoric's concept of *locus*).

A topic can be any of the concepts in the previous class, for instance, an action (*Eating Fruit*), a state (*Cancer*) or an instance of a relation (*Eating Fruit can Prevent Cancer*). We can then express sentences like the following.

- *Eating Fruit can be seen from the Health perspective.*
- *Cancer is Bad from the Health perspective.*
- *The fact that Eating Fruit can Prevent Cancer is Good from the Health perspective.*

An important restriction is that the attribution of a value to a topic has to be made via a perspective: so it cannot be said that cancer is good or bad *per se*, but that it is so from the health perspective. This, we believe, facilitates a more natural style of argumentation, in which the same topic can have simultaneously good and bad values with respect to different points of view.

4. Finally, a core of *speech acts* allow the two parties to communicate about the above concepts, for example, Assert, Ask, AskIf, AskWhy, Acknowledge, DontUnderstand, etc.

5.1.1. More about perspectives and values

Our definition of perspectives is very dynamic: every topic in our domain can be viewed from any number of different perspectives, and the association topic/perspective is not fixed *a priori*, but it is a matter of the agents' opinion, and can change during the conversation.

For instance, Figure 2 shows the snapshot of an agent's viewpoint during a dialogue about fruit at a certain stage of the conversation. Perspectives are shown in Courier font, and a bold arrow associates them with topics. Some associations are defined as a default (e.g. cancer and stroke can be seen from the health perspective). Others derive from the conversation. In the figure, for example, it appears that the agent thinks that having a good self-image concerns one's social life (arrow linking Social life with GoodSelfImage). Whereas, the fact that eating fruit cannot be done easily at work influences one's quality of life: in this case the topic associated with the perspective is not a single object (as Cancer) but a more complex proposition (arrow linking Quality of Life with a semantic link).

The perspective associated with a topic is considered "passed" to all the other topics which are linked to it, in a backward fashion. For instance, as Eating Fruit is eventually linked to Cancer, the agent can say that Eating Fruit can be seen from the health perspective.

Similarly, the attribution of a value is not fixed *a priori*. In Figure 2 the value attributions are indicated with a good or a bad on the perspective arrows. The same way as perspectives, values can be passed up through the links. For example, in Figure 2, as stroke has a negative value from the health perspective, Eating Fat will have the same negative value from the health perspective, as it is "useful to achieve" a stroke. That is why Eating Fruit, which does not involve Eating Fat, has a positive value, again from the health perspective.

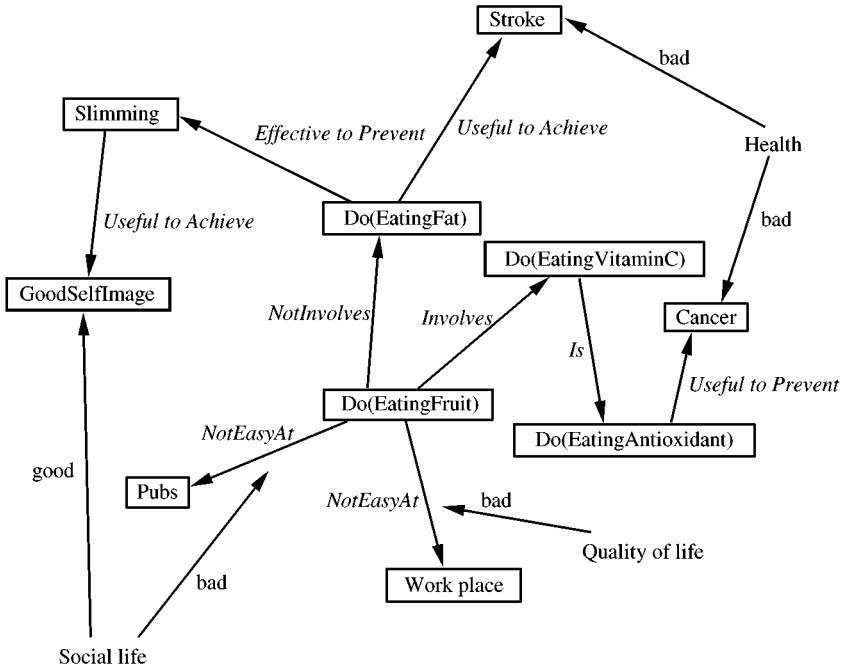


FIGURE 2. Perspectives and values associated with a topic.

With this mechanism, it might well happen that a topic has opposite values from the same perspective: in the example shown, Eating Fruit has a positive value from the social life perspective, as it helps improving one’s self-image, but also a negative value from the same perspective, as it cannot be done at pubs.

The particular form used for expressing values allows Daphne to conceive worlds in which an agent can have mixed feelings about a topic, without the world being inconsistent. In the example, the two beliefs about eating fruit are represented in the system as follows.

- HasVal(EatingFruit, SocialLife, Good).
- HasVal(EatingFruit, SocialLife, Bad).

These are not considered as opposite (nor related), so can be both held by the same agent at the same time, each with its own justifications.

5.2. DAPHNE AS A PARTIALLY COOPERATIVE AGENT

To successfully interact, in addition to her† own knowledge, an agent needs to explicitly represent her opponent’s set of beliefs. Representing also her opponent’s assumptions about her own beliefs can be necessary too. This requirement can be extended indefinitely, and, in principle, an unlimited number of nested levels may be expressed: an agent

† In what follows, we arbitrarily refer to the agent we are modelling as *she* and to her opponent as *he*.

may reason about her own beliefs (level 0), about what the other agent believes (level 1), about what the other agent believes that she believes (level 2) and so on.

It is possible to represent virtually every nesting level, for instance in Ballim (1992) an ascribing mechanism is used to dynamically create new nesting classes. This approach is especially suitable when modelling situations in which misconceptions and deceitful behaviour are possible (Lee & Wilks, 1997).

However, in most practical systems which use nested beliefs, the actual representation does not include every possible level, but only a fixed number, and conflates all the levels beyond these into a unique class representing *mutual beliefs*. Mutual beliefs are especially necessary, as Blanford (1993) noticed, when dealing with domains, like ours, in which claims are not necessarily right or wrong. The level of accuracy of the representation, that is the number of nesting levels actually represented in a model, determines how sophisticated the interaction can be.

In a situation of cooperative dialogue, agents are only interested in communicating what they believe, and assume their opponent does the same. This is the simplest communicative situation, and it has been argued (Kobsa, 1990; Taylor, Carletta & Mellish, 1995) that only three levels of nesting are sufficient to represent it.

When deception is allowed, or it is needed to “plan for plan recognition”, that is to figure out what the other agent’s plan is, deeper nesting levels are needed. The two mentioned studies maintain, however, that even in those cases only a few levels of nesting need to be added.

But what are the characteristics of our arguing agent? Can she be defined as cooperative?

It should be noticed that some sort of deception is in a sense inherent in the definition of dialectic argumentation. The fact that the arguer bases her justification on the audience’s beliefs and not on what she really thinks on the matter can be seen as a subtle kind of deception. In particular, the arguer may like a claim to pass as her own even if it is only a projection of the audience’s mind (for example the claim *Eating fruit is important because it helps slimming* can be made by an arguer who knows the audience gives a high value to being slim, even if she does not).

We wanted our agent to show this kind of behaviour, but at the same time we did not want to add any extra complexity, which would have been needed in order to deal with deception and lack of trust. We therefore introduced only a limited scope for lie, by allowing our agent to deceive only about her own (level 0) beliefs. Therefore, it is not permitted to be unsincere when speaking about “inner” beliefs (that is beliefs in levels 1, 2, 3, etc.).

Following the approach of Kobsa (1990) and Taylor *et al.* (1995), we defined a fixed number of basic belief sets. Given that the flouting of the cooperative principle is done only in a restricted sense, we maintain that three levels are still sufficient, as each agent will need to represent only one well-defined type of lie. However, we defined the third level, concerning mutual beliefs, in a more fine grained way, by distinguishing between the mutual beliefs about what the agent herself believes and those about what the other agent believes. This allows us to represent a situation in which, even after the two agents openly manifest their beliefs, a conflict is still present. The classes we defined are therefore.

Class 0: the agent’s own beliefs BA_1 (e.g. *I believe apples are fruit*).

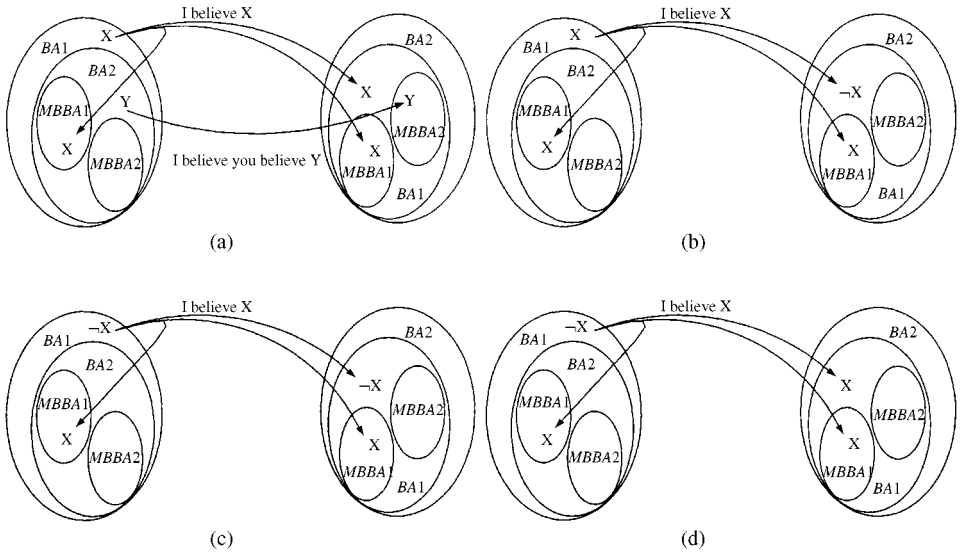


FIGURE 3. Communication types: (a) Sincere communication. (b) Sincere communication misinterpreted as insincere. (c) Insincere communication detected. (d) Insincere communication non-detected.

Class 1: the agent’s beliefs about the other agent’s own beliefs BA_1BA_2 (e.g. *I believe you believe apples are fruit*).

Class 2.1: the agent’s beliefs about what the other agent believes is mutually believed about the first agent’s own beliefs $BA_1BA_2MBA_1$ (e.g. *I believe you believe that we both believe I believe apples are fruit*). In other words, in this section the agent stores her “public image”, or at least what she thinks (hopes) this image is. Such an image might differ from what the agent actually believes, which is kept in BA_1 (class 0).

Class 2.2: the agent’s beliefs about what the other agent believes is mutually believed about the second agent’s beliefs $BA_1BA_2MBA_2$ (e.g. *I believe you believe that we both believe you believe apples are fruit*). In this space, the agent keeps the “public image” that the other agent (the user) gave of himself. For instance, all the beliefs that the other agent communicated to hold are stored here. Again, this image may contrast with what the agent actually thinks of her partner, which is kept in BA_1BA_2 (class 1).

We might think of the whole class 2 as the *public* appearance of the agents’ beliefs (or, better, what the agent thinks this public image is in her opponent’s mind), whereas classes 0 and 1 represent what the agent really thinks, her *private* beliefs.

Figure 3 represents how the various parts of the belief models of two agents are updated after different sorts of communication, showing that the four sets proposed are necessary and sufficient to represent the different situations.

When a communication is sincere, and is perceived as sincere, all the sections of the belief model of both agents contain the same piece of information (case a).

In cases in which either a misinterpretation or a lie occur, differences exist between the public and the private sections of both agents, where the public sections (classes 2.1 and 2.2) mirror the communication, and the private ones (classes 0 and 1) express the beliefs which are actually held.

- The speaker always assumes that the hearer will believe in what she said, otherwise she would not have said it,[†] therefore her public image of the belief is updated accordingly.
- The hearer always stores the communicated information in his public image of the speaker, to express the fact that he believes that the speaker is confident about the success of her communication. However, if he thinks that the speaker was lying, the opposite piece of information will be stored in his private section (cases b and c).

It should be noticed that case (a) is the only situation in which an inner belief other than the first level (i.e. a belief about others beliefs) is shown: the assumption is that agents can only lie about their private beliefs about themselves (class 0), and a communication about inner classes will always be considered honest by the addressee. Relaxing this constraint would require the representation of additional belief classes.

5.2.1. *Distinguishing facts and presumptions*

In what we have discussed so far, we have not taken into account the important distinction in the Perelman and Olbrechts-Tyteca theory between facts and presumptions[‡].

In our approach, the distinction is not made explicit, but it is implicitly rendered by considering the belief class to which a claim belongs.

An agent considers as facts, for the purpose of the argumentation, all the beliefs which are classified as public (classes 2.1 and 2.2); presumptions are then an agent's private beliefs about either the domain (class 0) or the opponent (class 1).

It may seem unusual that the system's own beliefs (for instance, system's knowledge about nutrition) are considered as just "presumptions", but we believe that this more realistically represents an argumentative situation. In most advice giving systems the user's beliefs are considered fewer than the system's beliefs, both in qualitative and in quantitative terms. The user can only know about those concepts which are known by the system, and, as the system is considered the *expert* in the conversation domain, every disagreement between the system's and the user's beliefs about a concept is attributed to user misconceptions. However, as Hustadt (1994) pointed out, this is not a good approach when the personal attitude towards the topic is an important subject of discussion.

In our model, the definition of facts and presumptions is given in terms of the considered argumentation theory, so it relies on how the claims are perceived by the interlocutor. The spirit of this choice is that the system only communicates its *opinions*, being prepared to give evidence for them, if necessary, or even to withdraw them. As the New Rhetoric observes, "what [the arguer] says is not "Gospel truth", he does not possess that authority [...]. He acknowledges that he must use persuasion, think of argument capable of acting on his interlocutor, show some concern for him and be interested in his state of mind" (Perelman & Olbrechts-Tyteca, 1969, p. 16).

The result of a such a differentiation between facts and presumptions is used in the planning module, as will be shown later, so that belonging to one class or another makes a given claim a stronger or a weaker premise for an argument.

[†] We ignore here more sophisticated styles of communication, such as irony, in which the speaker says false statements with the assumption that the hearer will recognize it (Lee & Wilks, 1997).

[‡] The New Rhetoric's concept of Truth, e.g. scientific theories, is not represented in this model.

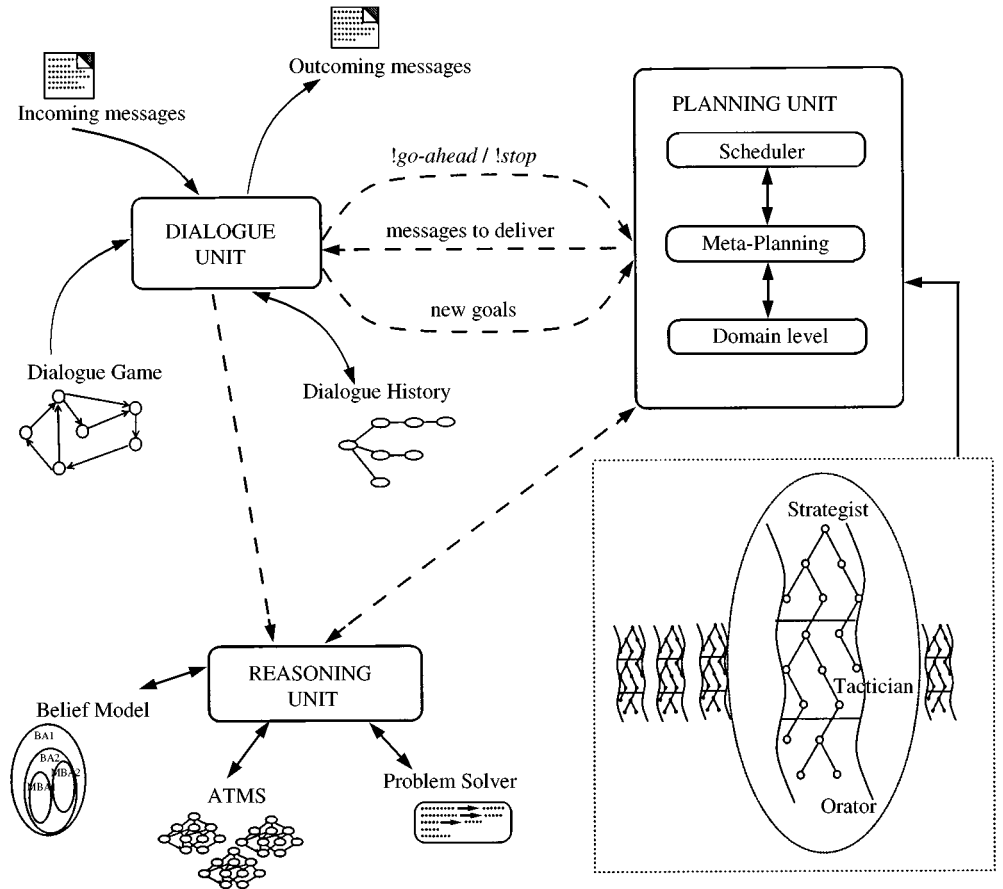


FIGURE 4. Daphne's architecture.

6. Daphne's architecture

The belief model presented in the previous sections constitutes the core of an architecture for an arguing agent, as shown in Figure 4.

The main components of the agent are the following.

- Reasoning unit responsible for managing the factual knowledge of the agent, about the domain and the opponent.
- Planning unit responsible for producing the piece of advice to be delivered to the user.
- Dialogue unit responsible for managing the actual communication between the two parties.
- Agent central unit whose aim is to achieve and coordinate all the others; there is no explicit mention of it in Figure 4, but its role can be represented by the dashed lines between the various components.

6.1. REASONING UNIT

The Reasoning Unit deals with the agent's beliefs about the domain and the opponent. It consists of a problem solver, coupled with a reason maintenance system to help deal with potentially withdrawable information.

The problem solver is based on a clausal form logic, but restricted to *Horn clauses*.[†] In addition to gaining efficiency in the problem-solving process (Dowling & Galliers, 1984), it has also been argued (Taylor, Carletta & Mellish, 1996) that representing justifications by means of Horn clauses, together with the use of limited depth of belief nesting, allows the reasoning process to be more "human-like". Humans do not easily deal with disjunctions of elements: for instance, the sentence *Either it rains or I will go out* may sound somewhat odd to many people.

However, humans make extensive use of negations, which are not considered valid in the Horn formalization: the phrase above is certainly less unusual if expressed as *If it doesn't rain, I will go out*. It therefore appears important to allow Daphne's opponent to give justifications using plain clause form, not just Horn clauses. Justifications given by the user are not used by the problem solver, so there is no need to change the problem solver's way of reasoning; but they are massively used by the reason maintenance module to determine the set of beliefs which are currently held.

The reason maintenance mechanism used is de Kleer's Assumption-based Truth Maintenance System (ATMS) (de Kleer, 1986a). In particular, the algorithm for label propagation used is the one described in de Kleer (1988). To deal with the opponent's general way of expressing justifications, an extension to the ATMS is used (de Kleer, 1986b) that can represent clauses containing negations, by introducing the concept of disjunction, or *Choose*, among assumptions.

6.2. PLANNING UNIT

The conversation in Daphne is the result of a planning process (Carletta, 1992; Cawsey, 1992; Moore and Paris, 1993; Chu-Carroll & Carberry, 1994; Taylor, 1994, Allen *et al.*, 1995).

Both Daphne's overall goals (e.g. argue about a topic, let the addressee progress from one stage of change to another with respect to an action, etc.), and the goals deriving from the discourse's obligations (e.g. answer questions) are solved by means of a planning mechanism, which decomposes them to eventually obtain a sequence of communicative actions. Several goals can be active at the same time, and are treated as an *agenda* of tasks to choose from.

6.2.1. Structure of the planner

Planning is hierarchical, both with respect to the organization of the planning operations at the domain level, and with respect to the planning process itself.

- A conceptual hierarchy is defined among the domain-level planning operators, by classifying them according to *abstraction spaces*. Distinct sets of abstraction spaces are

[†] That is, clauses formed by a conjunction of positive elements as premises and just one or no positive element as conclusion.

historically used (Sacerdoti, 1974, Washington, 1994; Fox & Long, 1995) in order for a problem to be first solved in a high-level space, and then refined at successively more detailed levels. Each space represents, in a sense, a different type of knowledge. The neat distinction among Daphne's tasks and sources of knowledge, which led us to identify different types of knowledge, as seen in Section 5.1, naturally suggested three abstraction spaces.

- The high-level space, or *strategist*, is concerned with the general goals of the system. In Daphne's case in particular, these goals concern the promotion of a healthy behaviour and embed the knowledge from both the Stages of Change and the Health Belief models. A different set of strategic operators can be 'plugged in', for instance, expressing tutoring strategies (Freedman, 1996).
- The middle-level space, or *tactician*, is concerned with the argumentative skills of the system, and embeds knowledge from the New Rhetoric domain. These skills are independent of the subject of the argumentation, so this set of operators would work equally well with a different strategic environment.
- The bottom-level space, or *orator*, has strictly linguistic knowledge: it is concerned with how to express, in natural language, the argument proposed by the tactician.
- A structural hierarchy is defined which divides the planning process into three *planning layers* (Stefik, 1981). At the bottom layer (Stefik's *domain space*) there are the actual planning operators related to the domain of the abstraction space currently under consideration. On the top of this, there is a *meta-planning* layer, whose aim is to control the applicability of domain operators, by checking prerequisites and executing goal decomposition. At the top of the hierarchy, a *scheduler* layer is responsible for controlling the overall process, by activating new goals, choosing among alternative operators, and so on.

6.2.2. Planning operators

Operators at any abstraction level have the following structure.

- *Goal*: the goal it aims to achieve; it can be either a belief to induce or a task to accomplish.
- *Prerequisites*: the conditions that have to be verified before the operator can be applied.
- *Decomposition*: the description of how the goal is decomposed into subgoals.

Some example operators are given in Figure 5. In the figure, the notation *?variable-name* represents a variable to instantiate, and the notation *!Constant Value* represents the constant whose value is Constant Value.

The first operator in the figure is a strategic operator: to encourage the addressee to pass from the Precontemplator to the Contemplator stage of change with respect to the action *?act*, a possibility is to argue pragmatically about the action.

Five tactician operators follow. The second operator in Figure 5 represents a pragmatic argument in favour of the action *?act*. The decomposition for this particular operator consists in looking for a state (*?state*) the action is useful to prevent, which has a constant value (*!Bad*) from a certain perspective (*?perspective*). Other possible Pragmatic Argue operators involve finding, for instance, a positive state (*!Good*) the action helps achieve, a negative state the action helps change, etc.

<i>Goal</i>	PrecToCont (Do(?act))	} Strategist's operator to try and let the addressee to progress from the Precontemplator to the Contemplator stage with respect to an action.
<i>Prerequisites</i>	B_3B_U (Precontemplator(Do(?act)))	
<i>Decomposition</i>	PragmaticArgue(Do(?act))	
<i>Goal</i>	PragmaticArgue(Do(?act))	} Tactician's operators realizing the New Rhetoric's schemas: <i>Pragmatic Argumentation</i> , <i>Argumentation by Model</i> and <i>Argumentation by Dissociation</i> .
<i>Prerequisites</i>	B_3 (Useful(Do(?act), Prevent(?state)))& B_3B_U (HasVal(?state, ?perspective, !Bad))	
<i>Decomposition</i>	Claim(HasVal(Do(?act), ?perspective, !Good))& Support(Useful(Do(?act), Prevent(?state)))	
<i>Goal</i>	ArgueByModel(Do(?act), ?perspective)	} General tactician operators.
<i>Prerequisites</i>	B_3 (HasHabit(Do(?act), ?individual))& $B_3B_U M B_U$ (Model(?individual, ?perspective))	
<i>Decomposition</i>	Claim(Habit(Do(?act), ?individual))	
<i>Goal</i>	Dissociate(?topic, ?perspective)	} General tactician operators.
<i>Prerequisites</i>	$B_3B_U M B_U$ (HasVal(?topic, ?perspective, !Bad))& $B_3B_U M B_U$ (Is(?subtopic1, ?topic))& B_3B_U (Is(?subtopic2, ?topic))& B_3B_U (HasVal(?subtopic2, ?perspective, !Good))	
<i>Decomposition</i>	Concede(HasVal(?subtopic1, ?perspective, !Bad))& Distance(?subtopic2, ?subtopic1)	
<i>Goal</i>	B_3B_U (HasVal(Do(?act), ?perspective, !Good))	} General tactician operators.
<i>Prerequisites</i>	<i>N I L</i>	
<i>Decomposition</i>	ArgueByModel(Do(?act), ?perspective)	
<i>Goal</i>	Promote(Do(?act))	} General tactician operators.
<i>Prerequisites</i>	<i>N I L</i>	
<i>Decomposition</i>	PragmaticArgue(Do(?act))	
<i>Goal</i>	Claim(?topic)	} Orator's operator to claim a topic.
<i>Prerequisites</i>	<i>N I L</i>	
<i>Decomposition</i>	Assert(?topic)	

FIGURE 5. Some planning operators.

The third operator represents an argument establishing the structure of reality: argumentation by model. An action is promoted by claiming that it is a habit of a person who represents a model for the listener (from a certain perspective). This operator illustrates how we use the notions of fact and presumption, as we anticipated in Section 5.2.1: that an individual is a model has to be a *fact* for the listener (it has to belong to the mutual section of the arguer's beliefs), otherwise the strength of the argument will be weakened.

In the fourth example, an argument by dissociation is presented. The prerequisites state that it is a fact (mutual section) that the given topic is Bad from a given perspective. The operator therefore looks for two subtopics of the given topic, such that one of them is Good from the same perspective, in order to differentiate them. One of the subtopics (?subtopic1) should be easily recognizable (e.g. it is a fact) by the addressee as an instance of the given topic, so that the negative value can be inherited from the general instance. On the other hand, the second subtopic can even be unknown to the addressee, and the

tactician can plan for making it known and showing that it is valuable, if the prerequisites concerning it are not met.

The fifth and sixth example show two more general tactician operators: in the former, the goal is a belief to induce, rather than a task resulting from subgoal decomposition. This is the most general type of tactic operator, and there is one of them for each argumentative technique implemented. The tactician will typically use them whenever it has not received a precise strategic suggestion. The latter shows a decomposition for the general goal to promote an action (by means of a pragmatic argumentation).

Finally, the last operator in Figure 5 is taken from the orator abstraction level, and states that in order to claim a certain topic it is sufficient to Assert it, where Assert is one of the basic speech acts.

Examples of the application of these operators are given in Section 7.

6.2.3. *Planning algorithm*

The planning process is incremental, as execution and planning are interleaved. This approach, first envisaged by Sacerdoti (1974) and thoroughly developed by Ambros Ingerson (1987) for general planners, is particularly useful in the planning of dialogue moves, as the speaker does not want to commit herself too much before knowing her partner's reaction (Cawsey, 1992; Freedman, 1996). Moreover, speaking without planning too much ahead makes the system's behaviour more humanlike. Interleaving planning and execution may cause a plan to be abandoned even if not completely executed, if new circumstances arise. This could mean for the system to interrupt a discussion about a topic without having finished to make its point, if the new circumstance leads to more urgent goals. The interrupted plan is then set aside, and, if its priority stays low, it may never be recovered, but, again, this makes Daphne's behaviour more similar to what real people do.

Goals are provided to the planner by different sources. The first, top-level goal derives from the system's dialogue aim, and is expressed by means of a strategic goal. In Daphne's case, it is typically concerned with moving the user one step forward in the Stages of Change chain with respect to some good nutrition practice. In another setting, for example tutoring, it could concern teaching a subject, and so on. Goals are also provided by the dialogue manager, for instance to fulfil a discourse obligation (e.g. answering to a user's question). Finally, goals can be self-produced by the planner as a result of the decomposition of goals in subgoals.

At each planning step, a goal is selected from the agenda, and, unless it can be directly executed (i.e. it is a primitive speech act), an attempt is made to plan for it. The appropriate operator library is considered in order to find an operator whose goal can be unified with the current goal. If the operator's prerequisites are not satisfied, the planner may try to plan for them as well. They will be put in the agenda, together with the subgoals in the decomposition slot, and the process will start again.

In Daphne, there is no explicit "replanning" or "repairing" to a plan. However, if the arguer is not satisfied with the efficacy of her argument (for instance, because a set of beliefs that should be true at a certain stage, are false), a further attempt can be made. This attempt involves a brand new planning task, even though, of course, the information gained in the previous attempts will be used. A history of the interaction is maintained, containing the goals attempted and the operator used, in order to prevent

the planner to repeat exactly an already used plan. However, the planner can use the same operator twice, with different instantiations.

6.2.4. Planning choice points

The planner has to perform a choice in two occasions: it has to choose the goal to accomplish when several goals are present in the agenda at the same time, and it has to choose the operator to apply when several operators are equally suitable to accomplish a goal.

Goals are kept in the agenda partially ordered with respect to several different priority scales. The goals proposed to fulfil system's high-level aims have a High value with respect to the *strategy* scale. The goal proposed by the dialogue manager have, on the other hand, a High value with respect to the *tactic* scale and a High or Medium value with respect to the *focus* scale, depending on whether they involve the reply to a user's question or not. The choice of which goal to activate first is made by means of metarules, and depends on how the system arranges the different scales. We can then run Daphne with a higher emphasis on the strategy or the tactic, or we can choose to let her be more or less cooperative, with respect to the discourse obligations, by giving the *focus* scale a higher or lower priority. At the moment, only these three scales are used, in addition to a default *decomposition order* scale, which ranks the subgoals of a given goal so that they are attempted in the order they appear in the decomposition of the operator used. Nevertheless, this mechanism allows the presence of an indefinite number of scales, as long as the appropriate metarules exist to evaluate the new goals.

If several operators are available to solve a goal, metarules are used again to choose among them. The set of metarules is specific to the operator library. This is reasonable if it is accepted that three different theories are expressed in the three libraries: the parameters influencing the choice of a strategy in the Stages of Change theory are necessarily different from those influencing the choice of an argumentative schema, or a linguistic construct. Currently, we only concentrated on designing guidelines for the tactician level, on the basis of the New Rhetoric's principles. For instance, operators based on "facts", that is whose prerequisites are public beliefs, are preferred to those based on "presumptions" (private beliefs). Or, operators mentioning more valuable perspectives (in the user's point of view) are preferred to the others. In the other two operator libraries, or any time there are equivalent rankings, the choice is made at random.

6.3. DIALOGUE UNIT

As the emphasis of the system is on dialectics, dialogue ability is of crucial importance and it should show the characteristics that would be reasonably expected in a discussion between two arguers. However, it is beyond the scope of this work to take into account more complex issues about natural language dialogues. Therefore, some assumptions are made.

- *Non ambiguity*: the two participants must agree on the meaning of the concepts being communicated.
- *Limited set of utterances*: specific communication acts are defined, and utterances including other types of communication acts will not be considered.

- *Explicit back reference*: whenever an agent wants to refer back to a previous dialogue move (e.g. to answer a question or to express disagreement), the previous move has to be explicitly mentioned.

These assumptions permit us to put the problem of understanding the incoming messages aside for the time being. In fact, the system's input and output consists at the moment of sentences in a first-order language.

The dialogue manager module is completely independent from the others. Its tasks are to keep track of the conversation and to ensure that some coherence is maintained, according to a defined set of rules (e.g. a question should be followed by a reply, etc.). However, it is not concerned with the *content* of each utterance, and bases its behaviour only on syntactic rules involving the different speech acts. This contrasts with other approaches, such as the one by McRoy, Ali and Haller (1998), which encodes both syntactic and conceptual information in the same structure.

A consolidated approach to dialogue modelling is inspired by the metaphor of a game (Carlson, 1983): the two dialoguing partners are viewed as players, each of them having a certain number of allowed moves in every situation of the game. Most commonly (Carletta, 1992; Kowtko, Isard & Doherty, 1992; Pilkington, Hartley, Hintze & Moore, 1992), different games are defined for each possible type of interaction (e.g. Inform game, Wh-ask game, etc.), with the constraint that a game has to be completed before passing to another one, or at least the partner has to explicitly agree to start a new game, nested in the current one.

We follow Burton and Brna (1996) in taking a different approach, modelling the whole dialogue as a single, large game. In this way, all the possible theoretical games are interconnected, and the players can pass from one "game" to another. This easily models, for instance, the behaviour of a non-cooperative agent which avoids answering a question.

Daphne's games can be represented as a state transition diagram, in which each move is represented by an arc connecting two states.† Similar constraints on turn sequencing, whether represented as a network or as a table, are widely used in conversational systems [e.g. see Frohlich & Luff (1990) who propose a general table of turn sequence constraints based on human conversation].

An example of one possible game is shown in Figure 6. In the figure, the two participants are identified as first and second speaker. States associated with the first speaker are identified with a solid box, those associated with the second speaker with a dashed box.

The games are defined by the following objects.

- The set of *allowed states* for each participant, some of which are identified as *start states*. The shadowed boxes in the figure represent valid starting states for the first speaker.
- The sets of *allowed moves*. Each state is associated with three sets.

† In what follows, unlike the usual notation, the actual move performed by the agents is associated with the state rather than with the arc leading to it. So a "Y/N" state corresponds to the state which the system is in after one of the agents performs a *yes* or a *no* speech act. Similarly, when speaking of, for example, an "assert move", this corresponds to the crossing of the arc which, from the current state, leads to the "assert" state for the same speaker. The convention permits a simplified encoding of the state diagram, avoiding superfluous labels on the arcs.

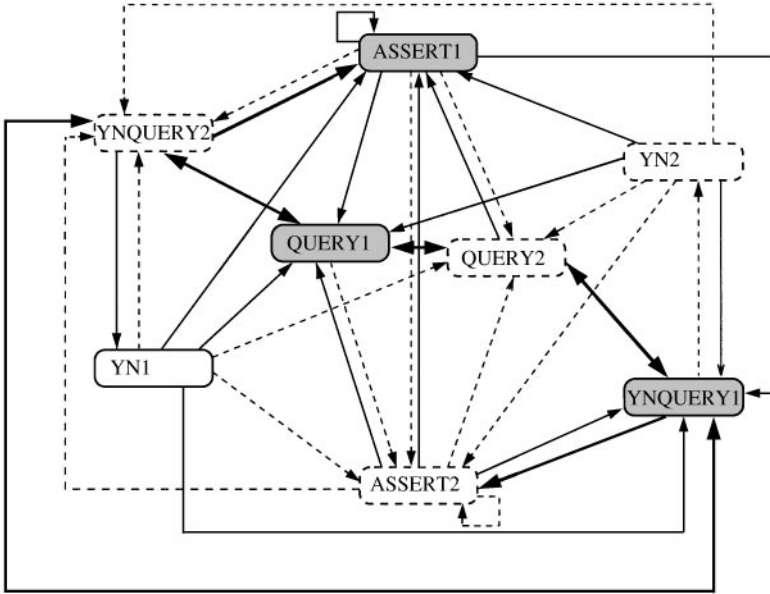


FIGURE 6. A dialogue game. \longrightarrow 1st speaker moves \square 1st speaker states \dashrightarrow 2nd speaker moves \square 2nd speaker states \longrightarrow changing game moves \square start states

1. Set of moves which the current speaker is allowed to make to keep on talking; in Figure 6 they are represented by arcs having the same type as the state they depart from.
2. Set of moves which the current speaker is allowed to make, on taking the turn, to follow-up the previous speaker's game; in Figure 6 are represented by arcs having the opposite type to the state they depart from.
3. Set of moves the current speaker can perform to change game; in Figure 6 they are represented by bold arcs.

Figure 6 represents the game currently used in our system. Four states are considered for each of the participants, having the following characteristics.

Assert: including all types of positive utterances an agent can perform (e.g. Inform, Reply, Acknowledge, etc.).

Query: including all types of request that can be made to the opponent apart from yes/no question (e.g. Why, Ask, DontUnderstand, etc.).

YN query: corresponding to yes/no questions.

YN: corresponding to the answer to a YN-question.

A characteristic of this game, as can be noticed from Figure 6, is that the game is perfectly symmetrical with respect to the two participants. The players are allowed to avoid answering questions (bold arcs after a query state), and more than one move can be made by the same participant before passing its turn, but players are constrained to pose

only one question per turn. To simplify the transition network, the question must be the very last move of a turn.

The concepts of “allowed”, “changing game” and, by exclusion, “forbidden” moves permit the Dialogue Manager to keep track of a sort of “syntactic” coherence of the dialogue, even though it is not concerned with the content of the sentences themselves. For instance, in the dialogue game in Figure 6, after an agent performs an “assert”, both a question and another assert can be admissible, and will preserve the coherence of the dialogue, whereas a “yes” or a “no” would be incoherent, and therefore are forbidden (there is no arc leading to them). Similarly, after a “Y/N query”, the normal and coherent move to perform is a “Y/N reply”, whereas any other move would represent a change in the game: it is permitted, but the Dialogue Manager keeps track that a “bold” arc has been traversed. Finally, after a “query”, the coherent way to reply is by means of an assertion; another query would be less coherent, but it is also allowed, but a “yes” or a “no” would be incoherent, and then forbidden.

The independence of the dialogue game definition from the rest of the modules allows us to use a variety of games, as long as the same formalism is used. For example, a game can be used that, as opposed to the game in Figure 6, is not perfectly symmetrical with respect to the two participants, or allows just one move per turn, etc.

The dialogue manager keeps track of a *dialogue history*, which is a stack of stacks. Each stack contains a thread of the dialogue, with a new stack initiated each time a bold arc is traversed. The top-stack represents the current thread of the conversation, with the current turn on the top. Whenever one of the participants refers back to a move in a previous thread of the conversation, the corresponding stack is popped.

The dialogue manager reads the incoming messages, sending messages for updating to the belief revision module. It traverses the dialogue net, ensuring that the move is allowed and starting a new thread in the history when a bold arc is traversed. It also updates the proper dialogue thread when answers are received.

When the system holds the turn, the dialogue manager establishes, on the basis of the allowed action sets related to the state, whether the planner can go ahead in producing new utterances or the system should pass the turn to the user. Similarly, a message is produced when the opponent reaches a state from which no further move is allowed, and an error message, with the request to reformulate the utterance, is produced whenever the user performs a forbidden move.

By traversing the network, the Dialogue Manager also triggers some rules which generate *action proposals*, that is suggestions about goals that can be useful to follow up the conversation. Two types of action proposals can be produced.

1. Actions to fulfil discourse obligations; for instance, after a question has been posed, the answer to that question is given to the planner as a suggestion to follow up.
2. Argumentative actions: the dialogue manager can propose to attack some user’s claims.
 - For each claim related to the real (that does not mention any value), if the system holds the opposite claim, a goal to assert its disagreement is put forward.
 - For each claim related to the preferable (that attributes a value to a topic), a way to counterargue to it is looked for.

The action proposals compete with the other goals already present in the agent's agenda, for the planner to choose among.

6.4. DAPHNE'S CENTRAL UNIT

Having described the various modules composing Daphne's architecture, a description of the coordinating central unit can be given. This consists of a fixed cycle invoking the various components in turn.

1. Activate the dialogue manager in order to read the incoming messages and update the dialogue history.
2. Evaluate dialogue manager's new goal proposals and pass the appropriate ones to the planner.
3. Activate the planner, as soon as a go-ahead message is received from the dialogue manager.
4. Manage the belief revision and query requests from both the planner and the dialogue manager.
5. Activate the dialogue manager to output the set of messages received by the planner to be delivered to the addressee.

7. An annotated example

This section shows Daphne's behaviour with respect to the following short dialogue (system and user's turns are denoted with S and U, respectively, followed by the turn number). The dialogue is expressed here in English for the sake of clarity, even though, as previously said, the actual input and output of the system are statements in first order logic.

-
- S1:** *Have you considered eating more fruit?*
U1: *No.*
S2: *It is good for your health, as it helps prevent cancer.*
U2: *Fruit is boring! It's granny's stuff!*
S3: *May be pears are boring. But avocados are different! Did you know it's Californians' favourite?*
-

The first two turns are produced in a pre-argumentation phase, involving the assessment of the user's stage of change with respect to some "healthy nutrition" action, such as, in this example, the action of eating more fruit. A brief questionnaire to establish the stage of change is used, proposed by Prochaska, Norcross and Clemente (1994).

After this stage, the actual argumentation begins.

S2. The first goal the system automatically poses is to let the user move one step forward in the States of Change chain (or reinforce a habit if the user is already in the last stage). Here, the system is trying to move the user from the precontemplation to the contemplation stage with respect to eating fruit. There are several ways to pursue this goal, dictated by the Stages of Change model, such as describing the benefits of the action, warning about possible consequence of not performing the action and so on. The choice among different operators at this first stage of the conversation (and for the

strategic level in general) is made at random. Let us suppose that the system instantiates the strategic operator in Figure 5.

<i>Goal</i>	PrecToCont (Do(!EatingFruit))
<i>Prerequisites</i>	$B_S B_U$ (Precontemplator(Do(!EatingFruit)))
<i>Decomposition</i>	PragmaticArgue (Do(!EatingFruit))

The decomposition refers to a tactician operator, so the control is passed to the next abstraction level. It is worth noticing here that the strategic level forced the tactician level to promote the action by means of a pragmatic argumentation, and this is because of the particular knowledge (in the health promotion domain) the strategic level expresses. In principle, the strategic level can leave the choice of the way to promote a particular action (or support a claim) entirely to the tactician level. The localization of the health promotion knowledge in the strategic level would permit us to replace this level with another set of strategic operators, and reuse the tactics from the New Rhetoric theory in another domain.

This pragmatic argumentation can be done in several ways, as we anticipated before. For example, if the system's belief set is as shown in Figure 2, the system can argue in favour of eating fruit by emphasizing that it is useful to prevent cancer, if it can assume that the listener too will attribute a bad value to cancer from the health perspective. The pragmatic operator the system chooses is then.

<i>Goal</i>	PragmaticArgue(?act, ?state)
<i>Prerequisites</i>	$B_S B_U$ HasVal(?state, ?persp, !Bad) & B_S (Successful (?act, Prevent(?state)))
<i>Decomposition</i>	Claim (HasVal(?act, ?p, !Good)) & Support (Successful (?act, Prevent(?state)))

instantiated as

<i>Goal</i>	PragmaticArgue(!EatingFruit, !Cancer)
<i>Prerequisites</i>	$B_S B_U$ HasVal(!Cancer, !Health, !Bad) and B_S (Successful (!EatingFruit, Prevent(!Cancer)))
<i>Decomposition</i>	Claim (HasVal(!EatingFruit, !Health, !Good)) and Support (Successful (!EatingFruit, Prevent (!Cancer)))

The Claim and Support operators refer to the orator abstraction level, producing two speech acts. The planner stops its activity and the first act is passed to the dialogue manager, which actually transmits it to the user, and traverses the dialogue game in Figure 6 reaching the "Assert1" state. As another move is allowed to the speaker, the dialogue manager communicates to the planner that it can carry on with its activity, and the second speech act is performed. After that, as the planner has no other goal to achieve, the turn is passed to the user.

U2. The user makes two assertions: the dialogue game is then traversed, going to state Assert2 (twice). The communicated messages are, in practice, the following.

- Is(!EatingFruit, !Boring).

- Habit(!EatingFruit, !Granny).

We assume in this example that the system considers the user truthful.† These beliefs are then recorded both in the mutual section and the private section of its beliefs about the user [see Figure 3, case (a)].

1. $B_S B_U$ Is(!EatingFruit, !Boring).
2. $B_S B_U$ Habit(!EatingFruit, !Granny).
3. $B_S B_U M B_U$ Is(!EatingFruit, !Boring).
4. $B_S B_U M B_U$ Habit(!EatingFruit, !Granny).

As the whole argumentation is based on values rather than facts, Daphne is primarily interested in ascribing an evaluation to what the user says. Therefore, when the user does not explicitly mention a value in his response, the reasoning module tries to make some inferences with the new pieces information to this aim.

A set of rules is used, and one or more steps of forward reasoning are made, by unifying the premises of these rules with the new beliefs. The resulting new beliefs, with their justifications, are recorded in the ATMS.

In the example, with the assumption that boredom is not a positive emotion, the system can infer that the same value is passed to EatingFruit; it is also inferred that in user's hierarchy the Emotional perspective has a higher position than the Health one; moreover, as the user said that EatingFruit is a Granny's habit, it is inferred that Granny is an anti-model for the user from the Emotional perspective. As the inferences are done by the system, they are considered "presumptions" and are stored in its private belief section about the user.

More formally, from 1 to 4 and with the assumption

5. $B_S B_U$ HasVal(!Boring, !Emotional, !Bad).

The following rules are used.

R1: $B_S B_U$ Is(?y, ?x) & $B_S B_U$ HasVal(?x, ?persp, ?val) \rightarrow $B_S B_U$ HasVal(?y, ?persp, ?val).

R2: $(B_S B_U M B_S$ HasVal(?y, ?persp1, !Good)) $_{Turn(x)}$ & $(B_S B_U$ HasVal(?y, ?persp2, !Bad)) $_{Turn(x+1)}$ \rightarrow $B_S B_U$ MoreValuable(?persp2, ?persp1)‡

R3: $B_S B_U$ Habit(?person, ?act) & $B_S B_U$ HasVal(?act, ?persp, !Bad) \rightarrow $B_S B_U$ AntiModel(?person, ?persp)

and the new beliefs are as follows.

6. $B_S B_U$ HasVal (!EatingFruit, !Emotional, !Bad).
7. $B_S B_U$ MoreValuable (!Emotional, !Health).
8. $B_S B_U$ AntiModel (!Granny, !Emotional).

S3. The dialogue manager, after analysing the input message, looks for action proposals to communicate to the planner.

Let us suppose that in this example the system cannot find any disagreement with the user's claims related to the real, and concentrates only on the user's claims (or system's

† If this was not the case, the record of the messages would have been different, but the behaviour of the system would have probably been the same: the argumentation schemas always tend to exploit what the opponent says, to build up new arguments, rather than contradict him.

‡ This rule exploits the knowledge of "when" something was said: if the user, when Daphne positively evaluates a topic from a certain perspective, immediately gives the opposite evaluation of the same topic from a different perspective, a hierarchy in the perspectives is assumed. This is achieved with an additional slot, in the ATMS record of the belief, expressing the turn number in which the belief was added.

inferences from the user's claims) which involve the attribution of a value to a topic. Therefore, the system focuses on the sixth element in the list above: $B_S B_U$ HasVal (!EatingFruit, !Emotional, !Bad). There are three ways Daphne can attack a claim about the attribution of a value with respect to a perspective.

1. By claiming that the topic has the opposite value with respect to the same perspective.
2. By claiming that the topic has the opposite value with respect to another perspective.
3. By performing a dissociative argumentation on the topic, with respect to the same perspective.

The three goals the dialogue manager proposes to the planner are then.

1. $B_S B_U$ HasVal (!EatingFruit, !Emotional, !Good).
2. $B_S B_U$ HasVal (!EatingFruit, ?persp, !Good).
3. Dissociate (!EatingFruit, !Emotional).

This concludes the work of the dialogue manager as far as the analysis of the user's turn is concerned.

At the beginning of its new turn, the system realizes the user has not passed to the contemplation stage, as there still are no "good" values attributed to the action of EatingFruit, from any perspective. The goal:

4. "PrecToCont(!EatingFruit)"

is then passed again to the planner, which puts it in its agenda.†

The planner has then four new goals in its agenda: three proposed by the dialogue manager and one strategic goal. Let us suppose for the current example that the arrangement of the scales is: {*focus, tactic, strategy*}. Goals 1, 2 and 3 are then preferred to goal 4. They have the same evaluation, so the choice among the three of them is made at random. Let us suppose the system chooses to activate the Dissociate goal first.

The operator that will be chosen is

<i>Goal</i>	Dissociate (?act, ?persp)
<i>Prerequisites</i>	$B_S B_U M B_U$ HasVal(?act, ?persp, !Bad) & $B_S B_U M B_U$ Is(?act1, ?act) & $B_S B_U$ Is(?act2, ?act) & $B_S B_U$ HasVal(?act2, ?persp, !Good)
<i>Decomposition</i>	Concede (HasVal(?act1, ?persp, !Bad)) & Distance (?act2, ?act1) &

instantiated as

<i>Goal</i>	Dissociate (!EatingFruit, !Emotional)
<i>Prerequisites</i>	$B_S B_U M B_U$ HasVal(!EatingFruit, !Emotional, !Bad) & $B_S B_U M B_U$ Is(!EatingPears, !EatingFruit) &

†Remember that in Daphne there is no explicit replanning or repairing.

	$B_S B_U$ Is(!EatingAvocados, !EatingFruit) &
	$B_S B_U$ HasVal(?EatingAvocades, !Emotional, !Good)
<i>Decomposition</i>	Concede (HasVal(!EatingPears, !Emotional, !Bad)) &
	Distance (!EatingAcovados, !EatingPears)

However, let us suppose, that the system cannot assume that the user will believe that EatingAvocados has a good emotional value, that is the prerequisite is not satisfied and needs to be planned for. The prerequisites which are not satisfied, together with the decomposition of the fired operator, are put in the agenda. They all will have the same value with respect to the *focus*, *tactic* and *strategy* scales. They will only differ in the *decomposition order* ranking. Unlike many discourse planners (Cawsey, 1992; Moore and Paris, 1993; Chu-Carroll and Carberry, 1994), however, it is not required that the prerequisites are actually planned for *before* the operator decomposition. If such a constraint may be crucial for general purpose planners (a robot has to pick up the hammer before using it!), we believe it can be safely relaxed in discourse planners, especially when the text generated is as short as a dialogue turn. To add variety, then, the precedence between prerequisites and decomposition goals is chosen at random. Let us suppose, in the current example, that the decomposition goals receive a higher value in the *decomposition order* scale than the prerequisite. The system will then first try to accomplish the two goals of Concede and Distance (which will be immediately executed, being speech acts), and then the prerequisite will be considered. The general tactician operator in Figure 5 will be fired, as the ArgueByModel operator can be instantiated and fired.

<i>Goal</i>	$B_S B_U$ HasVal(!EatingAvocados, !Emotional, !Good)
<i>Prerequisites</i>	NIL
<i>Decomposition</i>	ArgueByModel (!EatingAvocados, !Emotional)

<i>Goal</i>	ArgueByModel (!EatingAvocados, !Emotional)
<i>Prerequisites</i>	B_S HasHabit (!EatingAvocados, !Californians) &
$B_S B_U MB_U$	Model(!Californians, !Emotional)
<i>Decomposition</i>	Claim Habit(!EatingAvocados, !Californians)

Notice again that the fact that Californians are a model for the user should be a “fact” (mutually known) otherwise the argument will be weak (unless, of course, the tactician is lucky enough to guess a right model for the opponent!).

8. Evaluation issues

A Daphne prototype has been implemented, in the ML programming language. The prototype fully possesses the designed reasoning and dialoguing abilities, but has no natural language skills. It instead produces its output and accepts its input in a simple first-order logic. By using this prototype, an initial evaluation phase has been conducted.

Two experiments were set up to establish how the user would like/accept the idea of an argumentative advisor, and also to collect a corpus of data to base further details of the system on (e.g. selection of schemas, possible perspective-topic associations, etc). The first

experiment was carried out by using an email environment: this allowed us to contact people, from virtually any place in the world, at basically no cost, and also to make the participants in the experiments as comfortable as possible, as they could decide when and how to reply, when to stop and so on. The second experiment was principally aimed at testing a possible interface for the system, and was conducted by using the WWW facilities.

8.1 EXPERIMENT 1: HOW REAL USERS BEHAVE

Some “users” were contacted via email and solicited to engage in a conversation about diet. People were recruited by posting a message on a mailing list with an interest in nutrition matters. The recruiting message explained that a new nutrition education computer system was being tested, but for the experiment a real person would in fact interact with the participants via email, simulating the behaviour of the system. However, no mention was made of the interest in the argumentative nature of the interaction, to avoid the participants paying too much attention to this aspects, and losing naturalness.

User’s responses and questions were translated in a format Daphne could understand (sometimes user’s messages needed to be abridged), Daphne’s responses were generated using the model described in this paper, and, after an elaboration in English, sent by hand to the user, again via email. Having such an intermediary between the user and the system allowed us to start the experiment before having completed the task of populating the knowledge bases and the operator libraries. In fact, such a task was performed “on the run”, by adding new pieces of information and different New Rhetoric schemas whenever needed. Moreover, when the model could not handle user responses, we could decide whether to enlarge the knowledge we first envisaged Daphne to hold, without the user experiencing any delay and dis-functioning, as might have been the case if the users directly interacted with the system. An email environment meant also that the participants were free to express their responses in any way they considered appropriate, to keep the dialogues natural and smooth.

The experiment lasted a month, involving 46 participants, with more than 500 email messages exchanged overall. The average “dialogue” length was 11 turns, with the longest dialogue consisting of 45 turns. Snippets of four of these dialogues are shown in Figure 7. A final questionnaire was sent to every participant to let them express their opinions about the experiment. The response can be considered positive: 81% of the interviewed enjoyed participating in the experiment, 55% found that the program made them more conscious about their diet, and the same percentage thought the system made them contemplate changing their habits. A large part of the sample (64%) liked the dialogue environment for getting advice, and even more (74%) liked getting tailored advice. There were mixed feelings about the usefulness of the messages received (51% found them useful and informative and 45% said they had not learnt anything new), but this is probably due to the nature of the *users*. Many of them had an interest in nutrition, and were therefore different from the target users we envisage for the system. Moreover, most of the users (71%) admittedly forgot very soon that they were supposed to pretend to use a computer system. Perhaps for this reason, a good percentage of the interviewed (42%) found that sometimes the messages received were somewhat *irritating*: in more extensive comments, many people said the system’s messages were too *imperious* and

Daphne	<i>Do you like cooking?</i>
User	<i>Not especially. [...] Cooking feels to me like a lot of effort for something (ie. eating) that?s over quite quickly. Also, I often feel tired at the end of a day?s work and don?t want to spend too much time in the kitchen.</i>
Daphne	<i>You do not cook just because you have to eat! Cooking can be a very relaxing and interesting activity, better than the watching TV!</i>
User	<i>I know you?re right but that still doesn?t make it easy to do!</i>
Daphne	<i>Have you ever considered having some fruit for breakfast or as a snack? (instead of chocolate)</i>
User	<i>I should do that, yes, I?ll have to go and buy some ...</i>
Daphne	<i>Don?t you have to go and buy chocolate as well?</i>
User	<i>I didn?t mean it would take an extra effort to buy fruit on my trips to the supermarket. However [...] it?s much easier to get hold of unhealthy snack food near work than it is to get fruit.</i>
User	<i>I do enjoy fruit, but it is not as appealing as say a biccie, and vegetables in summer aren?t really the go, I would eat vegies in the winter time at least 4 times a week.</i>
Daphne	<i>Maybe pears and apples are not that appealing, but what about satsumas, or a cup of strawberries, or a bunch of grapes? Nice and refreshing, especially in summer!</i>
User	<i>Yummy, I think if someone was to give me a plate of cut up fruit ie: like below then I would definitely eat it, it is always more appealing when it is all done for you.</i>
Daphne	<i>Good choice, but maybe you should increase your target! (of bread intake) I would advice at least 6 slices per day.</i>
User	<i>Gulp. It?s the fibre polis. Do you have any idea how long it would take to eat six slices of said bread?</i>

FIGURE 7. Extracts from Experiment 1 dialogues.

knowing (one of them said they were too *personal and criticising*) and some of them felt they were not listened to. It would be interesting to understand whether these negative feelings would have been triggered if the users were actually sitting in front of a computer system.

Curiously enough, the choice of an email setting for the experiment has revealed unforeseen advantages: many users liked the small diversion in their working day, and one found that *the daily nature of the game was good in providing some reinforcement and continuity to any diet changes I might have considered*. This encourages a more thorough investigation of the potential of email in this respect.

8.2. EXPERIMENT 2: "MEAL TIPS" ON THE WORLD WIDE WEB

In the second experiment, people had the chance to see some of Daphne's pre-generated dialogues in an "interactive" way on the World Wide Web.

Selected dialogues from the previous experiment were analysed by two of the nutritionists participating in the pilot experiment, who gave some additional comments on the email advisor's behaviour, as well as some alternative ways to answer back to users' counterarguments and questions. The whole set of practical tips and counterarguments collected was then used to set up a web site in which the user can go through the dialogues dynamically.

From the home page,† mirroring the email dialogues, after asking a little information about a typical meal, some tips are presented, related to the user's choice. The user is then

†<http://www.cee.hw.ac.uk/~alison/meals2/start.html>.

asked whether he or she agrees with the suggestion, and, if not, a list of possible counterarguments (objections to the suggestion) are proposed to choose among. These are obtained from the set of counterarguments to Daphne's suggestions that the user expressed in the first experiment. After the web visitor selects one counterargument from this list, the reply associated with the objection is presented, to which new agreement is sought, and so on. If the user cannot find in the list a suitable counterargument, it is possible to input a new one manually, which is stored for future use. Whenever the visitor wants to leave the web site, a questionnaire is presented to fill in about the system.

The system is implemented in Prolog, with small Perl and shell scripts providing the interface with the World Wide Web. HTML code is dynamically generated by the program in response to the user's selections.

The web site was advertised on the same mailing list used for the email experiment, as well as among people in the authors' Departments. Over a month, 152 sessions were collected, and 43 questionnaires were filled in.

In general, the most positive comments came from those concerned with health education themselves. The general view from the comments was that it might well be a useful approach for "ordinary" users, but most of those who tried it found it was too simple for themselves.

Many users did not explore the counterargument options very much, and so often missed getting basic information backing up the tips. This suggests a possible problem in dialogue systems that gives small contributions, then pass initiative back to the user.† However, the questionnaire results suggest that even for these users, many thought more about their diet as a consequence of using the system (67%), found some of the tips helpful (60%), and liked the approach (66%). And most (90%) found it easy to use with few navigation problems. Other statistical and technical details for this experiment can be found in Cawsey, Grasso & Jones (1999).

9. Related work

A full comparison with related research would be very lengthy considering the variety of issues that were considered in developing Daphne. Therefore, we present here a selection of works related only to some Daphne's interesting peculiarities, and we omit comparisons with a large number of works in text planning, discourse representation and planning, user modelling, and argumentation theory (Hovy, 1990; Cawsey, 1992; Kay, 1994; Traum and Allen, 1994; Buchanan, Moore, Forsythe, Carenini; Ohlsson and Banks, 1995; Moore, 1995; Stein and Maier, 1995; Freedman, 1996; Paolucci, Suthers and Weiner, 1996; Poesio and Traum, 1997).

9.1. BEHAVIOURAL THEORIES FOR HEALTH EDUCATION

The idea behind Daphne is not entirely novel. The Stages of Change Model has already attracted interest from the artificial intelligence community (Reiter, Cawsey, Osman & Roff, 1997). The possibility of combining it with argumentative techniques in the

† A similar situation occurred in the first experiment too: in many cases, if the system had not concluded its turn with a question, the user just assumed the experiment was over, and stopped replying back.

domain of health communication was also suggested by Marcu (1996), even though only at a conjectural level. Daphne's makes a step forward, and combines the models in a clear computational framework.

9.2. ARGUMENTATIVE SYSTEMS

Many systems have approached the problem of presenting arguments to the user. Most of them are principally focused on the structure of "valid" arguments (Fox, Krause & Elvang-Goransson, 1993; Karacapilidis, 1996; Vreeswijk, 1997). Others are mainly interested in the generation of arguments in natural language (Maybury, 1993; Elhadad, 1995; Read & Long, 1997b). Fewer are the cases in which the effectiveness of arguments is considered, as opposed to their validity: Zukerman, Korb and McConachy (1996) describe the architecture of a system capable of generating *nice* arguments tailored to the addressee, but the only concession to validity here is allowing an argument to have some steps missing in its logic chain. In PERSUADER (Sycara, 1990) assumptions are made which are similar to ours: the argument generation process is always guided by argumentation goals, changing the importance a person attaches to things, and argumentation strategies used to achieve them. The theoretical basis on which the strategies are defined, however, is not made explicit.

9.3. BELIEF REPRESENTATION

Daphne's method of modelling beliefs are those of Kobsa (1990) and Taylor *et al.* (1995). However, the classification of belief nestings in these approaches do not account for behaviour which is not completely honest.

The TRAINS system (Traum, 1994) also represents the belief nestings as separated belief modalities. An interesting feature in this system is the distinction between actual beliefs about the domain and *proposals*, which may be seen as beliefs which still need to be consolidated. Separating these two sets, insincere user utterances may be represented. The belief and proposal modalities are implemented as a tree of belief contexts, where the root context represents general (assumed) mutual beliefs, whose child is the shared beliefs' context, and where the mutually believed user and system proposals inherit directly from the shared context. Other belief spaces have their own context, inheriting from their ancestors.

Daphne has a different semantic for the concept of mutual belief. In fact, the usual notion of mutual beliefs (Clark & Carlson, 1982) is not represented, but rather the agents' assumption about what is mutually believed. This is, again, in the spirit of representing a more realistic argumentation scenario, in which the arguer cannot know in advance what is mutually believed, nor can be absolutely certain that something is mutually believed, even after a communication occurred.

9.4. USE OF PERSPECTIVES

A similar approach to perspectives and values was adopted in Karacapilidis, Trousse and Papadias (1997), where a broader *discussion domain* is associated with every viewpoint of a case. In that approach, however, the association is in a sense "top down" from the more general topics to their subtopics. We, conversely, also proceed in a "bottom up"

fashion. This may admittedly lead, as we move further and further away from the topic directly linked to the perspective, to associations which have little sense. But this is just the sort of “tricks” that dialectic and rhetoric argumentation involve.

A different use of perspectives is found in McCoy (1989): perspectives are used to dynamically highlight selected attributes of a knowledge base. In Daphne’s case, the knowledge base is always wholly highlighted in the dialogue, but a concept of perspective is superimposed to the knowledge base itself to attach an evaluation of the concepts from different points of view. Perspective are then not views (portions) of the knowledge base, but evaluations of knowledge base concepts.

Jameson *et al.* (1994) use a notion of perspective closer to ours, in a system to provide “biased” information to potential buyers of a used car. The bias consists in focusing user’s attention on particular aspects of the car which are thought to be of particular interest to the user. In this case too, however, the principle is to focus only on a subpart of the information available, but no value judgments are made.

9.5. PLANNING LEVELS

Various systems have used similar hierarchical planning frameworks for discourse planning and interpretation, but Daphne is unusual in the way it combines the use of these two hierarchies. Reed, Long and Fox (1996), for instance, decompose the task of planning argumentative texts in three subtasks.

- Choosing the argument structure (AS level).
- Refining the argument by employing stylistic techniques (EG or Eloquence Generator-level).
- Refining the interclausal structure, through the use of the Rhetorical Structure Theory (Mann, Matthiessen & Thompson, 1989).

Daphne isolates all the knowledge about argumentation techniques and tactics (Read and colleagues’ AS level and part of the EG level) in its tactician component, and builds on the top of this a strategic component whose aim is to decide “why argue in the first place”, by establishing the purposes of the argumentation on the basis of behavioural models of change. All the linguistic knowledge (EG and RST levels) is then concentrated in the orator.

Another similar architecture, but with the purpose of discourse understanding rather than generation, is due to Lambert & Carberry (1999). The authors propose a tripartite model of dialogue, in which three types of actions are distinguished.

- Domain actions, related to the goals of the participants (e.g. having a university degree).
- Problem-solving actions, related to the planning process itself (e.g. evaluating alternatives).
- Discourse actions, related to the actual dialogue, conveying or obtaining the information needed by the other levels’ actions.

Daphne’s conceptual hierarchy can be seen as a more fine-grained representation of Lambert and Carberry’s domain actions, in which the discourse level actions are also included (Daphne’s orator), whereas all the knowledge concerning planning and problem

solving is separated from the application domain and is concentrated in the meta-planning and scheduler layer, in accordance with Stefik's view of the planning process.

10. Conclusions and further work

This paper presented the formalization of an agent, Daphne, able to provide advice on a controversial subject by using dialectical argumentative tactics. These tactics are, we believe, general enough to be easily adopted in different contexts: the knowledge about the specific domain (nutrition) is concentrated in well-defined subparts of the system (the nutrition knowledge base) therefore the argumentative tactics can safely be expressed in terms of general "actions", "states" and "beliefs". Moreover, the argumentative tactics are also independent from the general goals of the whole dialogue, which are expressed in the strategic planner. New strategies can be easily added, e.g. for a tutoring dialogue, using the same argumentative tactics. The planner is associated with a dialogue manager able to be armed with different kinds of dialogue games: it would be easy, as an extreme example, to provide a dialogue game in which only one player is represented, in order for Daphne to produce monologues rather than dialogues.

The design of Daphne was guided by the principle of practicality and simplicity, and the current architecture proved to be suitable for our purposes. Nevertheless, we believe that some interesting parts of the design deserve a deeper investigation, which we will leave as further work.

There is currently no explicit natural language generation component in the system — the orator simply presents the utterances using Daphne's speech act formalism. We plan a more thorough investigation into how to produce more effective argumentative utterances. In particular, the applicability of the theory by Anscombe and Ducrot (1983) to our architecture is currently being studied.

A weak point in Daphne's argumentative style is that not much effort is put into understanding the opponent's responses. They are generally considered separately and used to infer attitudes and beliefs of the addressee, but there is no explicit plan recognition of the opponent's argumentative schema. Suggestions about the counter-moves to use are captured in the action proposals sets collected by the Dialogue Manager, but it would be more effective to try figuring out whether the opponent is using a particular argumentative schema, in order to find a counter-schema to respond with.

As to the user modelling issue, although it is not a major focus of this work, it is obvious that a system whose aim is to address its argument to a particular user has to have a fairly sophisticated model of this user. For the time being, Daphne only relies on a stereotypical set of beliefs and values that the average user can be assumed to have (e.g. apples are fruit, cancer is a severe disease, diseases are bad from the health perspective, etc.) and on the information inferred over the interaction. The possibility of augmenting the user model by considering, for instance, the user's goals, or preferred argumentation style (especially in view of a more sophisticated plan recognition mechanism) is worth a more thorough study.

Finally, some considerations about architectural choices can be made. As shown before, the agent's coordinating central unit consists of a fixed cycle invoking the various components in turn. For the time being, this arrangement has appeared to be practical and suitable. Nevertheless, the modularity of the components' design would allow us to

implement the whole agent as a blackboard structure, and this possibility should be investigated.

On the evaluation side, after having assessed the usefulness and interestingness of the system for health care, we can now concentrate on evaluating more technical aspects of the system, such as the coherence and effectiveness of the dialogue generated. Comparative studies of different strategies are being planned, in which the output of Daphne, parameterized with respect to different characteristics, can be rated in randomized trials.

In conclusion, Daphne represents a novel approach to “computational argumentation”, which espouses the need for expressing everyday arguments in an advice giving system, and the use of a consolidated theory of *informal* argumentation which combines logic and dialectic. This is integrated in a well-formalized framework including behavioural change strategies and techniques deriving from well established theories in the health promotion field.

The system is fully implemented, and has been successfully applied in a nutrition education scenario. The results from the evaluation phase seem to confirm the validity of the framework, and suggest that it is general enough to extend the system, and apply it to new scenarios.

References

- ALLEN, J., SCHUBERT, L., FERGUSON, G., HEEMAN, P., HWANG, C., KATO, T., LIGHT, M., MARTIN, N., MILLER, B., POESIO, M. & TRAUM, D. (1995). The TRAINS Project: a case study in building a conversational planning agent. *Journal of Experimental and Theoretical Artificial Intelligence*, 7, 7–48.
- AMBROS INGERSON, J. (1987). *IPEM: integrated planning, execution and monitoring*. Master’s Thesis, Department of Computer Science, University of Essex.
- ANSCOMBRE, J. & DUCROT, O. (1983). *L’Argumentation dans la Langue*. Bruxelles: P. Mardaga.
- ARISTOTLE (1928). In W.D. ROSS, Ed. *Topics*. Oxford: Clarendon Press.
- BALLIM, A. (1992). *ViewFinder: a framework for representing, ascribing and maintaining nested beliefs of interacting agents*. Ph.D. Dissertation, Université de Genève, Faculté de Sciences.
- BECKER, H., Ed. (1974). *The Health Belief Model and Personal Health Behavior*. Thorofare, NJ: C.B. Slack.
- BLANDFORD, A. (1993). An agent-theoretic approach to computer participation in dialogue. *International Journal of Man–Machine Studies*, 39, 965–998.
- BUCHANAN, B., MOORE, J., FORSYTHE, D., CARENINI, G., OHLSSON, S. & BANKS, G. (1995). An intelligent interactive system for delivering individualized information to patients. *Artificial Intelligence in Medicine*, 7, 117–154.
- BURTON, M. & BRNA, P. (1996). Clarissa: an exploration of collaboration through agent-based dialogue games. *Proceedings of EuroAIED*.
- CARLETTA, J. (1992). *Risk-taking and recovery in task-oriented dialogue*. Ph.D. Dissertation, Department of Artificial Intelligence, University of Edinburgh.
- CARLSON, L. (1983). *Dialogue Games: an Approach to Discourse Analysis*. Synthese Language Library. Dordrecht: D. Reidel.
- CASTELFRANCHI, C. (1996). Conflict ontology. In H. MÜLLER, R. DIENG, Eds. *Proceedings of the ECAI-96 Workshop on: Modelling Conflicts in AI*, pp. 8–14.
- CAWSEY, A., GRASSO, F. & JONES, R. (1999). A conversational model for health promotion on the World Wide Web. In P.W. HORN, Y. SHAHAR, G. LINDBERG, S. ANDREASSEN, J. WYATT. Eds. *Artificial Intelligence in Medicine. Proceedings of the Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making, AIMDM’99*, Vol. 1620 of Lecture Notes in Artificial Intelligence, pp. 379–388. Berlin: Springer-Verlag.

- CAWSEY, A. (1992). *Explanation and Interaction: the Computer Generation of Explanatory Dialogues*. Cambridge, MA: MIT Press.
- CHU-CARROLL, J. & CARBERRY, S. (1994). A plan-based model for response generation in collaborative task-oriented dialogues. In *Proceedings of the 12th, National Conference on Artificial Intelligence (AAAI94)*, pp. 799–805. AAAI Press.
- CHU-CARROLL, J. & CARBERRY, S. (1996). Conflict detection and resolution in collaborative planning. In M. WOOLDRIDGE, J. MÜLLER, M. TAMBE, Eds. *Intelligent Agents II—IJCAI-95 Workshop on Agent Theories, Architectures, and Languages*, Vol. 1037 of Lecture Notes in Artificial Intelligence, pp. 111–126. Berlin: Springer-Verlag.
- CLARK, H. & CARLSON, T. (1982). Speech acts and hearer's beliefs. In N. Smith, Ed. *Mutual Knowledge*, Chapter 1, pp. 1–36. New York: Academic Press.
- DE KLEER, J. (1986a). An assumption-based truth maintenance system. *Artificial Intelligence*, **28**, 127–162.
- DE KLEER, J. (1986b). Extending the ATMS. *Artificial Intelligence*, **28**, 163–196.
- DE KLEER, J. (1988). A general labeling algorithm for assumption-based truth maintenance. *Proceedings of the 7th National Conference on Artificial Intelligence (AAAI88)*, pp. 188–192. San Francisco CA: Morgan Kaufmanns Publishers.
- Department of Health (1992). *The Health of the Nation: a Strategy for Health in England*. London: HMSO.
- DOWLING, W. & GALLIERS, J. (1984). Linear time for testing the satisfiability of propositional Horn formulae. *Journal of Logic Programming*, **3**, 267–270.
- EASTERBROOK, S., BECK, E., GOODLET, J., PLOWMAN, L., SHARPLES, M. & WOOD, C. (1993). A survey of empirical studies of conflict. In S. EASTERBROOK, Ed. *CSCW: Cooperation or Conflict?*, pp. 1–68. London: Springer-Verlag.
- ELHADAD, M. (1995). Using argumentation in text generation. *Journal of Pragmatics*, **24**, 189–220.
- FORSLUND, G. (1995). Toward cooperative advice-giving systems. *IEEE Expert—Intelligent Systems and their Applications*, **10**, 56–62.
- FOX, M. & LONG, D. (1995). Hierarchical planning using abstraction. *IEE proceedings—Control Theory and Applications*, **142**, 197–210.
- FOX, J., KRAUSE, P., ELVANG-GORANSSON, M. (1993). Argumentation as a general framework for uncertain reasoning. In D. HECKERMAN, A., MAMDANI, Eds. *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, pp. 428–434. San Francisco, CA: Morgan Kaufmann Publishers.
- FREEDMAN, R. (1996). *Interaction of discourse planning, instructional planning and dialogue management in an interactive tutoring system*. Ph.D. Dissertation, Computer Science Department, Northwestern University.
- FRIES, E. & CROYLE, R. (1993). Stereotypes associated with a low-fat diet and their relevance to nutrition education. *Journal of the American Dietetic Association*, **93**, 551–555.
- FROHLICH, D. & LUFF, P. (1990). Applying the technology of conversation to the technology for conversation. In P. LUFF, D. FROHLICH, Eds. *Computers and Conversation*, Chapter 9, pp. 187–220. New York: Academic Press.
- GALLIERS, J. (1992). Autonomous belief revision and communication. In P. GÄRDENFORS, Ed. *Belief Revision*, Vol. 29 of *Cambridge Tracts in Theoretical Computer Science*, pp. 220–246. Cambridge: Cambridge University Press.
- GOLDEN, J. PILOTTA, J. Eds. (1986). *Practical Reasoning in Human Affairs. Studies in honor of Chaim Perelman*. Dordrecht: Reidel.
- GOODMAN, B. Ed. (1994). *Proceedings of the 4th International Conference on User Modeling*. The MITRE Cooperation.
- HOVY, E. (1990). Pragmatics and natural language generation. *Artificial Intelligence*, **43**, 153–197.
- HOVY, E. (1993). Automated discourse generation using discourse structure relations. *Artificial Intelligence*, **63**, 341–385.
- HUSTADT, U. (1994). A multi-modal logic for stereotyping. In B. GOODMAN Ed. *Proceedings of the 4th International Conference on User Modeling*. The MITRE Cooperation. pp. 87–92.
- JAMESON, A., KIPPER, B., NDIAYE, A., SCHÄFER, R., SIMONS, J., WEIS, T. & ZIMMERMANN, C. (1994). Cooperating to be noncooperative: the Dialog System PRACMA. In B. NEBEL,

- L. DRESCHLER-FISCHER, Eds. *KI-94: Advances in Artificial Intelligence, Proceedings of the 18th German Conference on Artificial Intelligence*, pp. 106–117. Heidelberg: Springer.
- JOKINEN, K., MAYBURY, M., ZOCK, M. & ZUKERMAN, I. Eds. (1996). *Proceedings of the ECAI-96 Workshop on: Gaps and Bridges: New Directions in Planning and NLG*.
- KARACAPILIDIS, N., TROUSSE, B. & PAPADIAS, D. (1997). Using case-based reasoning for argumentation with multiple viewpoints. In D. LEAKE, E. PLAZA, Eds. *Case-Based Reasoning: Research and Development, Proceedings of the 2nd Intern. Conference on Case-Based Reasoning (ICCB'97)*, Vol. 1266 of Lecture Notes in Artificial Intelligence, pp. 541–552. Berlin: Springer-Verlag.
- KARACAPILIDIS, N. (1996). An argumentation based framework for defeasible and qualitative reasoning. In JOKINEN, MAYBURY, ZOCK, ZUKERMAN, Eds. *Proceedings of the ECAI-96 Workshop on: Gaps and Bridges: New Directions in Planning and NLG*, pp. 37–42.
- KAY, J. (1994). Lies, damned lies and stereotypes: pragmatic approximations of users. In B. GOODMAN, Ed. *Proceedings of the 4th International Conference on User Modeling*. The MITRE Corporation. pp. 175–184.
- KOBSA, A. (1990). Modeling the user's conceptual knowledge in BGP-MS, a user modeling shell system. *Computational Intelligence*, **6**, 193–208.
- KOWTKO, J., ISARD, S. & DOHERTY, G. (1992). *Conversational games within dialogue*. Technical Report HCRC/RP-31, Human Communication Research Centre, University of Edinburgh.
- LAMBERT, L. & CARBERRY, S. (1999). A process model for recognizing Communicative acts and modeling negotiation subdialogues. *Computational Linguistics*, **25**, 1–53.
- LEE, M. & WILKS, Y. (1997). Eliminating deceptions and mistaken belief to infer conversational implicature. In K. JOKINEN, D. SADEK, D. TRAUM, Eds. *IJCAI 1997 Workshop on Collaboration, Cooperation and Conflict in Dialogue Systems*.
- MANELI, M. (1994). *Perelman's New Rhetoric as Philosophy and Methodology for the Next Century*. Dordrecht: Kluwer.
- MANN, W. & THOMPSON, S. (1988). Rhetorical structure theory: toward a functional theory of text organization. *Text* **8**, 243–281.
- MANN, W., MATTHIESSEN, C. & THOMPSON, S. (1989). *Rhetorical structure theory and text analysis*. Technical Report, USC/Information Sciences Institute.
- MARCU, D. (1996). The conceptual and linguistic facets of persuasive arguments. In JOKINEN, MAYBURY, ZOCK, ZUKERMAN, Eds. *Proceedings of the ECAI-96 Workshop on: Gaps and Bridges: New Directions in Planning and NLG*. pp. 43–46.
- MAYBURY, M. (1993). Communicative acts for generating natural language arguments. In *Proceedings of the 11th National Conference on Artificial Intelligence (AAAI93)*, pp. 357–364. AAAI Press/The MIT Press.
- MCCOY, K. F. (1989). Generating context sensitive responses to object-related misconceptions. *Artificial Intelligence*, **41**, 157–195.
- MCROY, S., ALI, S. & HALLER, S. (1998). Mixed depth representation for dialog processing. In M. GERNSBACHER, S. DERRY, Eds. *Proceedings of the 20th Annual Meeting of the Cognitive Science Society*, pp. 687–692. Hillsdale, NJ: Hillsdale, Lawrence Erlbaum Associates.
- Ministry of Agriculture, Fisheries and Food (MAFF) (1995). *Manual of Nutrition* (10th edn). Reference book 342. London: HMSO.
- MOORE, J. & PARIS, C. (1993). Planning text for advisory dialogues: capturing intentional and rhetorical information. *Computational Linguistics*, **19**, 651–695.
- MOORE, J. (1995). *Participating in Explanatory Dialogues*. Cambridge, MA: MIT Press.
- PAOLUCCI, M., SUTHERS, D. & WEINER, A. (1996). Automated advice-giving strategies for scientific inquiry. *Proceedings of the 3rd international Conference on Intelligent Tutoring Systems (ITS'96)*.
- PERELMAN, C. & OLBRECHTS-TYTECA, L. (1969). *The New Rhetoric: a Treatise on Argumentation*. University of Notre Dame, IN: Notre Dame Press.
- PILKINGTON, R., HARTLEY, J., HINTZE, D. & MOORE, D. (1992). Learning to argue and arguing to learn: an interface for computer-based dialogue games. *Journal of Artificial intelligence in Education*, **3**, 275–285.
- POESIO, M. & TRAUM, D. (1997). Conversational actions and discourse situations. *Computational Intelligence*, **13**, 309–347.

- PROCHASKA, J. & CLEMENTE, C. D. (1992). Stages of change in the modification of problem behavior. In M. HERSEN, R. EISLER, P. MILLER, Eds. *Progress in Behavior Modification*, Vol. 28. Sycamore, IL: Sycamore Publishing Company.
- PROCHASKA, J., NORCROSS, J. & CLEMENTE, C. D. (1994). *Changing for Good*. Avon Books, New York.
- REED, C. & LONG, D. (1997a). Collaboration, cooperation and dialogue classification. *Working notes of the IJCAI'97 Workshop on Collaboration, Cooperation and Conflict in Dialogue Systems*.
- REED, C. & LONG, D. (1997b). Content ordering in the generation of persuasive discourse. In *Proceedings of the 15th International Joint Conference of Artificial Intelligence (IJCAI-97)*. San Francisco, CA: Morgan Kaufmann Publishers.
- REED, C., LONG, D. & FOX, M. (1996). An architecture for argumentative dialogue planning. In D. GABBAY & H. OHLBACH, Eds. *Practical Reasoning: Proceedings of the 1st International Conference on Formal and Applied Practical Reasoning (FAPR'96)*, Vol. 1085 of *LNAI*, pp. 555–566. Berlin: Springer-Verlag.
- REITER, E., CAWSEY, A., OSMAN, L. & ROFF, Y. (1997). Knowledge acquisition for content selection. *Proceedings of the 6th European Workshop on Natural Language Generation*, pp. 117–126.
- ROE, L., HUNT, P., BRADSHAW, H. & RAYNER, M. (1997). *Health Promotion Interventions to Promote Healthy Eating in the General Population: a Review*. Health Education Authority—The Cromwell Press.
- SACERDOTI, E. (1974) Planning in a hierarchy of abstraction spaces. *Artificial Intelligence*, **5**, 115–135.
- SADALLA, E. & BURROUGHS, J. (1981). Profiles in eating: sexy vegetarians and other diet-based social stereo types. *Psychology Today*, **15**, 51–57.
- SILLINCE, J. (1994). Multi-agent conflict resolution: a computational framework for an intelligent argumentation program. *Knowledge-Based System*, **7**, 75–90
- STEFIK, M. (1981). Planning and Meta-Planning (MOLGEN: Part 2). *Artificial Intelligence*, **16**, 141–170.
- STEIN, A. & MAIER, E. (1995). Structuring collaborative information-seeking dialogues. *Knowledge-Based Systems*, **8**, 82–93.
- SYCARA, K. (1988). Resolving conflicts via negotiation. In *Proceedings of the 7th National conference on Artificial Intelligence (AAAI88)*, pp. 245–250. San Francisco, CA: Morgan Kaufmann Publishers.
- SYCARA, K. (1990). Persuasive argumentation in negotiation. *Theory and Decision*, **28**, 203–242.
- TAYLOR, J., CARLETTA, J. & MELLISH, C. (1995). *Requirements for belief models in cooperative dialogue*. Technical Report HCRC/RP-66, Human Communication Research Centre, University of Edinburgh.
- TAYLOR, J., CARLETTA, J. & MELLISH, C. (1996). *Combining power with tractability in belief models*. Technical Report HCRC/RP-76, Human Communication Research Centre, University of Edinburgh.
- TAYLOR, J. (1994). *A multi-agent planner for modelling dialogue*. Ph.D. Dissertation, University of Sussex.
- TOULMIN, S. (1958). *The Uses of Argument*. Cambridge: Cambridge University Press.
- TRAUM, D. & ALLEN, J. (1994). Obligations in dialogue processing. *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL94)*, pp. 1–8.
- TRAUM, D. (1994). *A computational theory of grounding in natural language conversation*. Ph.D. Dissertation, Computer Science Department, University of Rochester.
- US Department of Health and Human Services (1991). *Healthy people 2000: national health promotion and disease prevention objectives*. DHHS publication PHS 91-50212.
- VAN EEMEREN, F., GROOTENDORST, R., HENKEMANS, F., BLAIR, J., JOHNSON, R., KRABBE, E., PLANTIN, C., WALTON, D., WILLARD, C., WOODS, J., & ZAREFSKY, D. (1996). *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- VREESWIJK, G. (1997). Abstract argumentation systems. *Artificial Intelligence*, **90**, 225–279.
- WALTON, D. (1989). *Informal Logic: a Handbook for Critical Argumentation*. Cambridge: Cambridge University Press.

- WARNICK, B. & KLINE, S. (1992). The new Rhetoric's argument schemes: a rhetorical view of practical reasoning. *Argumentation and advocacy*, **29**, 1-15.
- WASHINGTON, R. (1994). *Abstraction planning in real time*. Ph.D. Dissertation, Department of Computer Science, Stanford University.
- WERCH, C. & CLEMENTE, C. D. (1994). A multi-component stage model for matching drug prevention strategies and messages to youth stage of use. *Health Education Research*, **9**, 37-46.
- WHO (1986). World Health Organization. Ottawa Charter.
- YANG, Q. (1992). A theory of conflict resolution in planning. *Artificial Intelligence*, **58**, 361-392.
- ZIEBLAND, S., THOROGOOD, M., YUDKIN, P., JONES, L. & COULTER, A. (1998). Lack of willpower or lack of wherewithal? 'internal' and 'external' barriers to changing diet and exercise in a three year follow-up of participants in a health check. *Social Science and Medicine*, **46**, 461-465.
- ZUKERMAN, I., KORB, K. & MCCONACHY, R. (1996). Perambulations on the way to an Architecture for a nice argument generator. In JOKINEN, MAYBURY, ZOCK, ZUBERMAN, Eds. *Proceedings of the ECAI-96 Workshop on: Gaps and Bridges: New Directions in Planning and NLG* (1996), 32-36.