A Belief Representation for Understanding Deception

Gregory B. Taylor
and
Stephen B. Whitehill

Artificial Intelligence Project
Computer Science Department
University of California at Irvine
Irvine CA 92717

ABSTRACT

Identifying deception in stories requires an understanding of the beliefs of the characters. A model must include both beliefs about facts and beliefs about other characters. This paper presents a method for representing such belief structures. Arbitrary levels of embedded beliefs are represented by cvclic structures with Shared common beliefs. With these structures we show how different instances of deception can be recognised with a single deception template. We will illustrate these concepts by applying them to several example stories.

I INTRODUCTION

When Eve caused mankind to be ejected from the Garden, the world became a place where truth could no longer be taken for granted. Not only has deception played a critical part in history, but it is central in the development of fictional stories as well.

For a clear understanding of most stories, an understanding of deception is essential. We must understand how people can have differing beliefs, how they can successfully lie to each other, and why they lie to each other.

Deception ia difficult to understand because it deals with beliefs rather than with factual information. Most of the research done in language understanding has dealt with unquestionable facts (physical actions, attributes of objects, and so on). However, to understand deception, we must model the beliefs of characters. Some work has been done on modeling simple beliefs [1]; however, characters must also have models of other characters" beliefs in order to deceive.

In this paper, we will present a single method for modeling both characters" own beliefs and their beliefs about other characters. We will then show how stories containing deception can be analyzed using these models.

II BELIEF MODELS

In this section, we introduce the belief models that are used throughout the paper. Here, we will use them to model a simple situation with no deception:

Maggie tells Andy that she was once married.

Example 1.

In order to model this situation, we must make some inferences about what people believe after they say or hear something. These default rules of conversation [3] are crucial to the construction of the belief structures. After reading the sentence in example 1, we can make the following inferences:

Maggie was married.
Maggie believes she was married.
Andy believes Maggie was married.
Maggie believes Andy believes Maggie was married.
Andy believes Maggie believes she was married.

These are default inferences; some may not be made if conflicting information is already known.

Notice that the last two inferences contain nested beliefs, that is, beliefs about other beliefs. We could continue the list of possible inferences indefinitely by making each inference more nested than the previous one. The belief structures appear to be infinitely recursive. How many levels of nested beliefs should we model? For example, why not infer "Andy believes Maggie believes Andy believes Maggie was married"? In a notation similar to the one in Bruce and Newman 12], we can infer Andy and Maggie's beliefs (A and M, respectively) and their beliefs about each other's beliefs (AM and MA). However, people have difficulty understanding deeper beliefs such as the complex one shown above in quotes, which would be written as AMA. As we shall see in later examples, it is sometimes necessary to model these recursive beliefs in order to detect deception.

In example 1, models A and AMA can be viewed as being the same belief model. This simply means that Andy thinks Maggie has modeled him correctly; that is, Andy thinks that his beliefs are identical to Maggie's model of his beliefs. To the reader of

this one-sentence story, his view of A and AMA are not merely equivalent, they are the same belief model. For this simple case of mutual beliefs, if we realize that AM - M and MA - A, then we can create a cyclic structure which captures all the information that would appear in the infinitely recursive model.

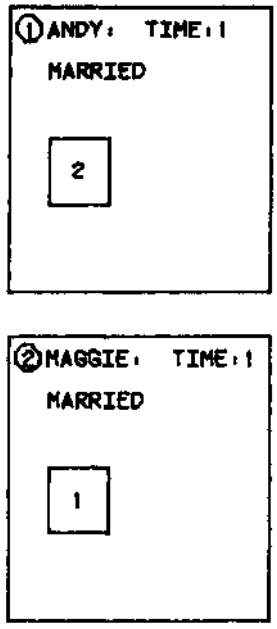

Figure 1.

Figure 1 illustrates this cyclic model. Here, the two belief models are the large boxes with numbers in the upper-left corner. The numbers are illustrative purposes and are not part of the representation. Each model contains a list of facts (only MARRIED in this case) and possibly one or more nested models. These models are the smaller boxes that contain only a single number. This single number is meant to illustrate that this is not a new model, but is merely a reference (in the LISP implementation, a pointer) to the existing model with the same number. In other words, in the static structure shown in figure 1, there are only two belief models.

For the sake of clarity within our figures, we have abbreviated individual facts. For example, the fact that Maggie was married is written as MARRIED. Actually, each fact in the model would be a knowledge representation such as Conceptual Dependency [10].

The structure in figure 1 is one of mutual belief [2J. This is the default belief structure when information is transferred from one person to another. It occurs only when neither person has conflicting knowledge about the subject.

III SIMPLE DECEPTION

Suppose that whan we were reading example 1, we knew that Maggie had never been married. The story would effectively have been:

Maggie has never been married.
She tells Andy that she was once married.

Example 2.

Figure 2 shows the belief models for this modified story. The same inference rules have been used to construct this model as we used in example 1. However, not all of the default inferences applied-to the model. For example, Maggie does not believe that she was ever married, even though she tells Andy she was. However, using the same default rule (if someone says something, he believes it), Andy constructs a model of Maggie that contains the fact that she was married. Andy's beliefs about Maggie's beliefs now differ from Maggie's actual beliefs (AM does not equal M).

Showing that AM does not equal M only yields that an inconsistency in the belief structure exists. A lie is an intentional act. There is deception in this example because Maggie knows she has lied to Andy. Maggie believes that Andy has an incorrect model of her; MAM does not equal M. The belief boxes corresponding to MAM and M are labeled 3 and 1, respectively. Notice that they differ about the 'MARRIED' fact.

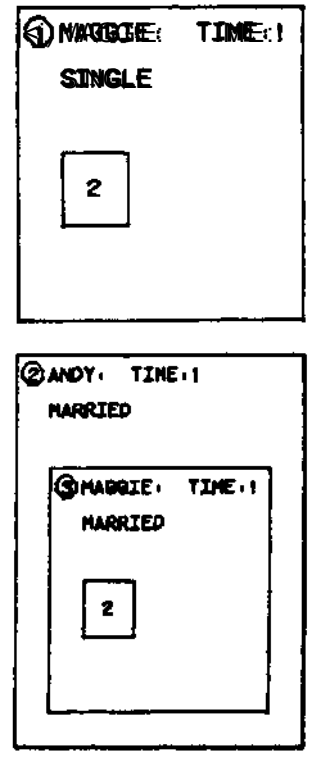

Figure 2.

This example is the pattern which we will hereafter refer to as the deception template. It includes two main ideas; first, the intentionality of the deception (MAM not equal to M) and secondly, the actual success of the deception (AM not equal to M). We will show that this pattern fits all examples of deception.

IV PSYCHOLOGICAL JUSTIFICATION OF THE MODEL

Although the need for some type of nested structure in a belief model is very clear, understanding stories that contain many levels of beliefs can become very difficult. For example, when one tries to analyse the beliefs of another, one can imagine himself as the other person. This involves replacing his top level beliefs with the beliefs he holds about the other person. This process of "getting into someone else's head" is part of the reason why people have difficulty in understanding stories with many levels of beliefs. With these stories, this replacement of top level beliefs occurs several times.

In example 1, A, AHA, and AMAMA describe several applications of this replacement process. However, in terms of the physical structure of the model, these belief boxes are the same model. In spite of this, we are unable to answer questions about AMAMA. This is because even though the structural nesting of AMAMA is only 2 levels deep, its dynamic nesting level (i.e., the number times the above process is used) is 6 levels deep. Our representation allows the dynamic nesting level to increase without bound. However, we believe that this is a very complex process, and for each person, the maximum depth in the model at which the process can take place is the determining factor in their ability to recognize deception.

V MODELING THE READER

The previous model was implicitly that of a reader of the story. If we make this fact explicit in our model, we can see that the author can deceive the reader in precisely the same way that story characters deceive one another.

Consider the following story fragment:

Maggie tells Andy she was married.
Maggie tells Andy she lied about being married.

Example 3.

Notice the representation (figure 3) now includes an outermost model of the reader of the story. Like example 2, this story contains an instance of deception. However, in example 2, the reader knew that Maggie was lying all along. In this example, the reader is as surprised as Andy by Maggie's confession.

Two new items of notation are introduced in figure 3. The two types of arrows (S-arrows and U-arrows) indicate a belief change as a result of new input. The transition from box 3 to box 5 shows that the reader held an incorrect view of Maggie. Maggie has not changed her beliefs; the reader has changed his mind about what Maggie's beliefs have always been. At this point, the reader has admitted to himself that he was wrong about what Maggie believed. Therefore he has (in the language of [6]) supplanted his model of Maggie. The 'S' (supplant) on the arrow indicates this. Notice that Andy's model of Maggie has changed in exactly the same way that the reader's has. This is because they were both deceived by Maggie in the same way. Thus, the reader's model is not the only one that can be supplanted; characters' models can be supplanted, too.

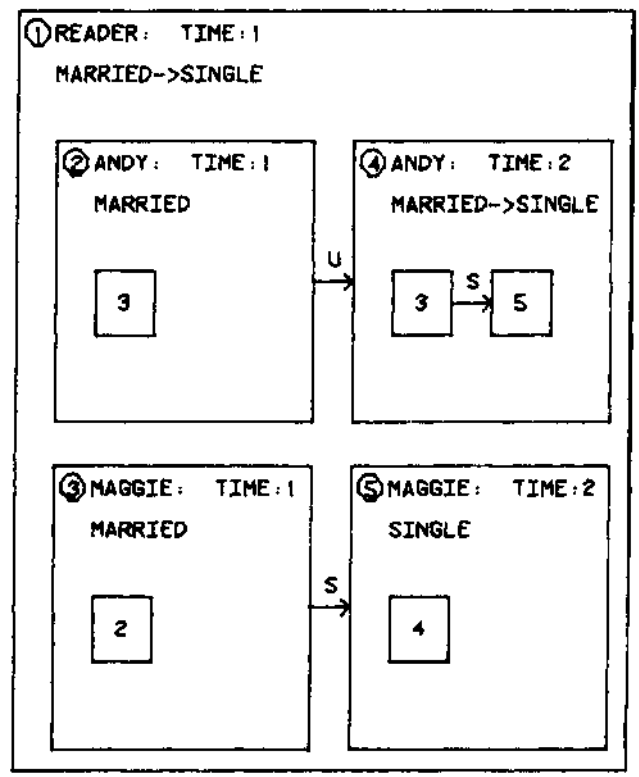

Figure 3.

The transition from box 2 to box 4 indicates that the reader has altered his beliefs about Andy. The reader was never wrong; he has merely recorded the fact that Andy has changed his mind. This is indicated by the 'U' (update) on the arrow. Andy's old belief (that Maggie was married) remains in the current model (box 4).

To summarise, when a belief model is changed because a character was modeled incorrectly, it is supplanted; when a belief model is changed to reflect that a character changed his mind, it is updated.

We will now explain how the simple deception template described earlier allows us to detect the presence of deception in this more complicated model. At the end of the first sentence, there is

a state of mutual belief. In figure 3, this is the cyclic structure involving boxes 2 and 3 (this is the same as figure 1). At the end of the second sentence there as is also a state of mutual belief. This is the structure involving boxes 4 and 5. There appears, at first glance, to be no deception in the model. But notice the "supplant" history in box 4. At the end of the second sentence the reader realizes that he was wrong about the apparent mutual belief that existed after the first sentence. The deception template matches the structure involving boxes 3, 4, and 5. These boxes represent the reader's revised view of the state of affairs after the first sentence. It was the second sentence that gave the reader a clue about what had actually happened.

Let us briefly clarify how boxes 3, 4, and 5 indicate that Maggie has deceived Andy. In box 5, Maggie has a reference to box 4. She believes that at time 1 Andy believed box 3. The left side of the supplant in box 4 refers to a belief held at time 1. Box 3 contains 'MARRIED' while Maggie has always believed 'SINGLE' (from box 5).

This is an example of discovered deception. Before the discovery, at time 1, there is a state of mutual belief. After the deception is discovered, there is a state of mutual belief, but also a realization that the previous state was not one of mutual belief.

VI DOUBLE DECEPTION

Double deception [2] occurs when a person allows another to think that a lie has been successful. An example will clarify:

Maggie tells Andy she was once married.
Maggie tells Andy she lied about being married.
Andy says he knew it all along.

Example 4.

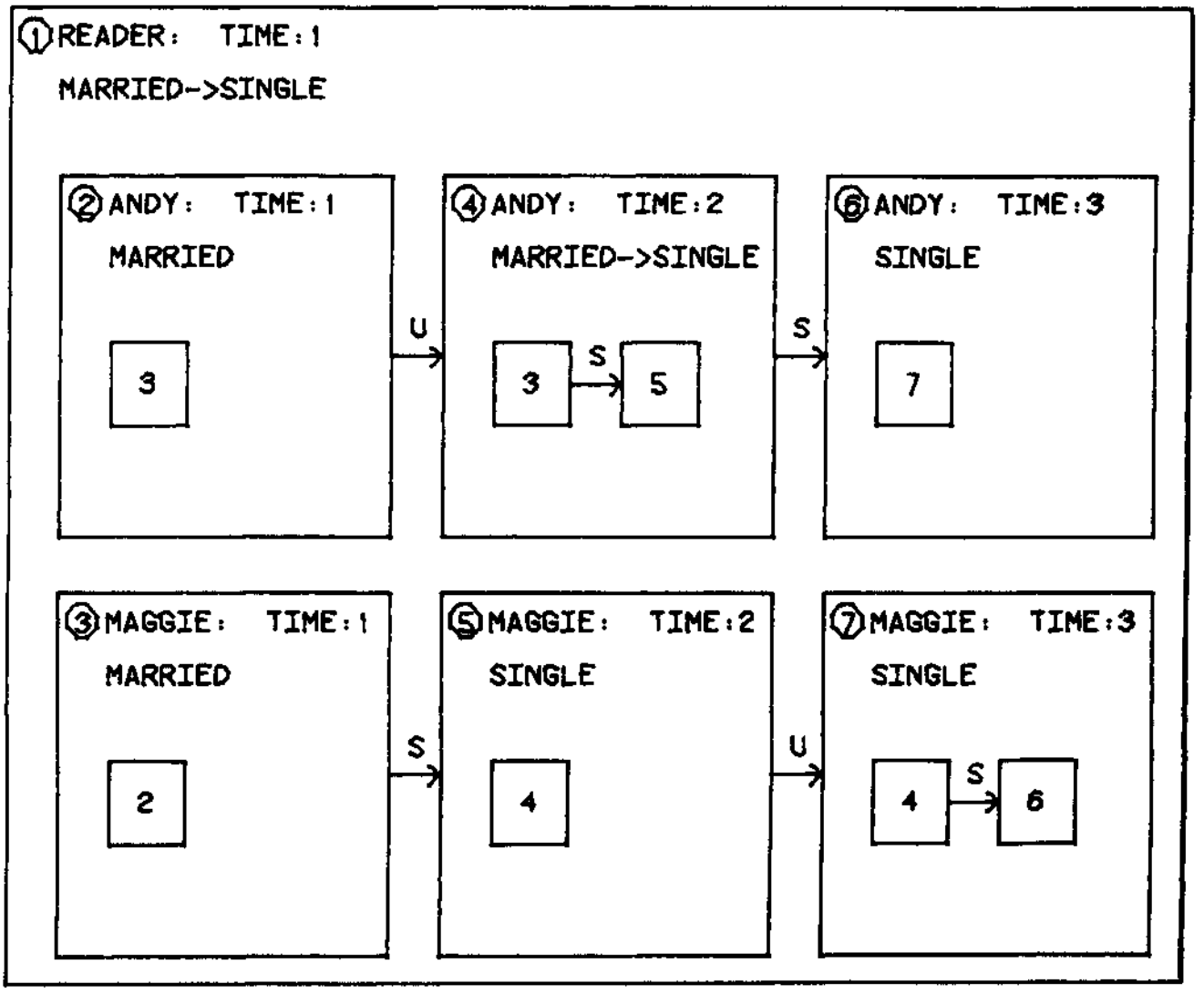

Figure 4.

The first two sentences are the same as those in example 3. In the third sentence Andy admits that he knew of Maggie's deception ail along. In figure 4, this if represented in the reader's model by the transition from box 4 to box 6. The reader has replaced the relatively complicated structure in box 4 (that Andy believed one thing and then changed his mind) with the simple view in box 6 (that he knew ail along).

Maggie changes her model of Andy in precisely the same way. This is because the reader and Maggie are both surprised by Andy's revelation in the third sentence. This illustrates how our model can capture the common elements of the superficially different activities of deceiving the reader and deceiving a character.

VII DETECTING DECEPTION

Because deception can occur at certain times in the story and be noticed at other times, we now describe the use of what we call the "TIME" slot of each belief. The time slot is used to differentiate between similar beliefs held by the same person at different times. When a belief is "created" we place a number in the time slot corresponding to the number of the sentence from which the belief was inferred.

At any point in the story, we can ask questions relating to the supplant/update histories of our model. We may ask, "When did a deception occur in the past?" and "When did you first find out about it?" In example 3, the reader is aware of the deception at the time that it happens. In example 4, the reader finds out later, at time 3, about a deception that occured at time 2. The reader's use of the belief histories that are modeled within each character allows him to answer these questions. The reader may also use his own belief histories to examine what he believed in the past.

VIII READER-AUTHOR DECEPTION

If we want to study how a story with deception is written, we can create models of the author and the reader. Each model has a hypothetical model of the other. As the reader reads the story, he constructs a view of the author in exactly the same way as he models characters in the story. The reader uses essentially the same default inference rules when dealing with the author. However, there are some differences. For example, we know of no case in fictional stories where the author lies directly to the reader. Instead, the author achieves this effect by having a character lie, thereby lying to the reader indirectly, as in example 3, or by setting up a situation with strong inferences that later are found to be incorrect. It is interesting to note that in other forms of written material where fewer restrictions apply, such as propaganda and campaign literature, the author may deliberately lie to the reader.

IX LIMITATIONS OF THE INFERENCING MECHANISM

We initially introduced (in section 2) the inference rules that were necessary to construct all of the given deception examples. This static set of rules may not apply in all conditions. For example, when one suspects that another is dishonest, his inference rules may drastically change. That is, if one believes he is being deceived, he may believe something different from what he hears. If he is being told the truth then he may misconstrue the good intent of the speaker based on his own erroneous inference.

Later, when we try to understand more complicated stories, it will be necessary for each person to have reasons for each of his beliefs. We propose to add what we call 'WHY' tags to each belief. These tags indicate the reasons for the belief. A tentative list of reasons for a belief includes personally experiencing it, being told about it, and inferring it. These tags will aid in tracing bad sources of information and correcting bad inferences.

There is a specific form of deception which has not been considered here. This is when the deceiver take6 advantage of the fact that a person can be led to draw incorrect inferences when he is presented with correct but incomplete data. In this case, the why-tag on the fact of the deception indicates that the incorrect information was inferred and not told by anyone. An area of future research is to make a program answer questions about how a particular character was deceived. This will involve extensive use of the why tags and is beyond the scope of this paper.

X IMPLEMENTATION OF *THE* THEORY

A program is presently being written in UCI MLISP to implement the theory presented here. It is divided into 3 separate parts. The first part will use applications of the inference rules to transform the initial input (in Conceptual Dependency) into the internal representation. This transformation process has been partially described in this paper; however, it has not yet been implemented. The second process consists of recognizing the deception using the deception template. It could also be used as the basis for a question-answering system. This portion of the system is in the implementation stages, as well. The third and only completed portion of the system was used to print the deception diagrams in this paper from the proposed internal belief representation which was explicitly input by hand.

The main goal of our representation has been to adhere to a reasonable theory of memory organisation. The first problem solved is that of nested beliefs and mutually nested beliefs. By using references to belief boxes in creating the cyclic structures, we have economized on the number of beliefs. We create a fixed-sized model which appears to be dynamic when referenced.

Another goal of our representation involves storing knowledge about the past. In [6] we see that a history of inferences is needed in order to understand the intermediate inferences of characters in a story at any arbitrary time. Our representation also exhibits this necessary capability We have shown how time dependent information is "remembered" in our representation and how questions concerning such information can be answered.

Our belief representation is evolving. Presently it is used only to detect deception. We believe that it will be useful in interacting with a planning mechanism. The effects of changing beliefs on plan-creation and goal-setting are widespread. It is our goal to design a belief structure that will satisfy the interaction requirements of a planning mechanism so that we can achieve a better understanding of the underlying reasons for deception.

ACKNOWLEDGEMENTS

We would like to thank Rick Granger for his helpful guidance in this research, Jim Meehan for useful information on belief systems, and Stephen Willson for his participation in the early stages of this research.

REFERENCES

[1] Abelson, R. "Differences between belief and knowledge systems", Cognitive Science Technical Report 1, Yale University, 1980.

[2] Bruce, B. and Newman, D. "Interacting plans", Congnitive Science 2 (1978) 195-233.

[3] Bruce B. "What makes a good story?" Reading Education Report 5, Bolt, Beranek and Newman, Cambridge, Massachusetts, 1978.

[4] Carbonell, J. Subjective Understanding: Computer Models of *Belief Systems* Ph.D. thesis. Yale Computer Science Department Research Report 150, 1979.

(5] Cohen, P. January on Knowing What to Say: Planning Speech Acts Ph.D. thesis. Department of Computer Science, University of Toronto, Technical Report 118, 1978.

[6] Granger, R. "When expectation fails: Towards a self-correcting inference system." Proceedings of the First National Conference on Artificial Intelligence. Stanford, CA, 1980.

[7] Meehan, J. The Metanoval: Writing Stories *by* Computer. Ph.D. thesis. Computer Science Department, Yale University, 1976.

[8] Minsky, M. Semantic Informative Processing Massachusetts Institute of Technology, 1968.

[9] Perrault, C. and Cohen, P. "Planning speech acts", AI-Memo 77-1, Department of Computer Science, University of Toronto, 1977

[10] Schank R. Conceptual Information Processing North Holland, Amsterdam, 1975.

[11] Schank, R. and Abelson R. Scripts. Plans. Goals. and Understanding. Lawrence Erlbaum Associates, Hillsdale, N.J, 1977.