Reasoning on a Highlighted User Model to Respond to Misconceptions

Kathleen F. McCoy

Department of Computer and Information Sciences
University of Delaware
Newark, DE 19716

Responses to misconceptions given by human conversational partners very often contain information refuting possible reasoning which may have led to the misconceptions. Surprisingly there is a great deal of regularity in these responses across different domains of discourse. For instance, one reason a user might have given an object a property it does not have is that the user confused the object with another similar object. In correcting such a misconception, a human conversational partner is likely to point out this possible confusion.

This work describes a method for generating responses like the one just described by reasoning on a highlighted model of the user to identify possible sources of the error. Through a transcript study a number of response strategies were abstracted. Each strategy was associated with a structural configuration of the user model. For example, the above mentioned strategy of pointing out a similar confused object is associated with a configuration of the user model that indicates the user believes there is an important similar object that has the property involved in the misconception. Upon finding that configuration in the highlighted user model, the system can respond with the associated strategy.

Notice that the reasoning must be done on a highlighted user model since the perception of both an object's importance and its similarity with another object change with the perspective being taken on the domain. This paper investigates how domain perspective can be modeled to provide the needed highlighting and introduces a similarity metric that is sensitive to the highlighting provided by the domain perspective. Finally, the paper shows how the highlighting affects misconception responses.

1 Introduction

When people interact with a database or expert system, it is reasonable to expect that they might reveal a misconception about an object modeled by the system. Since a human conversational partner would correct such a misconception if it was important to the current goals of the conversation, our database and expert systems should also be equipped with this ability.

In order to investigate how the process of correcting misconceptions might be automated, a study of transcripts of both humans interacting with what they thought were expert systems (Malhotra 1975, Malhotra and Sheridan 1976, Schuster 1982), and humans interacting with other humans to achieve some goal (Pollack, Hirschberg, and Webber 1982) was undertaken. The transcripts, which varied greatly in their domains of discourse, were analyzed to determine if there was any regularity in the content and rhetorical force of re-sponses given to misconceptions. The intention of this analysis was not to mimic the actual behavior found in the transcripts, but to use them as a source of intuitions about the context and textual shape of responses as well as the process of generating them.

The study revealed that a response to a misconception important to the current discourse goals of the participants can be viewed as consisting of three parts: 1. a denial of the incorrect information; 2. a statement of the correct information; and 3. justification for the denial and correction given. For a particular type of misconception (i.e., one involving a particular kind of knowledge), variations in responses could be found in the form of the justification given. The justification often seemed to refute support that might have led to the misconception. While the kind of support someone might have for a misconception seems unrestricted, the form of the justification was limited for misconceptions

involving a particular kind of knowledge. A large number of responses found could be accounted for by a small number of correction strategies based on the kind of justification given (and hence the faulty reasoning refuted).

If a principled reason for using one strategy over another could be developed, these strategies could be used by a natural language generation system for responding to a misconception. In this work the faulty reasoning refuted by several of the found correction strategies is characterized in a domain-independent fashion in terms of the user's beliefs about the domain. Therefore, given a highlighted model of the user's beliefs about the domain, a generation system can look for possible support for the misconception. The response strategy that refuted the kind of support found could then be instantiated. Notice that the domain-independent characterization of the faulty reasoning enables the same strategies and same method for choosing a strategy to be used given a highlighted user model for any domain.

It is crucial that the user model given to the misconception corrector be highlighted by previous discourse since the kind of response given by a human conversation partner is apparently not only dependent on the beliefs about the person being corrected, but also on the context in which the misconception occurred. For instance, we could imagine the following dialog where the user exhibits the misconception that T-bills have a penalty. A reasonable response is shown.

U: I am interested in investing in some securities to use as savings instruments. I want something short-term and I don't have a lot of money to invest, so the instrument must have small denominations. I am a bit concerned about the penalties for early withdrawal. What is the penalty on a T-Bill?

R: T-Bills don't have a penalty. Were you thinking of Money Market Certificates?

This response might be prompted by R thinking that U came to the misconception by confusing T-Bills with Money Market Certificates.

On the other hand, it is reasonable that the response might be different given a different preceding dialog. For example:

U: I am interested in investing in some securities. Safety is very important to me, so I would probably like to get something from the government. I am a bit concerned about the penalties for early withdrawal. What is the penalty on a T-Bill?

R: T-Bills don't have a penalty. Were you thinking of T-Bonds?

This response may have been prompted by R thinking that a confusion between T-Bills and T-Bonds was the source of the misconception. The two different re-

sponses to the same misconception suggest that the user model should be influenced by previous discourse. I will show how part of this influence can be achieved by a highlighting due to the perspective being taken on the domain.

Currently, when a user model containing what the system takes to be the user's model for the domain is accessed, all user knowledge has equal importance. When people engage in a conversation, however, certain aspects of their domain model become more important than others. This importance is more than just a highlighting of those things that have been explicitly mentioned. Rather, certain things that are somehow related to those things explicitly mentioned in previous discourse are also highlighted. In fact, an orientation on the domain is usually established.

In section 8 I will introduce a notion of object perspective that will enable this highlighting effect of previous discourse to be incorporated into the user model. Thus when the user model is accessed, certain things in it will be highlighted while other things will be suppressed. I will show how this variable highlighting of the user model can explain why the response to a particular misconception by a particular user may vary.

2 RELATED MISCONCEPTION WORK

The method of correcting misconceptions outlined above should be contrasted with the way that misconceptions are handled by AI systems today. For the most part misconceptions have been left to the Intelligent Computer Aided Instruction systems, which basically use an a priori listing of misconception-response pairs (see, e.g., Brown and Burton 1978; Stevens, Collins, and Goldin 1979; Stevens and Collins 1980; Woolf and McDonald 1983; Woolf 1984). The major problem with these systems is due to their inability to reason about the misconception itself, they are completely at a loss when faced with a misconception absent from their a priori listing.

The work of Sleeman (1982) on inferring defective algebra rules (mal-rules) is based on the observation that the a priori listing of misconceptions is a difficult, if not impossible, task. Sleeman proposes on-line inference of mal-rules based on the answer the student has given to a particular problem. Although Sleeman's work is a major improvement over the a priori listing approach, it still has several problems. First, there is no measure of how *reasonable* or *likely* a particular mal-rule is. In addition, once a mal-rule has been inferred, no indication is given concerning how the misconception should be corrected.

The work of Kaplan (1979) and Mays (1980) is closer to the work described here in that they were concerned with handling and reasoning about whole classes of misconceptions, thereby giving the system the ability to handle a potentially infinite number of misconceptions. Kaplan and Mays were concerned with responding to

certain types of misconceptions in the context of a natural language interface to a database system. They worked on detecting and correcting such misconceptions based on domain independent linguistic cues from the user and an enhanced model of the domain. For instance, the query "Which faculty take courses?" indicates a *presumption* failure. A truthful response of "none" to this query would confirm the user's erroneous belief that faculty can take courses. Mays suggests correcting a query like the one above by 1. denying that the "take" relation can hold between faculty and courses, and 2. describing all correct alternatives which can be reached by abstracting on each of the objects and the relation involved. This method would produce the following kind of response:

R: I don't believe that faculty can take courses. Faculty teach courses. Students take courses.

Although responses such as this would probably be helpful to the user, they have the potential (given a complicated domain) for being overly verbose and containing information that the user does not care about. One goal of this work is to provide a more pointed and natural response to these same kinds of errors.

3 KNOWLEDGE AVAILABLE

The work being done here is in the context of a natural language interface to a database or expert system. It is an attempt to define a module of an interface that could generate a cooperative response to a misconception— the kind of response that would be generated by a helpful human conversational partner. In this work, a *misconception* is defined to be some discrepancy between system beliefs and user beliefs (as exhibited through the conversation). Upon encountering such a misconception, the assumption is that the system knowledge is correct, and therefore the job of the misconception corrector module is to attempt to bring the user's knowledge into line with the system's knowledge.

The scope of this work is limited by several assumptions about the kind of knowledge available.

• The system's model of the world contains an object taxonomy with attribute/value pairs attached to the objects.
• The system has available to it a user model that includes the user's beliefs about the world.[1] This is what the system takes to be the user's model of the domain. Although the content of the user model and the system's model of the world may differ greatly, the user model is in the same form as the system's model of the world. Thus while both of these models contain an object taxonomy with attribute/value pairs attached to the objects, the set of objects in the taxonomies and the way these objects are classified, as well as the particular attribute/value pairs associ-

ated with an object, may vary. This model of the user may be updated as the conversation progresses.
• The system has available to it certain pieces of contextual and discourse information that serve to highlight the user model. This highlighting (explored below) is gained from a new notion of object perspective and from a record of items and attributes which have been explicitly focused on in the discourse.

Given the kind of information assumed in the system's and user's models of the world, there are two kinds of misconceptions that may occur: misclassifications (a user may classify an object wrong) and misattributions (a user may give an object an attribute/value pair it does not have). For each of these kinds of misconceptions, a small number of response strategies[2] were found in the transcript study. These were abstracted into response schemas. In section 4, examples of the response strategies found for misclassifications will be examined. For each strategy an abstract specification of the content will be given. Next we will look at what beliefs about the user might have prompted the use of each strategy, and characterize these beliefs in terms of the structure of a highlighted user model. With this pairing of user model structures and response schemas, a misconception can be responded to by looking in the highlighted user model for one of the user model structures and instantiating the associated schema. Section 5 examines response strategies for misattributions. The sections following that will concentrate on the highlighting from object perspective. A new notion of object perspective will be defined and it will be shown how object perspective aids in generating context sensitive responses to misconceptions.

This paper is concerned with reasoning on the user model to decide how to respond to a misconception. It is not concerned with inferring the user model; it is assumed that the user model is already available. The emphasis in this work is on using the user model, in a domain-independent fashion, to respond to a misconception in a manner similar to a human conversational partner's response.

The methods described here have been implemented in the ROMPER system (Responding to Object-related Misconceptions using PERspective). The system takes as input a specification of the information that is inconsistent with the system's model of the world, the current perspective (described below), and a record of past focus. It produces a formal specification of the response. This response specification is passed into the Mumble system (McDonald 1980), which, using a grammar and dictionary written by Robin Karlin (1985), produces an actual English response.

The implemented system works on the financial securities domain. In order to show the generality of this approach, two different domains will be used in this paper. The motivation for the system's method of choosing a response strategy will use examples from the domain containing whales and fish. In the last sections

of the paper (those pertaining to the new notion of object perspective), the examples will be taken from the domain of financial securities, since the ideas can be better motivated with a more complex domain.

4 MISCLASSIFICATIONS

One kind of misconception concentrated on in the transcript study was that involving the classification of an object. From that study three major correction strategies, which could be supported with the kind of knowledge available in our knowledge base, were abstracted. These are used to illustrate the type of analysis advocated; no claim is being made about the completeness of this set. The first strategy is exemplified by the following dialog.

U: I thought whales were fish.

R: No, they are mammals. You may have thought they were fish because they are fin-bearing and live in the water. However, they are mammals since, (while fish have gills) whales breathe through lungs and feed their young with milk.

Let us first examine the content of this response. It can be seen as consisting of three parts. In the first sentence we have the denial of the incorrect information and the statement of the corresponding correct information. The remaining sentences comprise the justification for the denial and correction given.

The content of this response can be abstracted into the following:

$$MISCONCEPTION = X \text{ is-a } Y$$

$$RESPONSE =$$

1. X is-NOT-a Y
2. X is-a Type-of(X)
3. X is like Y because both share attributes-of(X) $\cap$ attributes-of(Y)
4. BUT X has attributes-of(X) — attributes-of(Y)
5. WHILE Y has attributes-of(Y) — attributes-of(X)

The justification is contained in (3–5) of the rule. Notice that it can be seen as a concede/override pair. First, information is conceded that actually supports the misconception. This consists of attributes that whales and fish have in common. These attributes are a potential source of the misconception since an object that shares many attributes with a superordinate is likely to belong to that superordinate class. One way to view this conceding of intersecting attributes is as R acknowledging the apparent support for the misconception lest U think that that part of his/her domain model is wrong.

The response goes on to override the conceded information by offering attributes that differentiate whales and fish. These are attributes that whales have but fish cannot, and attributes that fish have that whales do not.

By analyzing the justification in this way, not only do we get an abstract representation that can be used by a generation system, but we get some insight into when it would be reasonable to use such a response. It makes sense to use a response that first concedes a similarity between the misclassified object and the object that was wrongly given as its superordinate (from now called the *posited superordinate*) and then overrides that conceded information, when such a similarity is a possible source of the misconception. This might be the case when *the user model shows that the user believes that such a similarity exists*. Thus a generation system may generate its response by instantiating the above rule if the user has misclassified the object X as a Y, and an analysis of the user model shows that the user believes X is similar to Y. Because of this reason for using this response strategy, it has been termed the *like super* response strategy.

A second kind of strategy is exemplified by the following dialog. Imagine a recent conversation centered on sharks being large fish that people are typically afraid of. Following that, the conversation turns to classifying whales. The user says:

U: I thought whales were fish, too.

R: No, they are mammals. You may have thought they were fish since they are like the fish, sharks, in that both are large aquatic creatures and both scare people. However, whales are mammals since, (while fish have gills) whales breathe through lungs and feed their young with milk.

Again, this response can be seen as consisting of a denial, a correction, and a justification. The denial and correction are contained in the first sentence. As in the previous example, the remaining sentences comprise the justification in the form of a concede/override pair. The conceded information, while dependent on attributes, differs from that given above. The abstract content rule can perhaps best illustrate this difference:

$$MISCONCEPTION = X \text{ is-a } Y$$

$$RESPONSE =$$

1. X is-NOT-a Y
2. X is-a Type-of(X)
3. $\exists Z$ (Z *is-a* Y $\wedge$ X is like Z because both have attributes-of(X) $\cap$ attributes-of(Z))
4. HOWEVER X is-a Type-of(X) because X has attributes-of(Type-of(X)) — attributes-of(Y)

The major difference between this response and the like super response is that instead of conceding a similarity between the misclassified object and the posited superordinate itself, a similarity between the misclassified object and a descendant of the posited superordinate is conceded. This response is reasonable if the user came to the misconception by reasoning: "I don't know how to classify whales, but I do know how to classify sharks.

Sharks are fish and whales are very similar to sharks. Therefore it is reasonable to believe that they are classified the same." Evidence for this kind of reasoning might be a similarity found in the user model between the misclassified object and some descendant of the posited superordinate. Finding such a similarity would justify the use of this *like-some-super* strategy.

A final strategy found for misclassifications is exemplified by the following:

U: I thought whales were fish.

R: No, they are mammals.

This is the degenerate *no support* strategy, which contains no justification for the correction. At first glance the use of this strategy seems rather puzzling. Since I have assumed that the misconception is important, one would expect a full response to be given. However, no justification was included in the response.

This last response is an example of a response in a situation where no evidence was found in the highlighted user model for any of the preidentified kinds of faulty reasoning. If R could not figure out what led to the misconception, then R could not include justification to refute the support! The most that R could do in this situation was deny the incorrect information and state the corresponding correct information. This is exactly what was done in the response above. For this reason the response has been termed the no support response, and it can be used when the user model contains none of the evidence that would key off any of the other correction strategies.

5 MISATTRIBUTIONS

Misattributions are the second kind of misconceptions that may arise given the kind of knowledge assumed. A misattribution occurs when the user gives an object an attribute/value that the object does not have in the system's domain model.

The transcript study revealed that a misattribution response could also be broken into three parts. In this case, however, the justification took the form of offering another statement that would change what the user said in a small way in order to make it a correct statement. This is illustrated in the following dialog, which is an example of the first kind of correction strategy found. Imagine that we are discussing sea animals and have previously discussed sharks and their attributes. We move to whales.

U: I thought whales had gills.

R: No, they breathe through lungs. Were you thinking of sharks?

The response might go on to discuss the similarities and differences between whales and sharks.

Notice what was done in this response. We have the denial and the correction as before. This time the

justification offers another object that is similar to the object involved in the misconception but for which the misattributed attribute/value pair is true. This content is captured in the following rule:

MISCONCEPTION = X has attribute Y with value V

then RESPONSE =

1. X NOT have attribute Y with value V
2. X has attribute Y with value W
3. ∃ Z (Z has attribute Y with value V), did you mean Z?

This kind of strategy would be effective if the user's misconception results from either confusing the object being discussed with some other object with the misattributed property, or through an incorrect analogy with such an object. Mention of the object in the case of a confusion prevents continuation of the confusion. In the case of an incorrect analogy, mention of the other object alerts the user to the inappropriateness of analogy and also prevents the continued usage of the analogy. Notice that in either case, a failure to mention the second object will probably bring protest on the part of the user. For instance, suppose that the preceding misconception was arrived at because the user was confusing whales and sharks. If the expert just responds, "No, they breathe through lungs", it is likely that the user will counter, "But you just told us that they had gills!" By mentioning the confused object, the user becomes aware of the confusion.

It makes sense to use this *wrong object* strategy when a possible source of the misconception is either an object confusion or the usage of a bad analogy. Evidence for this would be the existence of some similar object from the highlighted user model that has the misattributed property.

A second kind of correction strategy used for misattributions is shown here:

U: I thought whales had gills.

R: No, whales breathe through lungs. Did you mean you thought they had fins?

This is very similar to the last response except, instead of offering another object, the strategy offers a similar attribute which makes the user's statement correct. The content rule for this response follows:

MISCONCEPTION = X has attribute Y with value V

then RESPONSE =

1. X does *not* have attribute Y with value V
2. X has attribute Y with value W
3. ∃Q∃U(X has attribute Q with value U ∧ similar(Y,Q), did you mean X has attribute Q with value U?)

This strategy, termed the *wrong attribute* strategy, is used when there is a similarity of attributes within the

highlighted user model. (The no support strategy also occurs in the case of misattributions.)

6   User Model Analysis

Given the association of user model configurations with response strategies given in the previous sections, we can come up with a method for deciding how to respond based on the type of misconception along with an analysis of the highlighted user model. Basically the user model analysis looks for one of the preidentified configurations and, if one is found, suggests instantiating the associated strategy. The following rule captures the way that the ROMPER system implements what has been discussed so far.

> IF misconception = "X is-a Y"
> THEN
> IF similar(X,Y)
> THEN instantiate like super schema
> ELSEIF $\exists Z$ (Z is-a Y) $\wedge$ similar(X,Z)
> THEN instantiate like-some-super schema
> ELSE instantiate no support schema
> ELSEIF misconception = "X has attribute Y with value V"
> THEN
> IF $\exists Z$ ((Z has attribute Y with value V) $\wedge$ similar(X,Z))
> THEN instantiate wrong object schema
> ELSEIF $\exists Q \exists U$ ((X has attribute Q with value U) $\wedge$ similar(Y,Q))
> THEN instantiate wrong attribute schema
> ELSE instantiate no support schema

Notice that each of the tests for instantiating a schema hinges on the similarity assessment of two objects. These assessments must be context dependent. Because of this, it is crucial that the user model analysis be done on a user model highlighted by previous discourse and that the similarity metric take advantage of this highlighting. The highlighting and similarity metric used by ROMPER will be discussed below.

This method for correcting misconceptions suggests a model of natural language generation that is similar to that put forth by McKeown (1982) but which differs from McKeown's model in several ways.

Both McKeown and this work concentrate on determining the content and textual shape of a response. McKeown is concerned with responses to questions about the structure of a data base. Upon encountering such a question McKeown first delimits a *relevant knowledge pool* using fairly simple mechanisms. This relevant knowledge pool contains that information from the knowledge base that could possibly be included in the response; the actual generated response need not exhaust this pool. Next, based on the goal of the discourse (as determined by the question type) and a characterization of the relevant knowledge pool (again, a *simple test*) a response schema is chosen for the response. The response schema dictates the textual structure of the response. This schema is filled by stepping through it and matching its predicates against the relevant knowledge pool. A focusing mechanism is used to mediate between choices arising during this process.

The schemas used by the ROMPER system are more complicated than those advocated by McKeown. In effect, ROMPER's schemas are responsible both for determining the textual shape of a response and for determining what information from the knowledge base to include in the response. Thus they are applicable to a much more restricted generation problem (e.g., responding to a misconception of a particular type). Because of this, the test for determining which schema to use can be much more specific than the tests employed by McKeown.

7   Highlighting and Object Similarity

We claim that in order for the above strategy for correcting misconceptions to work, the similarity metric that is used to assess object similarity must be affected by the preceding discourse.

To date, most AI systems do not assess object similarity in a way that is context dependent. Several systems that do assess object similarity (Rumelhart and Abrahamson 1973, McKeown 1982, Carberry 1984, Weiner 1984) use a metric based on distance in some space. Most often, this space is the generalization hierarchy. Basically, two objects that have a common immediate superordinate (i.e., are siblings in the hierarchy) are seen as very similar, while objects whose lowest common ancestor is several levels up in the hierarchy are seen as quite different.

One problem with this metric arises when objects can be classified in more than one way and there are several lowest common ancestors of the objects being compared. A decision must be made about which of these lowest common ancestors should be considered since the similarity assessment of the objects might vary widely as a result. For instance, a treasury bond and a corporate bond may be assessed as being very similar since they have a common immediate superordinate of bonds. On the other hand, both of these objects can be classified along other dimensions. If we look at a treasury bond as being a treasury issue, which is a type of US Government security, it seems quite different from a corporate bond, which is a corporate security.

A second major problem with a similarity metric based on distance in the generalization hierarchy is that it is *context invariant*; contextual information has no way of affecting the assessments. As shown by Tversky (1977) and others, human judgments of object similarity have been found to shift both when the set of objects under discussion are altered (e.g., a violin and an electric guitar may be judged quite similar when in a group with a clarinet and an oboe, and may be judged quite different when the other members of the group are

a cello and an electric bass), and when the salience of attributes are altered (e.g., in a group containing a red triangle, a blue triangle, and a red square, the red triangle might be judged similar to the blue triangle when attribute shape is stressed, but may be judged similar to the red square when attribute color is stressed).

One metric that avoids these problems was introduced by Tversky (1977). Tversky's metric, rather than relying on distance in some space, is based on the common and disjoint features of the objects involved. The metric, termed a contrast model, allows context to be taken into account in several places.

Suppose we have two objects a and b where A is the set of properties associated with object a and B is the set of properties associated with object b. Tversky's measure can be expressed as:

$$s(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$$

for some θ, α, and $\beta \geq 0$.

In the above equation θ, α, and β are parameters which represent the importance of each piece of the equation. The function f maps over the features and yields a salience rating for each. In essence, the contrast model states that the similarity of two objects is some function of their common features minus some function of their disjoint features. The importance of each particular feature involved (determined by the function f) and the importance of each piece of the equation (determined by θ, α, and β) may change with context.

Although Tversky discusses in general terms how these functions might be set, he gives no concrete methods for doing so. For instance to set θ, α, and β he turns to the relative prominence of objects a and b in the discourse. The more prominent an object is, the more its attributes should have an impact on the similarity rating. Thus, finding the relative prominence of objects a and b in the discourse would help set these values. If a is relatively more important, then functions θ and α should be greater than β resulting in the attributes of the more prominent object having a greater influence over the similarity assessment. While I would conjecture that information about the focus of the discourse (Grosz 1981, Sidner 1983, Grosz, Joshi, and Weinstein 1983) might give an indication of an object's prominence and would therefore be useful in setting the values of θ, α, and β, in this work I have assumed a value of 1 for the θ, α, and β and have concentrated on setting the f function.

We turn, then, to the problem of finding a value for the f function: the measure of salience for each property of the objects involved. Other work, such as Carbonnell and Collins (1970) and Weiner (1984), has hand-encoded salience values for attributes of individual objects directly into the knowledge base, permanently setting the f function. This approach is not sufficient for setting the f function for Tversky's metric, since it is crucial that the f function be able to change with context. In order to

make this happen, the salience values computed by f must change with context.

To see this, consider our ability to explicitly mention an attribute to increase its salience. In the example earlier with the red and blue triangles and the red square, if the request for a similarity judgment had been preceded by "Look at the pretty colors of these objects", the red triangle and red square would have probably been judged to be more similar than the red triangle and the blue triangle. Thus we see that explicit mention of an attribute in a discourse is one way that the f function might be affected by previous discourse.

Explicit mention of an attribute is only one way in which the salience of the attributes may change dynamically. Another aspect of dynamic salience comes from the point of view, or perspective, applied to the domain. For instance, a building can be referred to as being an architectural work, for example, or as being someone's home. The two different views of the building cause different sets of attributes to become salient. Notice that this set of attributes is in addition to the attributes that have been explicitly mentioned in the discourse. From an architectural work point of view, attributes like the architect's name, date of building, and particular architectural features become salient. On the other hand, from the home point of view, attributes like the kitchen size, number of bedrooms, and living space become important. If we can find a way of modeling how these "precompiled" sets of attributes become highlighted in a discourse, we will have a principled method for setting the f function needed for Tversky's similarity metric. The next section discusses how this highlighting can be modeled by a computer system.

8 OBJECT PERSPECTIVE

The notion of point of view or object perspective has been noted by other researchers in artificial intelligence. Perspective's ability to explain the changing attribute salience has been attributed to a limited inheritance mechanism (see, e.g., Grosz 1977; Bobrow and Winograd 1977; Tou et al. 1982). An object viewed from a particular perspective is seen as having one particular superordinate, although in fact it may have many. The object inherits properties only from the superordinate in perspective. Therefore different perspectives on the same object cause different properties to be inherited (and therefore highlighted).

While explaining object perspective via a limited inheritance mechanism is intuitively appealing, it is unable to handle some effects which intuitively should be handled by object perspective. The first has to do with the availability of object attributes. A limited inheritance mechanism makes attributes inherited from superordinates other than the one in perspective unavailable. This seems a bit too strong. When we discuss a building as an architectural work, I may comment on the number of bedrooms the building has. While you

may think my comment irrelevant to the current conversation, you would still be able to understand it and even evaluate its truth or falsity. In the limited inheritance account of object perspective, however, this would not be possible. As far as the system would be concerned the number of bedrooms would not even be an attribute to the building.

A second problem with the limited inheritance account of object perspective has to do with deciding what attributes the superordinate in perspective has. The superordinate itself may have multiple classifications and thus potentially multiple perspectives. Thus, in order to figure out what attributes a particular concept should inherit, we must figure out not only what perspective it is being viewed from, but also what perspective the perspective superordinate is being viewed from, and so on. But this seems to be much more work than is necessary.

Another problem is that a limited inheritance mechanism explains the perspective for a single object only. However, during the course of a conversation it is usually the case that more than one object will be discussed. When this happens, usually the same kinds of things are discussed about the objects. In essence, a particular highlighting of attributes (or point of view) seems to be in force during the conversation. Yet, this highlighting is applied to different objects—some of which may not even have the same superordinates. What seems to be happening is that the conversational partners are viewing an entire group of objects from the same perspective. A limited inheritance mechanism cannot account for this unless each of the objects under discussion can be said to have the same (immediate) superordinate.

A final effect that is not accounted for by the limited inheritance mechanism, yet seems to hinge on the view being taken on the domain, has to do with the heightened importance of some objects during a discourse. Like the importance of attributes, the relative importance of some objects in the discourse cannot solely be accounted for by explicit mention. Some objects are more likely to be mentioned and discussed in a discourse than others. For instance, when discussing a particular building as an architectural work, I might reasonably mention the library down the street that was designed by the same architect. On the other hand, I will probably not mention my apartment. Along the same lines, in discussing that building as a home, my apartment is a likely candidate for mention in the conversation. Although this effect seems to be in some way tied to the notion of object perspective, the limited inheritance mechanism does not address this issue.

We want to retain the dynamic highlighting of "precompiled" groups of attributes. Instead of the limited inheritance mechanism, we propose that the following account be used:

1. Instead of tying perspective into the generalization hierarchy of objects as has been done in the past, the new notion of perspective is independent of that hierarchy. Perspectives that can be taken on the objects in the domain will be defined and will sit in a structure that is orthogonal to the generalization hierarchy.
2. A number of perspectives are available for any domain of discourse and *any* given domain object may be viewed from any one of several perspectives for that domain.
3. Each perspective comprises a set of attributes with associated salience values. It is these salience values that dictate which attributes are highlighted and which are suppressed.
4. One such perspective is designated *active* at any particular point in the discourse.[3]

Our solution is that any object that is accessed by the system is viewed through the current *active* perspective. However, instead of dictating which attributes an object inherits, the active perspective affects the salience values of the attributes that an object possesses (either directly or inherited through the generalization hierarchy). The active perspective essentially acts as a filter on an object's attributes. By raising the salience of the attributes, it highlights those attributes which have a high salience rating in the active perspective. By lowering the salience of the attributes, it suppresses those attributes that are either given a low salience value or do not appear in the active perspective.

By defining object perspective in this way, we have retained the desirable results of the limited inheritance account of object perspective while avoiding its problems. In addition, since any object accessed by the system is viewed through the active perspective, we gain the feeling of perspective on the entire domain. The object importance aspect of perspective is gotten by saying that those objects that contribute attributes which are highly salient to a perspective are important while that perspective is active.

We propose that the f function in Tversky's metric be set by taking into account the salience values derived from the active perspective. This would yield an f function that is context dependent and would help the similarity metric exhibit many desirable properties.

9 Modeling a Domain with Perspectives

In some natural language systems, a model of a particular domain includes a usual object taxonomy containing all of the objects in the domain and all of the attributes associated with those objects. We will show an example of building a domain model with perspectives. In order to do this, one must build the domain model as usual. In addition, the perspectives that can be taken on the domain objects must be defined. The result of viewing the domain model through the perspectives will be shown.

Money Market Certificates
Maturity: 3 months
Denominations: $1,000
Issuer: Commercial Bank
Penalty for Early Withdrawal: 10%
Purchase Place: Commercial Bank
Safety: Medium

Treasury Bills
Maturity: 3 months
Denominations: $1,000
Issuer: US Government
Purchase Place: Federal Reserve
Safety: High

Treasury Bond
Maturity: 7 years
Denominations: $500
Issuer: US Government
Penalty for Early Withdrawal: 20%
Purchase Place: Federal Reserve
Safety: High

Figure 1. Objects in "Flat" Domain Model.

Savings Instruments
Maturity—1.0
denominations—1.0
safety—0.5
yield—0.5

Issuing Company
issuer—1.0
safety—1.0
purchase-place—0.5
yield—0.5
tax—0.5

Figure 2. Sample Perspectives.

Figure 1 shows a small piece of a typical domain model. The domain is that of financial securities. Three of the objects from this domain, Money Market Certificates, Treasury Bills, and Treasury Bonds, are shown with the attributes they possess. In systems as they are defined today, a group of objects defined in this way and arranged in a generalization hierarchy would constitute the domain model. If a system were to access any one of the objects in the domain it would be given the object with the attributes as listed in the figure.

In order to get a dynamic highlighting of the domain model, we must build, in addition to the object taxonomy, a separate structure containing the perspectives that can be taken on the domain objects. This means that we must think about the different points of view that can be taken on the objects in the domain and compile sets of attributes from our model which capture the important domain concepts in that point of view. There will be attributes in each perspective that do not occur with all of the objects in the domain. At the same time, there will be attributes of individual objects that do not appear in a particular perspective. The perspective simply captures the attributes that are important in a particular point of view.

Figure 2 contains two perspectives that might be reasonable for this domain (here we are assuming salience values from low salience of 0 to high salience of 1). The perspective of Savings Instruments highlights maturity and denominations, and somewhat highlights safety and yield. This indicates that when people are discussing securities as savings instruments, they are most interested in how long their money will be tied up and in what denominations they can save their money. The perspective of Issuing Company, on the other hand,

highlights different attributes. When securities are discussed from this perspective, things like the name of the company and the stability of an investment in the company become important. Other attributes of the securities are ignored (recall that attributes not mentioned in the perspective get assigned a low salience rating).

Notice how the objects look when accessed, depending on which of the two different perspectives are active. For instance, through the savings instruments perspective the objects' attributes take on the salience values shown in Figure 3. Attributes are not shown in the figure that have 0 salience. When no static salience is included in the domain model, the attributes of the objects derive their salience directly from values given in the active perspective. Attributes in the perspective that the objects do not have (e.g., yield) are ignored. Attributes of the objects not occurring in the perspective are given the lowest salience rating. The very same objects look different when viewed through the other perspective. This is shown in Figure 4.

The salience values derived from perspective can be used for many tasks. Section 11 will show how they can be used in conjunction with Tversky's similarity metric to generate context sensitive responses to misconceptions.

10 CHOOSING THE ACTIVE PERSPECTIVE

In order for the notion of object perspective to be truly beneficial, there must be a mechanism for choosing the active perspective based on previous discourse. While this topic is still very much open to investigation, some preliminary research has revealed several factors that might influence the choice of active perspective.

Perhaps one of the most influential pieces of information useful in choosing a perspective is the user's current goal. In (McKeown, Wish, and Matthews 1985) the user's goal completely determines which perspective is active. In their work each perspective that can be taken on the domain objects is indexed by potential goals. Thus once the system has determined what the user's goal probably is, it has also determined what

Money Market Certificartes
Maturity: 3 months—1
Denominations: $1,000—1
Safety: Medium—0.5

Treasury Bills
Maturity: 3 months—1
Denominations: $1,000—1
Safety: High—0.5

Treasury Bond
Maturity: 7 years—1
Denominations: $500—1
Safety: High—0.5

Figure 3. Objects Through Savings Instruments
Perspective.

perspective the user has probably taken on the domain objects.

While the user's goal is a good source of information to use to determine the probable perspective, other factors may also influence this choice. These include the attributes and objects mentioned so far in the dialog. The mentioned attributes are obviously thought to be important and one would therefore expect them to be given a fairly high salience rating in the active perspective. Thus the choice of active perspective can be narrowed down to those in which the mentioned attributes appear with high salience.

By the same token, the objects mentioned so far in the dialog can also give a clue concerning the active perspective. One would expect that the active perspective would deem these objects important. Therefore the system might look for perspectives that give high salience ratings to many of the attributes associated with objects that have been mentioned in the discourse.

In this section I have identified several factors that influence the choice of active perspective. While the success of other systems (McKeown, Wish, and Matthews 1985) has shown that a reasonable choice of

Monrey Market Certificates
Issuer: Commercial Bank—1
Purchase Place: Commercial Bank—0.5
Safety: Medium—1

Treasury Bills
Issuer: US Government—1
Purchase Place: Federal Reserve—0.5
Safety: High—1

Treasury Bond
Issuer: US Government—1
Purchase Place: Federal Reserve—0.5
Safety: High—1

Figure 4. Objects Through Issuing Company Perspective.

perspective can be made based on discourse goals, the nature of establishing and perhaps "shifting" perspective during a discourse must still be investigated. Still unanswered are questions such as: When does a perspective change? How long is a perspective active? Is there a relationship between a discourse unit (Grosz and Sidner 1985) and perspective? Is there any structure to the space of perspectives that would put constraints on moving from one active perspective to another? These questions must be taken up in future research on perspective.

11 PERSPECTIVE'S INFLUENCE ON RESPONSES

In this section we will show how using the active perspective to highlight the user model, the misconception correcting algorithm from Section 6, and the Tversky similarity metric can account for context sensitive corrections to misconceptions. Recall that in correcting a misattribution one of the correction schemas used by ROMPER called for a similar object to be offered as a possible object of confusion. A study of transcripts reveals, however, that this schema may be instantiated in different ways depending on the context. Consider once again the following dialogs that were first seen in the introduction.

> U: I am interested in investing in some securities to use as savings instruments. I want something short-term and I don't have a lot of money to invest so the instrument must have small denominations. I am a bit concerned about the penalties for early withdrawal. What is the penalty on a T-Bill?

> R: T-Bills don't have a penalty. Were you thinking of Money Market Certificates?

In this case money market certificates were seen as being similar to T-bills and therefore included in the response. A different object might be used in a different context. Consider:

> U: I am interested in investing in some securities. Safety is very important to me, so I would probably like to get something from the government. I am a bit concerned about the penalties for early withdrawal. What is the penalty on a T-Bill?

> R: T-Bills don't have a penalty. Were you thinking of T-Bonds?

The difference in these two responses can be explained by different perspectives being taken on the objects. Suppose we are given the objects, attributes, and perspectives from Section 9. The dialog preceding the first example could lead to the establishment of savings instruments as the active perspective[4] since it mentions the perspective by name and explicitly mentions several of the attributes made important by the perspective. If ROMPER were given this information it would

proceed by attempting to instantiate the wrong object schema described in section 5. Recall that this schema is applicable when there is a similar object that has the property involved in the misconception. The system would collect all objects having the attribute in question and then test their similarity with the object involved in the misconception. In our knowledge base there are two objects that have the attribute involved in the misconception: Money Market Certificates and T-Bonds.

Assume that the f function in the metric is set solely on the basis of the salience values given by the perspective. Also assume that we have decided that two objects are highly similar if Tversky's metric returns a number greater than 0, and not similar otherwise. Applying the Tversky metric using the salience values attached by the savings instrument perspective (and assuming a value of 1 for θ, α, and β) we get:

$$s(\text{T-Bill, MM-Cert}) = f(\text{maturity, denom}) - f(\text{safety})$$
$$= 2 - .5 = 1.5 ===> \text{high similarity}$$

$$s(\text{T-Bill, T-Bond}) = f(\text{safety}) - f(\text{maturity, denom})$$
$$= .5 - 2 = -1.5 ===> \text{low similarity}$$

With these calculations the system would choose the Money Market Certificate as the possible object of confusion and respond:

R: Treasury Bills don't have a penalty. Were you thinking of Money Market Certificates?

Contrast the above calculations with calculations that might occur given a different active perspective. The discourse preceding the misconception utterance in the second example suggests the active perspective of "Issuing Company". Using the salience values attached by this perspective the similarity metric would produce the following calculations:

$$s(\text{T-Bill, MM-Cert})$$
$$= f() - f(\text{issuer, safety, purchase})$$
$$= 0 - 2.5 = -2.5 ===> \text{low similarity}$$

$$s(\text{T-Bill, T-Bond})$$
$$= f(\text{issuer, safety, purchase}) - f()$$
$$= 2.5 - 0 = 2.5 ===> \text{high similarity}$$

In this case a reasonable response by the system would be:

R: Treasury Bills don't have a penalty. Were you thinking of a Treasury Bond?

As the examples show, changes in the active perspective can account for the same misconception being responded to in two different ways.

12 Conclusions

This paper has described a new method for responding to misconceptions that relies on an analysis of a highlighted user model to generate a response that is most likely to benefit the user. A number of strategies were abstracted from a study of transcripts. Each strategy was associated with a distinguished structure in the user model which could explain its use in a given situation. A system can use this pairing to decide how to respond by looking in the highlighted user model for evidence of one of the distinguished structures. The corresponding strategy can be used to respond.

In addition, the paper has described a new notion of object perspective that is able to model one aspect of the dynamic highlighting of the user model due to previous discourse. It was shown how perspective could account for certain contextual affects on responses to misconceptions.

References

Bobrow, D.G. and Winograd, T. 1977 An overview of krl, a knowledge representation language. *Cognitive Science* 1(1): 3–46.

Brown, J.S. and Burton, R.R. 1978 Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science* 2(2): 155–192, 1978.

Carberry, Sandra M. 1984 Understanding pragmatically ill-formed input. In *10th International Conference on Computational Linguistics and 22nd Annual Meeting of the Association of Computational Linguistics*, Stanford University, CA: 200–206.

Carbonnell, Jaime R. and Collins, Allan M. 1970 *Mixed-Initiative Systems For Training and Decision-Aid Applications*. Technical Report ESD-TR-70-373, Electronics Systems Division, Laurence G. Hanscom Field, U.S. Air Force, Bedford, MA.

Grosz, B. 1977 *The Representation and Use of Focus in Dialog Understanding*. Technical Report 151, SRI International, Menlo Park, CA.

Grosz, B.; Joshi, A.K.; and Weinstein S. 1983 Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, Cambridge, MA: 44–50.

Grosz, B. 1981 Focusing and description in natural language dialog. In Webber B, Joshi, A, and Sag, I. (ed.), *Elements of Discourse Understanding*, Cambridge University Press, Cambridge, England: 85–105.

Grosz, B. and Sidner, C. 1985 Discourse structure and the proper treatment of interruptions. In *Proceedings of the 1985 Joint Conference on Artificial Intelligence*, IJCAI85, Los Angeles, CA.

Kaplan, S.J. 1979 *Cooperative Responses From a Portable Natural Language Database Query System*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.

Karlin, Robin 1985 *Romper Mumbles*. Technical Report, University of Pennsylvania, Philadelphia, PA.

Malhotra, A. 1975 *Design Criteria for a Knowledge-Based English Language System for Management*. Technical Report TR-146, Project MAC, MIT, Cambridge, MA.

Malhotra, A. and Sheridan, P. 1976 *Experimental Determination of Design Requirements for a Program Explanation System*. Tech-

nical Report RC 5831, IBM Research Center, Yorktown Heights, NY.

Mays, E. 1980 Correcting misconceptions about database structure. In *Proceedings of the 3rd CSCSI Biennial Meeting*, Victoria, BC, Canada.

McDonald, D.D. 1980 *Natural Language Production as a Process of Decision Making Under Constraint*. Ph.D. thesis, MIT, Cambridge, MA.

McKeown, K. 1982 *Generating Natural Language Text in Response to Questions About Database Structure*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.

McKeown, K.; Wish, M.; and Matthews, K. 1985 Tailoring explanations for the user. In *Proceedings of the 1985 International Joint Conference on Artificial Intelligence*, Los Angeles, CA.

Pollack, M.; Hirschberg, J.; and Webber, B. 1982 User participation in the reasoning processes of expert systems. In *Proceedings of the 1982 National Conference on Artificial Intelligence*, AAAI, Pittsburgh, PA.

Rumelhart, D.E. and Abrahamson, A.A. 1973 A model for analogical reasoning. *Cognitive Psychology* 5: 1–28.

Schuster, E. 1982 *Explaining and Expounding*. Technical Report MS-CIS-82-49, University of Pennsylvania, Philadelphia, PA.

Sidner, C.L. 1983 Focusing in the comprehension of definite anaphora. In Brady, Michael and Berwick, Robert (eds.), *Computational Models of Discourse*, MIT Press, Cambridge, MA: 267–330.

Sleeman, D. 1982 Inferring (mal) rules from pupil's protocols. In *Proceedings of ECAI-82*, Orsay, France: 160–164.

Stevens, A.L. and Collins, A. 1980 Multiple conceptual models of a complex system. In Federico, Pat-Anthony; Snow, Richard E.; and Montague, William E. (eds.), *Aptitude, Learning, and Instruction*, Erlbaum, Hillsdale, NJ: 177–197.

Stevens, A.; Collins, A.; and Goldin, S.E. 1979 Misconceptions in student's understanding. *International Journal of Man-Machine Studies* 11: 145–156.

Tou, F.; Williams, M.; Fikes, R.; Henderson, A.; and Malone, T. 1982 Rabbit: an intelligent database assistant. In *Proceedings of AAAI-82*, Carnegie-Mellon University, Pittsburgh, PA: 314–317.

Tversky, A. 1977 Features of similarity. *Psychological Review* 84: 327–352.

Weiner, E. Judith 1984 A knowledge representation approach to understanding metaphors. *Computational Linguistics* 19(1):1–14.

Woolf, Beverly P. 1984 *Context Dependent Planning in a Machine Tutor*. Ph.D. thesis, University of Massachusetts, Amherst, MA.

Woolf, B. and McDonald, D. 1983 *Human-computer discourse in the design of a pascal tutor*. In Janda, Ann (ed.), CHI'83 Conference Proceedings—Human Factors in Computing Systems, Boston, MA: 230–234.

NOTES

1. While in general it may be necessary for the user model to contain more information about the user (e.g., goals and plans), these are not required for the restricted task set out in this paper.

2. The dialogs contained in this paper were not taken directly from the transcripts. They are used to illustrate the kinds of responses found.

3. Saying that exactly one perspective is active is actually a simplification. It may be the case that a number of perspectives can be active. In this case the resulting "active perspective" will be some function of the individual active perspectives. The exact nature of this function is an open research question.

4. Note that the ROMPER system does not choose the active perspective—it is given as input to the system. This example is simply used to illustrate perspective's influence on misconception responses.