Towards More Flexible and Common-Sensical Reasoning about Beliefs

Gees C. Stein and John A. Barnden.

gstein@crl.nmsu.edu , jbarnden@crl.nmsu.edu
Computing Research Lab & Computer Science Dept.
New Mexico State University
Las Cruces, NM 88003-8001

Abstract

Some important research problems within reasoning about beliefs are how to deal with incomplete, inaccurate or uncertain beliefs, when to ascribe beliefs and how to make reasoning about beliefs more common-sensical. We present two systems that attack such problems. The first, called CaseMent, uses *case-based reasoning* to reason about beliefs. It appears that using cases as opposed to rules has advantages with respect to reusability, context-sensitivity of reasoning, belief ascription and introspection. Although the program is still in an initial stage, the first results from the existing implementation look promising. The second system, ATT-Meta, addresses the fact that *metaphorical descriptions of mental states* play an important role in the coherence of mundane discourse (including stories). ATT-Meta's reasoning module has an advanced prototype implementation. It incorporates powerful facilities for defeasible and uncertain reasoning. Both systems use *simulative reasoning* and thereby reduce the amount of explicit self-knowledge they need in ascribing, to other agents, beliefs and reasonings similar to their own.

1 Introduction

The call for papers addressed the issues of what is needed for story understanding, the way mental states should be represented, the reusability of representations, introspection, and the type/amount of explicit self-knowledge. In this paper two systems will be discussed that reason about beliefs and address certain aspects of these issues. The two systems are based on different reasoning techniques and focus on different problems of reasoning about beliefs. The first system uses case-based reasoning in order to grapple with some common-sensical aspects of reasoning about beliefs. The second system focuses on other common-sensical aspects in concentrating on discourse contexts that contain metaphorical descriptions of mental states.

Although both systems are currently confined to reasoning about beliefs as opposed to other mental states, we plan to expand to other mental states in the future.

Section 2 discusses those issues from the workshop's call for papers that our systems address to a significant extent. Section 3 discusses the general technique of case-based reasoning, which is the basis of our first system, CaseMent. Section 4 sketches CaseMent itself, and section 5 shows why case-based reasoning helps with some of the issues of section 2. Section 6 sketches our ATT-Meta system, and then section 7 discusses its benefits with regard to the issues of section 2. Section 8 concludes, and briefly points out how CaseMent and Att-Meta could be brought together in the future.

2 Belief-Reasoning Issues Addressed

In this section we discuss aspects of several of the crucial issues mentioned in the call for papers for the workshop. These aspects are addressed by one or both of the systems we will be describing.

2.1 Kinds of Reasoning Done by People

Classical logics rely on deductive inferences and most systems that reason about beliefs do the same (Creary 1979, Levesque 1984, Haas 1986, Hadley 1988, Lakemeyer 1991, Maida et al. 1991). In particular, believers are implicitly or explicitly cast as performing deduction. However, people do not only use deduction but also induction, abduction, analogy-based reasoning and other forms of plausible reasoning. They have the ability to infer general rules to explain some particular examples (induction) or generate plausible explanations for observed facts (abduction — see, for example, Hobbs et al. 1993). While deduction produces logically sound results, induction and abduction produce plausible results, which do not necessarily have to be true. Ideally, systems that reason about beliefs should be able to ascribe abduction, deduction, induction, analogy-based reasoning etc. to people.

2.2 Incomplete, Incorrect or Uncertain Information

Ideally, reasoning happens with all the relevant information available, and with all the information being correct and free of uncertainties. In reality people often have to reason with incomplete, incorrect and uncertain information. They generate defeasible conclusions, conclusions that seem likely to be true but later might turn out to be false. A system that reasons about beliefs has to deal with these issues at both the system level and within a belief space. Because agents might have incomplete or uncertain beliefs it is not enough for the system itself to know how to reason with its own incomplete, uncertain beliefs; it has to ascribe such reasoning to other agents as well. Only a minority of research on belief has endowed systems with this sort of capability. One example is the work of Cravo & Martins (1993), on a version of SNePS that uses default reasoning.

2.3 Shades of Belief

Most systems do not deal with different kinds of beliefs. However, beliefs can be of different varieties or "shades". For example, they can be held with different degrees of consciousness, they can be more or less active (i.e., more or less capable of affecting reasoning and action), and they can be momentary or long-lasting. The particular shade of a belief can be implied by its context. Take for instance *"Helga believes that Werner is intelligent,"* in a context where Helga is not just in love with Werner, but really worships him, and where Werner is not in fact intelligent. Then, plausibly, Helga's belief is a kind of "wishful thinking" belief: something halfway between wishing and believing. Contrast this with *"Mary believes olive oil is very healthy,"* given that Mary has done a lot of research on how olive oil affects health and that Mary is a well-respected scientist. Mary's belief is presumably not wishful thinking. Other mental states also come in various shades. For instance, hoping involves a context-sensitive mixture of expectation and desire.

Levesque (1984) and Lakemeyer (1991) make a distinction between "explicit" and "implicit" beliefs, mainly to solve the logical omniscience problem. It is common in philosophy to distinguish "tacit" beliefs as a special case (see Richard 1990 for discussion). However, these distinctions only scratch the surface of the issue of shades of belief (or of other mental states).

2.4 Ascribing Beliefs

To be able to reason about other agents' beliefs a system has to make assumptions about what the agents believe. Normally, relatively few if any of the beliefs of an agent are known with certainty, and the system has to "ascribe" beliefs: assume that the agent has particular beliefs.

People do a similar thing: they can interact even with someone totally unknown by assuming that this person has certain beliefs. People appear often to make a default assumption that other people have beliefs similar to their own . By default everybody is expected to believe that the earth is round. At the same time a person might have certain beliefs that he or she does not ascribe to others: an astronomer will not generally assume that non-astronomers have the same beliefs as he about stars.

A system has to know which of its beliefs it can ascribe to other agents and which not. The same is relevant for nested beliefs: the system has to reason about agent A deciding whether agent B can be expected to have one of A's beliefs or not. Whether to ascribe a belief or not should depend at least on the contents of the belief and the nature of the agent involved, as we see in the astronomer example. Now, some existing systems go a useful distance towards handling this issue. For example, Ballim &Wilks (1991)'s ViewGen system has a liberal, default method for belief ascription between agents. The default is that everything the system believes can be ascribed to another agent, unless contradicted by known beliefs of the agent. (Similarly in nested cases.) However, the default can be bypassed for a particular belief by specifying that the belief is only possessed by particular agents or types of agents.

However, existing systems do not take account of context in more sophisticated but clearly useful ways. For instance, it is plausible that, when trying to ascribe belief B to person P, a person can use the following heuristics:

(i) if B is similar in subject matter to an existing belief of P's, then ascribing B to P is reasonable.

(ii) if other people similar to P believe B, then ascribing B to P is reasonable.

(iii) if other people similar to P believe something similar in subject matter to B, then ascribing B to P is reasonable.

Thus, analogy might be a good way for a system to improve its belief ascription. Of course, a suitable notion of similarity would need to be developed.

2.5 Introspection: Knowledge of Own Mental Processes

It is useful for an intelligent system to be able to view other agents as reasoning in much the same way as it itself reasons. This view is a useful default, in the absence of specific evidence about how another agent is reasoning. Now, a particularly effective way of embodying this strategy in a system is *simulative reasoning.* This technique has often been suggested in AI, under various names (see, e.g., Attardi & Simi 1994, Ballim & Wilks 1991, Chalupsky 1993, Creary 1979, Dinsmore 1991, Haas 1986, Konolige 1986, Moore 1973); it is also closely related to the Simulation Theory in philosophy and psychology (Davies & Stone, in press). Simulative reasoning works as follows. When reasoning about another agent X, the system in effect *pretends* it is agent X by using X's beliefs as if they were its own, and applying its own reasoning mechanism to them. Results of such pretences are then wrapped inside an "X believes that" layer. Simulative reasoning has various advantages over the alternative methods that would otherwise be needed (such as reasoning explicitly about inference steps taken by X). One salient benefit is that *it does not require the system to have explicit information about its own reasoning abilities* in order to to be able to ascribe them to other agents. This advantage is especially strong in a system that has a variety of reasoning methods — e.g., induction, abduction, analogy-based reasoning, as well as deduction — or where an individual method is procedurally very complex — e.g., analogy-based reasoning. (See Haas, 1986, and Barnden, in press, for further discussion and other advantages.) Under simulative reasoning, the system's reasoning methods are automatically and implicitly ascribed to other agents.

Nevertheless, there are complications. Simulative reasoning should be defeasible. That is, the conclusions it provides about another agent's beliefs should be capable of defeat by other evidence. Also, it should be gracefully integrated with non-simulative reasoning about other agents' beliefs. Non-simulative reasoning is still needed, because some inferences about an agent's beliefs are not dependent on the agent's reasoning — e.g., there could be a rule that says that an agent who believes the earth is flat probably has many other incorrect beliefs — and because simulative reasoning is relatively weak at producing negative belief conclusions (conclusions of the form: X fails to believe Q).

2.6 Context-Sensitivity

Varieties of context-sensitivity were touched upon in section 2.3 and 2.4. In addition, the context of a reasoning problem can have a important influence on the outcome of the problem. People under stress might come to one conclusion about a certain problem, while those people in relaxed circumstances might come to a very different conclusion. Or, a doctor reasoning about a patient's problem will systematically try to avoid mistakes. When the same doctor is reasoning about what to eat (s)he can afford to be much less 'scientific' in the reasoning. Thus, context plays an important rule in reasoning about beliefs, in several ways. Currently, systems do not make much use of context when reasoning about beliefs.

2.7 Story Understanding, Mental States and Metaphor

In real speech and text, including stories, mental states are often described metaphorically. This is true even when the discourse is not particularly literary. The following are some examples of common, mundane forms of common-sense, metaphorical description of mental states/processes:

John was leaping from idea to idea.
George put the idea that he was a racing driver into Sally's mind.
His desire to go to the party was battling with the knowledge that he ought to work.
Peter hadn't brought the two ideas together in his mind.
Martin had a blurred view of the problem.
Veronica caught hold of the idea and ran with it.
One part of Sally was convinced that Mike was cheating on her.

For further examples and discussion of metaphors of mind see, for example, Belleza (1992), Bruner & Feldman (1990), Cooke and Bartha (1992), Fesmire (1994), Gallup and Cameron (1992), Gentner and Grudin (1985), Gibbs and O'Brien (1990), Jäkel (1993), Johnson (1987), Lakoff, Espenson and Schwartz (1991), Lakoff & Johnson (1980), Larsen (1987), Richards (1989), Sweetser (1987, 1990), and Tomlinson (1986).

Metaphors of mind can make a crucial difference to the understanding of discourse, including stories. In particular, metaphorical descriptions often provide important information about what inferences the characters in a story have made. For instance, if someone hasn't "brought" two ideas "together" in his or her mind, then it can be assumed that (s)he hasn't drawn even the obvious conclusions from them. Notice also that the last example sentence above implies that, probably, some other part of Sally did *not* have the mentioned conviction. It is our contention that description of complex or subtle mental states often requires metaphor, so that understanding of metaphorical descriptions of mind is an important consideration for the workshop. However, the matter has not been intensively addressed in AI other than in our own work on the ATT-Meta system (Barnden 1988, 1989a,b, 1992; Barnden *et al.* 1994a,b, and forthcoming).

2.8 Logical Form of Mental-State Representations

Our attention to metaphorical description of mental states, in the case of our ATT-Meta system, has had profound consequences for the style of internal logical representation that we use in that system. Usually, metaphor researchers tacitly assume that a system that interprets metaphorical sentences must construct internal meaning representations that cash out the metaphors in objective, non-metaphorical terms. (There are exceptions, notably Hobbs 1990.) However, we take an opposite tack, for reasons explained in Barnden *et al.* (in press). That is, the meaning representation for the sentence "In the recesses of her mind Veronica believed that Sally was wrong" takes Veronica's mind to be a physical space and places her believing in the recesses of that space; that is, roughly speaking, *the representation "pretends" that the sentence is LITERALLY true.* Although our system can in some cases partially translate such metaphor-imbued internal representations into metaphor-free ones, there are many cases when this is not possible and yet the metaphor-imbued representations can be used in inference processes that yield useful results. This point is clarified below.

2.9 Reusability of Knowledge

Finally there is the issue how knowledge about beliefs could be reused. Often systems try to focus on a specific domain within which reasoning about beliefs occurs. 'Belief' bases have to be developed together with reasoning rules or other reasoning mechanisms. It would be helpful if this knowledge could be easily reused by other systems. Not much attention has been paid to this issue though.

3 Case-Based Reasoning

The first system that will discussed, CaseMent, has as its main motivation that it tries to be a more common-sensical and realistic system with respect to the way people reason about beliefs. This system uses case-based reasoning (CBR) as opposed to rule-based reasoning to reason about beliefs. In this section we give a short description of CBR in general; we present the CaseMent system in the next section.

CBR is an approach that tries to solve a problem by using knowledge/experience with previous cases that have similarities with the problem at hand (Riesbeck and Schank 1989, Kolodner 1993). A case can be seen as a previous experience. Often people like judges, doctors, use their previous experience with similar cases when trying to solve a problem. If a certain approach in a previous similar case was successful, it might work again. If no rules are known to solve a problem, the only way left to deal with it might be looking at similar known and tried cases.

The main steps involved in CBR are finding relevant cases, analyzing how they differ from the problem at hand, choosing one or more best cases, and adapting solutions to fit the current problem. Because cases that are slightly different from the problem still can be used by adapting solutions, a case can be used for more than one problem. Depending on how 'smart' the system is, what kind of analogies it can find between a problem and a different case, and how good it is at adapting solutions, a case can be used in many ways. Because a case does not have to match a problem exactly it is relatively easy to deal with incomplete and incorrect information.

A case-based reasoning system's performance is improved in several ways. Cases are added which gives the system more experience to build on. The system gets feedback from its proposed answers. Not only correct answers are stored, but the system also keeps track of its failures, to prevent similar mistakes in the future. Cases inherently provide more context-sensitivity than rules through storing more information about individual situations.

CBR seems to be especially suited for those 'non-exact' domains that do not have clear rules that tell what to do in a certain situation. If the knowledge about certain problems is known but very complex and time consuming to use, case-based reasoning can improve a system's performance. Case-based reasoning systems have been developed for such different tasks as creating recipes, diagnosing heart failures, classifying hearing disorders, settling resource disputes and designing physical devices (Kolodner 1993). To our knowledge, it has never been applied to reasoning about beliefs.

4 The CaseMent System: Using CBR for Belief Reasoning

The current CaseMent system is implemented in Prolog and still in its initial stage. First a short description of the basic system will be given to give a feeling how the system works. The next section will discuss the expected advantages of using case-based reasoning.

A case in CaseMent is an experience that might or might not involve beliefs. When beliefs are involved, they can be nested. An agent might have a set of cases about a variety of things. He can have an experience about cooking a chicken of 5 pounds for 2 hours which came out very good. This would be one case. Another case might be the experience that his neighbor Mr. Jones believes that women always gossip. The fact that Mr. Jones has a wife that gossips a lot could also be part of that case. In the current system, these two cases, (c1) and (c2), would be represented as follows:

```
case_part(c1,chicken(food),1).
case_part(c1,weight(food,5_lbs),2).
case_part(c1,cooked(food,2_hrs),3).
case_part(c1,test(food,good),4).

case_part(c2,believes(jones,woman(X)-->gossips(X)),1).
case_part(c2,married_to(jones,wife),2).
case_part(c2,gossips(wife),3).
case_part(c2,because(1,[2,3]),4).
```

The third argument of a case part is a unique number that can be used to relate to case parts to each other. In case (c2) the 4th case part says that Jones believes that all women gossip *because* his wife gossips. Right now the only way to relate parts of cases to each other is the 'because' predicate.

Using the first case the system might decide that a turkey of similar weight probably needs 2 hours of cooking if it realizes that turkey and chicken are similar. Using the second case and reasoning about another man who believes that women gossip, it might assume that perhaps this man also has a wife that gossips a lot. In general a query consists of the question itself and contextual information. The answer right now is either 'yes', 'no', 'have insufficient information' or 'have unresolvably contradicting information', together with one or more case(s) on which the answer is based.

The system does not only have its own cases but can also contain other agent's cases, in other words case bases can be nested. Suppose agent Annette holds her own set of cases and the system has a question about Annette's beliefs, together with relevant information.

First the system can consider its own cases, that might be about Annette's beliefs, or about someone else's beliefs that seem to apply to the question and its context. This can be seen as *metareasoning* with cases. Next it will look at Annette's cases and see if any are relevant. This is where simulative reasoning comes in. What happens is that the system will consider Annette's cases as they were its own. It will apply its *own* case based reasoning to these cases. In other words, simulative reasoning in CaseMent amounts to ascribing its CBR to other agents.

An example of a nested case is:

```
case_part(c3,has_case(annette,a1),1).

case_part(a1,not-gossips(wife),1).
case_part(a1,married_to(jones,wife),2).
```

If the query is whether "believes(Annette,gossips(wife))" is true or not, and the context contains "married_to(jones, wife)", both Annette's case (a1) and the system's case (c2) will be considered. Both cases mention a belief about Jones' wife gossiping. (c2) mentions Jones' belief and (a1) contains Annette's belief. As the question is about what Annette believes, the system will consider (a1) a more relevant case and use it to say that Annette believes that Jones's wife does not gossip.

In general more than one case might seem relevant in an episode of CBR. The system has to be able to compare cases to each other with respect to their relevance to the query. This will happen within simulative reasoning, and also at the metareasoning level. If the question is about (nested) beliefs, both simulative reasoning and metareasoning might give an answer. The system then has to compare the cases used for both reasoning methods and decide which one seems to fit the query best.

5 Advantages of CBR for Belief Reasoning

Given the basic idea of the system, the expected advantages of using case-based reasoning to reason about beliefs will be discussed, with special attention to the issues listed in section 2.

As was said earlier, case-based reasoning is especially suited for those domains that are not very well understood and for which it is very hard to define precise rules. Reasoning about beliefs seems to fit that description. People might try to be reasonable but sometimes they are not. If not impossible it is at least very hard to define rules that tell us how different people having different sets of beliefs reason about beliefs in different circumstances. It is easier to describe *examples* of people having certain beliefs and making particular inferences: to create cases.

Cases do not only have to describe people's logically correct inferencing, but can also store mistakes people make. Cases can capture different kinds of reasoning, mistakes in reasoning or inferences not made. An example might be a case where someone believes that if the car's gas tank was just filled with gas it has to be full. But because he finds it empty he reasons that the tank was not filled. Of course this is not necessarily true, there might be a leak in the tank.

```
case_part(c4,believes(joe,filled(tank)-->full(tank)),1).
case_part(c4,believes(joe,not-full(tank)),2)
case_part(c4,believes(joe,not-filled(tank)),3)
case_part(c4,filled(tank),4).
case_part(c4,leaks(tank),5).
case_part(c4,sleepy(joe),6).
case_part(c4,in_hurry(joe),7).
case_part(c4,because(3,[2,1]),8).
```

The example above is an example of a person making a reasoning mistake. One example does not mean that this reasoning mistake is a general one. Context is very important. Within the context of being under a lot of pressure people might act or reason differently than under very relaxed circumstances. While rules try to be as general as possible, cases are meant to store relevant context. Because cases can be used not only for exactly matching queries but also for 'close' ones, the context does not restrict their usage but makes it possible to use context when available.

Because a case does not have to match exactly to be used, case-based reasoning can deal with inaccurate information or even missing information. The case (c1) about chicken might be used for turkey too if the system knows that chicken and turkey are similar. Case (c4) might still be used to reason about another person, Sally, even if it is not known whether Sally is sleepy or not. The system can assume those missing parts of a case if otherwise it matches the input. Together with an answer the system also tells on which case(s) it is based. These cases implicitly show how certain/uncertain the answer is by their degree of similarity to the input problem.

Beliefs in different kind of circumstances might have different intensity, or level of consciousness, or they might influence behavior in different ways. Our feeling is that there are different shades of belief (section 2.3). However, it is hard to make the differences explicit and to define when they are present. Because cases keep the context of a situation they *implicitly* can deal with different kinds of beliefs. Consider cases partially described as follows:

```
case_part(c5, worships(helga, werner), 1).
case_part(c5, in_love_with(helga,werner), 2).
case_part(c5, believes(helga, intelligent(werner)), 3).
case_part(c5, not-intelligent(werner), 4).

case_part(c6, done_research(marian,swimming), 1).
case_part(c6, does(marian,swimming), 2).
case_part(c6, believes(marian,healthy(swimming)), 3).
```

Both of these cases mention beliefs, but it is plausible that they involve different kinds of beliefs. The belief in (c5) could well be "wishful thinking." Accordingly, other parts of (c5) and (c6) might involve distinctively different types of facts about Helga and Marian.

If the system has a certain case/experience, when can it ascribe this to another agent? As pointed out in section 2.4, this should depend on the case **and** on the agent the ascription is done to. A general case about eating pizza in a restaurant can probably be ascribed to most agents. A case that is about specialist medical information should not be ascribed to other agents unless that agent is familiar with the medical field. Ascription of a case can also happen within belief spaces, for instance 'does agent X think that his case C is common enough to be ascribed to agent Y?

A similar problem is the reverse: suppose an agent has a certain case. The system does know about this, but will it use this case to solve a problem? It might if the system trusts this particular agent with respect to this particular case. A similar question can be asked within several belief layers: 'does agent X trusts agent Y enough with respect to agent Y's case C that it will use case C?'. If the case C is a medical case and the agent Y is a well respected doctor, it seems reasonable to 'trust' Y and X can use case C as it were its own.

The third related problem occurs in the following situation: suppose the system tries to reason about agent X's belief, given a certain context. If finds a case C similar to the problem, but it is a case within agent Y's case base. This case can only be used if it is likely that X might have had a similar experience. So the question is whether agent X is similar to agent Y with respect to the case C.

Although this feature is not implemented yet, the system will use case-based reasoning to decide what it should do in any of the three situations described above. Cases might mention situations of certain agents (not) believing other agents. Or they might mention certain facts that are not believed by certain people. When trying to ascribe a case C to person P, cases that P already has can be checked. If P has a very similar case C' it is probably acceptable to ascribe case C to P. If P has a contradicting case it does not make sense to ascribe case C to P. Another possibility is to look at other agents whose case bases contain the case C. If these agents are very similar to agent P the system might ascribe case C to P. The combination of both previous examples is also possible: another agent has a case C' similar to case C. If the cases are similar enough and the agents seem similar with respect to these cases, case C can be ascribed to P. Not only agent P's cases and other agent's cases can be used: *meta* cases *about* agent P and other agents can also contain useful information, for instance about the beliefs someone does not have.

Concerning the issue of introspection, most belief-reasoning systems use explicit rules to define positive and negative introspection. A case-based reasoning systems can base introspection on cases it has about agents' introspection, and hereby make introspection more context sensitive.

Because CBR is combined with simulative reasoning in CaseMent, the characteristics described above are true not just for the system's reasoning but also for other agents' reasoning. Not only the system can handle incomplete information, agents will handle incomplete information the same way. By default, an agent will be assumed to ascribe cases to other agents similar to the way the system does. The system does not need explicit information about how agents (including itself) can reason, which makes the system more efficient, flexible and modifiable.

It is not only with respect to the above issues that CBR seems promising. It also has some characteristics that make it suitable for *reuse*. First there is the point that a case can be used in many ways. If the CBR system is 'smart' enough that it can not just detect surface similarities but also more abstract similarities, cases can be used in creative ways. A case can present an experience from several points of view and can therefore be used in different ways. The creative use of cases makes it easier to reuse cases in different domains. After all, people often use their experiences in creative ways across domains.

Another important characteristic that makes it easier to reuse cases is the fact that they store the context of their main contents. Rules are often tailored for a specific domain, handcrafted in such a way that they interact in the right way with each other. Using rules in different domains and having to interact with new rules might cause problems. The context a rule should be used in is not known anymore, and the system might use it when it does not apply. Because a case contains more of the context for which it was meant to solve problems it has less danger of doing the wrong thing in a different domain. It will still fit those situations it was meant for best because of its context.

Finally a word about possible applications this kind of system would suit especially well. Any application where cases are available or can be extracted should be fine. A specific example would be a computer aided instruction system where a student interacts with a system to gain knowledge. For each student the system can keep track of what is learnt and what mistakes were made. When interacting with a student the system will use this knowledge to know what it has to explain and what it can assume as known. Previous mistakes would help to understand new mistakes and even anticipate mistakes. Not only cases of the student could be used, the system might also use other students' experiences to anticipate and understand.

6 ATT-Meta

The second system we present in this paper, ATT-Meta, is rule-based. It combines metaphor-based reasoning and simulative reasoning to reason about discourse coherence in certain specialized ways. In mundane text and speech, mental states are often described with the aid of metaphors, and the metaphorical descriptions can convey information that is crucial for understanding the discourse. Being able to reason with mental metaphors helps a system to understand characters that might forget certain things, might not make the obvious conclusions, might act in certain ways and might make certain inferences. Crucial information bearing on those points is often communicated via metaphors.

ATT-Meta consists of two parts. The first part (still under construction) has as input the discourse, and produces the logical representation of the discourse. The latter is the input for the reasoning part which has an advanced prototype implementation (demonstrated at the 1994 meeting of the Association for Computational Linguistics). ATT-Meta is described in more detail in Barnden et al. (1994a, 1994b).

To see the importance of metaphors in mental-state description, consider the following discourse fragment (1), with possible continuations (1a-c):

(1) *Veronica was preparing for her dinner party. Her brother's recipe had said to fry the mushrooms for one hour.*

(1a) *She did this even though she believed the recipe to be wrong.*
(1b) *She did this even though in the recesses of her mind she believed the recipe to be wrong.*
(1c) *She did this even though she thought, "The recipe's wrong".*

(1a) is a non-metaphorical continuation, (1b) contains the MIND AS PHYSICAL SPACE metaphor and (1c) uses the IDEAS AS INTERNAL UTTERANCES metaphor. In all examples there is the disparity between Veronica's following the recipe and her believing it to be incorrect. (1c) gives the impression that Veronica consciously believed that the recipe was wrong, so a reasonable assumption would be that she had a special reason to follow the wrong recipe. For example, perhaps she was ordered to follow it. By contrast, a reasonable interpretation for (1b) is that Veronica's belief is only minimally involved in her conscious reasoning, so that she did not very consciously believe that she followed a recipe that was wrong. Continuation (1a) does not have a metaphor that can guide us towards a certain interpretation, but it seems a reasonable (default) assumption that she believes it consciously and thus has a special reason just as in (1c). However, the pressure towards this explanation is less strong than in the case of (1c).

Apart from the two metaphors mentioned above, other important ones we have considered are MIND PARTS AS PERSONS and COGNIZING AS SEEING. Under the former, an agent's mind is viewed as being broken up into separate "sub-persons" which have their own mental states, emotions, etc., and are often cast as communicating with each other in natural language. A vivid example of the use of the metaphor is "One part of Veronica was vociferously insisting that the recipe was wrong." An example of COGNIZING AS SEEING is "Veronica only dimly saw that the recipe was wrong."

ATT-Meta is unusual with respect to other systems for reasoning about *mental states*, in that it copes with metaphorical descriptions. ATT-Meta is also unusual with respect to other systems that address *metaphor* (whether the metaphors are about mental states or not) in that ATT-Meta does not (in general) translate metaphorical descriptions into internal, non-metaphorical versions, but instead represents them literally (see section 2.8). For example the internal representation of (1b) itself casts Veronica's mind as a physical space that has recesses. This is done because it is hard to choose metaphor-free representations that can be used in the same ways as the metaphor-imbued ones. Also, keeping the metaphor makes the system more flexible with regard to novel manifestations of a metaphor. (Novelty to the system, not necessarily to the language community.) Suppose, for instance, the system is told that

> *"One part of John was vociferously insisting that Sally was being unfair, but another was calmly resigned to this idea."*

This sentence directly contributes premises saying that a subperson of John was [literally] vociferously insisting that Sally was being unfair, but another subperson was [literally] calmly resigned to this idea. Now, the defeasible inferences from these premises could include:

> *the first subperson believed that Sally was being unfair, and did not like her being unfair;* and
>
> *the second subperson believed that Sally was being unfair but did not mind this being the case.*

Through very general "transfer rules" that map from metaphorical representations to metaphor-free ones, the system could then infer that John (as a whole) believed that Sally was being unfair, and to some extent disliked and and to some extent did not mind her being unfair. The point is that the system does *not* need to have *any transfer rules* concerning insisting, vociferousness, calmness and resignation, that is, concerning the *intermediate* notions used in the chain of reasoning.

ATT-Meta uses a combination of simulative and non-simulative reasoning to reason about other agents' beliefs (see section 2.5. for why such a combination is necessary). To avoid logical omniscience, results of simulative reasoning are at most *default* results. Metaphorical information can be used to downgrade these results further. For instance, if agent X believes P and believes Q, and R follows from P and Q, then the system would normally conclude (as a default) that X believes R. However, if the input discourse implies that P and Q are "far apart" in X's mind, then the system retracts the conclusion. Thus, ATT-Meta's simulative reasoning can be modulated by metaphorical reasoning.

7 ATT-Meta and the Issues of Section 2

On *logical-form*: the system uses metaphor-imbued internal representations (as well as more standard, metaphor-free ones). As a result, its representation style is, in part, fundamentally different from that of other mental-state reasoning systems.

On *introspection*: the system can in principle take a metaphorical view of its own mental states/processes (although we have not actually made the system do this yet). More usefully, the system can cope with speakers who say things like *"Part of me wants to go to the party."* That is, the system can deal with metaphor-based introspection by other agents.

On *shades of belief*: by paying attention to metaphors, ATT-Meta represents a far richer variety of mental states than other systems are capable of. In particular, examples like (1b) and (1c) above illustrate that ATT-Meta addresses different shades of belief.

On *reusability*: when reasoning about a metaphorically-described mental state, ATT-Meta uses general knowledge about the source domain of the metaphor. E.g., if the metaphor is MIND AS PHYSICAL SPACE, ATT-Meta uses general knowledge about physical space (and physical objects). This knowledge could also be used for other, non-metaphorical purposes, and is thus more reusable than the knowledge the system would need to have if it were to get the same effect by reasoning entirely with metaphor-free representations of mental states. In fact, in developing ATT-Meta we have made a conscious effort to ensure that our knowledge bases for domains like physical space are useful for reasoning in those domains as such, rather than just for metaphorical reasoning about minds.

On *uncertainty* and *incompleteness*: ATT-Meta attaches qualitative certainty factors to the conclusions it draws. These factors result from certainty factors attached to rules in the system's knowledge base and from the certainty factors of the propositions the rules work on. Conclusions that are not absolutely certain can be defeated by contrary conclusions from other lines of reasoning. Such defeats arise partly with the aid of a specificity-based heuristic. In addition, agents' beliefs themselves have qualitative certainty factors. The factors at different levels are independent. Thus, it is possible for ATT-Meta to be certain that agent X believes P as a default; alternatively, ATT-Meta might assume by default that agent X is certain that P. Certainty factors that are "within agents' minds" are manipulated during simulative reasoning in just the same way as the certainty factors at the system's own level.

On *kinds of reasoning* and *self-knowledge*: by virtue of using simulative reasoning, ATT-Meta ascribes its own complex, sophisticated default/uncertain reasoning method to other agents, without needing to describe that reasoning explicitly to itself. It can also ascribe its simulative reasoning and its metaphor-based reasoning to other agents, to any degree of nesting. See Barnden et al. (1994b) for some discussion of the significance of various types of nesting.

Unlike CaseMent, ATT-Meta has no special handle on context-sensitivity and belief ascription.

8 Conclusion

The two systems discussed both contribute to improving reasoning about beliefs, making it more common-sensical, realistic and flexible, although they focus on different issues and are based on different techniques.

ATT-Meta was the first of the two systems to be developed and uses a complicated combination of simulative reasoning, metareasoning, metaphor based reasoning, multi-valued logic and truth maintenance. Making sure that rules interact in a desirable way becomes hard, and it is not always easy to predict how the system will reason. Even for relatively simple questions the system has to do a large amount of work.

With this in mind it was thought that a system based on cases instead of rules might result in a less complicated, easier to understand system for reasoning about mental states. Initial results look promising but still a great deal of work has to be done before final conclusions can be made. The current case-based system, CaseMent, does not focus on mental metaphors. Future work might include trying to give CaseMent similar metaphorical reasoning capabilities to those ATT-Meta demonstrates.

On the other hand, there is room for rules in a case-based system that reasons about mental states. There has been a considerable amount of work in the CBR community on integrating case-base and rule-based reasoning — see, e.g., Kolodner (1993). Currently, our case-based system is essentially free of rules, but that is largely in order to simplify the research issues.

References

Attardi, G. & Simi, M. (1994). Proofs in context. In J. Doyle, E. Sandewall & P. Torasso (Eds), *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference,* pp.15–26. San Mateo, CA: Morgan Kaufmann.

Ballim, Afzal and Yorick Wilks (1991). *Artificial Believers: The Ascription of Belief.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Barnden, J.A. (1988). Propositional attitudes, commonsense reasoning, and metaphor. In *Procs. 10th Annual Conference of the Cognitive Science Society*, Hillsdale, N.J.: Lawrence Erlbaum. pp.347–353.

Barnden, J.A. (1989a). Belief, metaphorically speaking. In *Procs. 1st Intl. Conf. on Principles of Knowledge Representation and Reasoning* (Toronto, May 1989). San Mateo, CA: Morgan Kaufmann. pp.21–32.

Barnden, J.A. (1989b). Towards a paradigm shift in belief representation methodology. *J. Experimental and Theoretical Artificial Intelligence 2*, 1989, pp.133–161.

Barnden, J.A. (1992). Belief in metaphor: taking commonsense psychology seriously. *Computational Intelligence, 8* (3), pp.520–552.

Barnden, John A. (in press). Simulative Reasoning, Commonsense Psychology and Artificial Intelligence. In *Mental Simulation*, edited by M. Davies and A. Stone.

Barnden, J.A., S. Helmreich, E. Iverson and G.C. Stein (1994a). An Integrated Implementation of Simulative, Uncertain and Metaphorical Reasoning about Mental States. *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference,* pp.27–38. San Mateo, CA: Morgan Kaufmann.

Barnden, J.A., S. Helmreich, E. Iverson and G.C. Stein (1994b). Combining Simulative and Metaphor-Based Reasoning about Beliefs. *Procs. 16th Annual Conference of the Cognitive Science Society* (Atlanta, Georgia, August 1994), pp.21–26. Hillsdale, N.J.: Lawrence Erlbaum.

Barnden, J.A., Helmreich, S., Iverson, E. & Stein, G.C. (forthcoming). Artificial intelligence and metaphors of mind: Advantages of within-vehicle reasoning. *Metaphor and Symbolic Activity.* Accepted with revisions encouraged.

Belleza, F.S. (1992). The mind's eye in expert memorizers' descriptions of remembering. *Metaphor and Symbolic Activity, 7* (3 & 4), pp.119–133.

Bruner, J. & Feldman, C.F. (1990). Metaphors of consciousness and cognition in the history of psychology. In D.E. Leary (Ed.), *Metaphors in the History of Psychology,* pp.230–238. New York: Cambridge University Press.

Chalupsky, Hans (1993). Using hypothetical reasoning as a method for belief ascription. *J. Expt. Theor. Artif. Intell.*, Volume 5, pp. 119 - 133.

Cooke, N.J. & Bartha, M.C. (1992). An empirical investigation of psychological metaphor. *Metaphor and Symbolic Activity, 7* (3 & 4), pp.215–235.

Cravo, Maria R. and João P. Martins (1993). SNePSwD: A newcomer to the SNePS family. *J. Expt. Theor. Artif. Intell.*, Volume 5, pp. 135 - 148.

Creary, Lewis G. (1979). Propositional attitudes: fregean representation and simulative reasoning. *Proceedings of the 6th International Joint Conference on Artificial Intelligence*, Tokyo, pp. 176-181. Los Altos, CA: Morgan Kaufmann.

Davies, M. and Stone, A. (Eds) (in press). *Mental simulation.* Oxford: Blackwell.

Davis, E. (1990). *Representations of Commonsense Knowledge.* San Mateo, California: Morgan Kaufmann Publishers, Inc.

Dinsmore, John (1991). *Partitioned representations.* Dordrecht: Kluwer Academic Publishers.

Fesmire, S.A. (1994). Aerating the mind: the metaphor of mental functioning as bodily functioning. *Metaphor and Symbolic Activity, 9* (1), pp.31–44.

Gallup, G.G., Jr. & Cameron, P.A. (1992). Modality specific metaphors: is our mental machinery "colored" by a visual bias? *Metaphors and Symbolic Activity, 7* (2), pp.93–101.

Gentner, D. & Grudin, R. (1985). The evolution of mental metaphors in psychology: A 90-year perspective. *American Psychologist, 40* (2), pp.181–192.

Gibbs, R.W., Jr. & O'Brien, J.E. (1990). Idioms and mental imagery: the metaphorical motivation for idiomatic meaning. *Cognition, 36* (1), pp.35–68.

Haas, Andrew R. (1986). A Syntactic Theory of Belief and Action. *Artificial Intelligence.* Volume 28, pp. 245-292.

Hadley, Robert F. (1988). A Process-Oriented, Intensional Model of Knowledge and Belief. *Proceedings of the 10th Annual Conference of the Cognitive Science Society*, (Montreal), pp. 354-360. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hobbs, Jerry R., Mark E. Stickel, Douglas E. Appelt and Paul Martin (1993). Interpretation as abduction. *Artificial Intelligence* 63, pp. 69-142.

Hobbs, J.R. (1990). *Literature and cognition.* CSLI Lecture Notes, No. 21, Center for the Study of Language and Information, Stanford University, Stanford, CA.

Jäkel, O. (1993). The metaphorical concept of mind: mental activity is manipulation. Paper No. 333, General and Theoretical Papers, Series A, Linguistic Agency, University of Duisburg, D-4100 Duisburg, Germany.

Johnson, M. (1987). *The body in the mind.* Chicago: Chicago University Press.

Kolodner, Janet (1993). *Case-Based Reasoning.* San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Konolige, K. (1986). *A Deduction Model of Belief.* London, U.K.: Pitman.

Lakemeyer, Gerhard (1991). On the Relation between Explicit and Implicit Belief. *Proceedings of KR '91*, pp. 368-375. San Mateo, California: Morgan Kaufmann Publishers, Inc.

Lakoff, G., Espenson, J. & Schwartz, A. (1991). Master metaphor list. Draft 2nd Edition. Cognitive Linguistics Group, University of California at Berkeley, Berkely, CA.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by.* Chicago: University of Chicago Press.

Larsen, S.F. (1987). Remembering and the archaeology metaphor. *Metaphor and Symbolic Activity, 2* (3), 187–199.

Levesque, Hector J. (1984). A Logic of Implicit and Explicit Belief. *Proceedings of the National Conference on Artificial Intelligence*, pp. 198-202. Los Altos, CA: Morgan Kaufmann.

Maida, Anthony S., Jacques Wainer and Sehyeong Cho (1991). A syntactic approach to introspection and reasoning about the beliefs of other agents. *Fundamenta Informatica* XV, pp. 333-356.

Mamede, Nuno J. (1993). SNePS$_R$ – A SNePS with resources. *Journal Expt. Theor. Artificial Intelligence*, Volume 5, pp. 199 - 213.

Moore, R. C. (1973). D-SCRIPT: A computational theory of descriptions. *IEEE Transactions on Computers,* C-25(4), 366–373 (1976).

Richard, M. (1990). *Propositional attitudes: an essay on thoughts and how we ascribe them.* Cambridge, U.K.: Cambrdige University Press.

Richards, G. (1989). *On psychological language and the physiomorphic basis of human nature.* London: Routledge.

Riesbeck, Christopher K. and Roger C. Schank (1989). *Inside Case-based Reasoning.* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Sweetser, E.E. (1987). Metaphorical models of thought and speech: a comparison of historical directions and metaphorical mappings in the two domains. In J. Aske, N. Beery, L. Michaelis & H. Filip (Eds), *Procs. 13th Annual Meeting of the Berkeley Linguistics Society.* Berkeley, CA: Berkeley Linguistics Society, pp. 446–459.

Sweetser, E.E. (1990). *From etymology to pragmatics: metaphorical and cultural aspects of semantic structure.* Cambridge, U.K.: Cambridge University Press.

Tomlinson, B. (1986). Cooking, mining, gardening, hunting: metaphorical stories writers tell about their composing processes. *Metaphor and Symbolic Activity, 1* (1), 57–79.