Combining Simulative and Metaphor-Based Reasoning about Beliefs

John A. Barnden Stephen Helmreich Eric Iverson Gees C. Stein

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

Abstract

An unprecedented combination of *simulative* and *metaphor-based* reasoning about beliefs is achieved in an AI system, ATT-Meta. Much mundane discourse about beliefs uses conceptual metaphors (e.g., MIND AS CONTAINER) productively, and ATT-Meta's metaphor-based reasoning accordingly leads to crucial discourse comprehension decisions. ATT-Meta's non-metaphorical mode of belief reasoning includes simulative reasoning (SR). In ATT-Meta, metaphor-based reasoning can block and otherwise influence the course of SR. Also, ATT-Meta can nest SR and metaphor-based reasoning within themselves and each other. As well as currently allowing ATT-Meta to simulatively reason about beliefs about beliefs ..., the nesting will in the near future allow the system to handle chained metaphors, ascribe its own metaphor-based reasoning to other agents, and apply simulative reasoning to purely metaphorical agents.

Introduction

Metaphors in discourse affect an understander's task of obtaining a coherent understanding. This is clear from, e.g., Hobbs (1990), Martin (1990) and Nayak & Gibbs (1990). Consider the discourse fragment (1), and contrast some possible continuations of it, namely (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."*

Continuation (1a) contains a non-metaphorical mental-state description. Sentence (1b) manifests the MIND AS PHYSICAL SPACE conceptual metaphor.[1] The sentence refers to a specific subregion of the whole "space" of the mind. Ideas or thinking episodes in one subregion can be incompatible with ideas in another. For instance, one subregion can contain the thought that a recipe is wrong, whereas another can contain the thought that the recipe is right. Alternatively, thoughts in one subregion can simply be absent from another. In (1c) we see the conceptual metaphor of IDEAS AS INTERNAL UTTERANCES (following Barnden, 1992). A thinking episode is portrayed as inner speech within the agent. We take IDEAS AS INTERNAL UTTERANCES to be a special case of MIND AS PHYSICAL SPACE, with the internal utterance being an event that takes place within the "space" of the agent's mind. MIND AS PHYSICAL SPACE and IDEAS AS INTERNAL UTTERANCES are two of the metaphors to which we have paid most attention in our work. Another is MIND PARTS AS PERSONS, briefly mentioned at the end of the paper. However, there are many commonly-used metaphors of mind. See, for example: Barnden (1992), Cooke & Bartha (1992), Lakoff, Espenson & Schwartz (1991), and Sweetser (1990).

If one looked only at (1b,c) one might dispute the above claims about metaphor, saying that (1b,c) just involve canned forms of language. However, consider the following productive variants of those sentences:

(1b′) *She did this after forcibly shoving the idea that the recipe was wrong to a murky corner of her mind.*

(1c′) *She did this even while whining to herself, "Oh no, this damned recipe's wrong."*

Consider also the immense potential for further varying these, e.g. using verbs other than "shove" and "whine." The most economical explanation of the sense that (1b′-c′) and their further variants make is that they appeal to the metaphors we mentioned above (cf. the observations in Lakoff 1993). Then, for uniformity and continuity, it is a short step to saying that (1b,c) also manifest those metaphors, though in a more pallid way. (1c) does differ markedly from (1c′) in not using an ordinary verb of speech. However, we make three observations. First, people commonly experience some thoughts as "inner speech," so that it is fair to take (1c) as saying that Veronica was experiencing inner speech. Secondly, the verb "think" is in fact often used to portray real speech in the following way: "Veronica thought aloud that the recipe was wrong." Thirdly, the idea that (1c) really is suggesting speech is reinforced by the effect of introducing the evaluative adjective "damned" into the quotation

[1] See Lakoff (1993) for the notion of conceptual metaphor — something that is important in the way people view things — as opposed to mere linguistic manifestations of metaphor. In this paper a metaphor is a conceptual view of something as something else, not a linguistic form.

in (1c).

In (1a,b,c) there is a disparity between Veronica's obeying the recipe and her belief in its incorrectness. The different ways the belief is described lead to different degrees of plausibility for various possible explanations of the disparity. One reasonable interpretation for (1b) is that Veronica's wrong-recipe belief was only minimally involved, if at all, in her conscious thinking, so that she did not consciously think (to any strong degree) that she was following a recipe that was incorrect. But (1c) places the wrong-recipe belief squarely in her conscious thinking, so it seems much more likely that Veronica deliberately went against her own strong doubts, for some reason. For example, she might have been ordered to obey the recipe. This is not to say that such an explanation for (1b) could not hold for (1c), or that explanations for (1c) could not hold for (1b). Rather, the *balance of reasonableness* of explanations is different between (1b) and (1c). The non-metaphorical (1a) is vaguer in its implications than (1b,c), but (1c)-type explanations seem more likely than (1b)-type ones.

We have developed an AI system, ATT-Meta, that seeks explanations such as the ones above. An advanced prototype of the reasoning module is implemented in Prolog. The reasoning system currently takes as input logical expressions constructed by hand, but we are completing the implementation of a natural language front-end.

The system currently does not deal with mental states other than belief, but we plan to broaden its scope in the future. It does not deal with novel metaphors: it already knows about the metaphors that are of interest in our research. However, it can use its metaphors in open-ended, novel ways. It deals only with metaphors of mind, but there appears to be nothing to prevent the extension to numerous other types of metaphor.

ATT-Meta uses an intimate combination of "simulative reasoning" (SR) and metaphor-based reasoning (M4BR) about beliefs. Simulative reasoning has been proposed within AI by, for instance, Chalupsky (1993), Creary (1979), Haas (1986), Konolige (1986) and Moore (1973); and see Dinsmore (1991) for the linguistic context. It also corresponds to an aspect of the philosophical/psychological Simulation Theory (ST) of how people reason about other people's beliefs and behavior (Davies & Stone, in press). In SR, a system temporarily pretends to adopt some of the beliefs of the agent in question, reasons from those pretend facts, and then (defeasibly) ascribes the results to the agent. SR/ST has some advantages over non-simulative reasoning (see Barnden chapter in Davies & Stone, in press; Haas, 1986).

Sketch of ATT-Meta's Reasoning

Here we outline the main reasoning steps ATT-Meta takes for examples (1–1a), (1–1b) and (1–1c), conveying the rough flavor of its simulative and metaphor-based reasoning and of their intimate interaction.[2] The section also illustrates the uncertainty and defeasibility of ATT-Meta's reasoning, largely apparent below through the use of the qualifier "presumably." A proposition so qualified is amenable to being defeated. In this paper we do not go into the formal details of the system. Many of these are contained in Barnden *et al.* (1994).

Case (1–1a)

ATT-Meta's explanation in this case is that she presumably had some special reason for following the recipe even though she was consciously aware of the disparity between doing so and the recipe's being wrong (according to her). This explanation arises as follows.

From the fact that Veronica followed the recipe, ATT-Meta infers:

(2) *presumably, Veronica consciously believed she was following it.*

This is by a rule that says that when people follow instructions they presumably do it consciously. ATT-Meta also infers:

(3) *presumably, she believed it wasn't good to follow the recipe.*

This happens as follows. ATT-Meta sets up a special reasoning environment for conducting a partial simulation of Veronica's (hypothesized) thought processes. Within the environment, ATT-Meta adopts the premise that *the recipe is wrong.* This is a belief of Veronica's, according to the second clause of (1a). Because ATT-Meta can be viewed as temporarily *pretending to adopt Veronica's beliefs*, we call the environment a "simulation pretence cocoon." Using a rule that says that if a body of instructions is wrong it's not good to follow it, ATT-Meta infers within the cocoon that

(3′) *it's not good to follow the recipe.*

Since this conclusion is within the cocoon, it is a hypothesized belief of Veronica's, and ATT-Meta makes this point explicit by automatically concluding (3).

Further, by virtue of (2), ATT-Meta adopts the following premise within the cocoon:

(2′) *Veronica follows the recipe.*[3]

Then, within the cocoon, ATT-Meta infers the conjunction of (2′) and (3′), namely

(4′) *Veronica follows the recipe AND it's not good to follow the recipe.*

Concomitantly, ATT-Meta infers, outside the cocoon, that Veronica presumably believes this conjunction. That is:

[2] At the time of writing, ATT-Meta has only a simplified implementation of the complex "specificity" principles appealed to in case (1–1b), so that the description here goes slightly beyond the current capabilities of the implementation. Another departure by ATT-Meta from the description in this section is that ATT-Meta proceeds in a backwards-chaining, goal-directed way, whereas the description portrays a more easily presentable forward-chaining process.

[3] Since ATT-Meta is simulating Veronica, it would be better to couch this premise as "I am following the recipe." However, ATT-Meta does not yet have a treatment of indexicals.

(4) *presumably, Veronica believes that: she follows the recipe AND it's not good to follow it.*

Notice here that an SR conclusion such as (3) or (4) is always qualified by "presumably," reflecting the fact that ATT-Meta merely *presumes* that the agent actually does the inferencing corresponding to ATT-Meta's within-cocoon reasoning.

Now, ATT-Meta has the following (admittedly oversimplified) rule:

(R.1) IF *someone is using something (e.g., a set of instructions) and believes something, P, about it,* THEN, presumably, *(s)he consciously believes P.*

Therefore, ATT-Meta infers that Veronica's belief in the wrongness of the recipe was presumably conscious. Thus, both premises used in the simulation cocoon (namely: *the recipe is wrong*; *Veronica follows the recipe*) reflect *conscious* beliefs of Veronica's. As a result, ATT-Meta presumes that any belief resulting from the simulation is also conscious. Therefore, a further result of the simulation is

(4c) *presumably, Veronica* **consciously** *believed that: she follows the recipe AND it's not good to follow the recipe.*

This feeds into a rule that can be paraphrased as follows:

(R.2) IF *agent X does action A and consciously believes that [(s)he does A AND it's not good to do A]* THEN, presumably, *the explanation is that (s)he has a special reason for doing A despite having that conscious belief.*

Thus, ATT-Meta is able to infer the main result of the example, namely:

(5) *presumably, Veronica had a special reason for following the recipe even though consciously believing that [she's following the recipe AND it is not good to follow it].*

Case (1–1c)

We defer (1–1b) as it involves more complex reasoning than (1–1c) does. ATT-Meta's general approach to metaphor is to "pretend" to take a metaphorical utterance at face value (i.e "literally"). That is, in the case of (1c), ATT-Meta pretends that

(6) *there was a* **real** *utterance of "The recipe's wrong" within Veronica's mind.*

Concomitantly, ATT-Meta pretends that

(7) *Veronica's mind was a PHYSICAL SPACE.*

These pretences are embodied as the adoption of premises (6) and (7) within a special environment that we call a *metaphorical pretence cocoon* for Veronica's-IDEAS AS INTERNAL UTTERANCES. Now, the real force of such cocoons is that inference processes can take place within them, much as within simulation cocoons. This will happen in case (1–1b). However, in the present example, the only important action that ATT-Meta bases on the metaphor is to use the following "transfer rule" linking certain metaphorical cocoons to reality:

(TR.1c) IF [within a cocoon for agent X's-IDEAS AS INTERNAL UTTERANCES] *there is an utterance of a declarative sentence S within X's mind* THEN, presumably, *X consciously believes the proposition stated by S.*

Thus, ATT-Meta infers that, presumably, Veronica consciously believed the recipe to be wrong. This is also inferred by means of R.1 just as in case (1–1a). Thus, the remainder of the reasoning is essentially the same as that for (1–1a), and ATT-Meta again constructs the main result (5).

However, ATT-Meta is theoretically in a position to deem (5) to be more strongly supported in the (1–1c) case than in the (1–1a) case. This is for reasons detailed below, but is to do with TR.1c providing a more specific basis than R.1 does for the conclusion that Veronica consciously believes the recipe to be wrong.

Case (1–1b)

In case (1–1c), the M4BR was minimal, and furthermore did not affect the course of simulative reasoning as such. Things are markedly different in the case of (1–1b). ATT-Meta still tries to do the same simulative reasoning about Veronica as above, and comes up with hypothesis (4) again. However, this hypothesis is defeated. This in turn means that neither (4c) nor (5) is inferred. The reason for the defeat of (4) is as follows.

Suppose ATT-Meta comes to a within-cocoon conclusion Q, and that this was directly based on within-cocoon propositions P1, ..., Pn. ATT-Meta concomitantly sets up the external conclusion that the agent (X) believes Q, as was implied above. However, another action is to record that this conclusion is dependent upon the hypothesis that

(I) *X performs some inference process yielding Q from P1, ..., Pn.*

This hypothesis is, normally, deemed by ATT-Meta to be *presumably* true. It turns out that for examples (1–1a) and (1–c) there is nothing that defeats this presumption. However, one use of M4BR in ATT-Meta is precisely to defeat presumptions of form (I). But if all instances of I for a given Q are defeated then (in the absence of other support) ATT-Meta abandons the conclusion that X believes Q, and concomitantly abolishes Q within the cocoon. Two instances of I are set up in case (1–1b), for different propositions Q:

(I.1) *Veronica performed some inference process yielding [it is not good to follow the recipe] from [the recipe is wrong];*

(I.2) *Veronica performed some inference process yielding [Veronica follows the recipe AND it is not good to follow the recipe] from [Veronica follows the recipe] and [it is not good to follow the recipe].*

These arise because of within-cocoon reasoning steps just like those in cases (1–1a) and (1–1c). (I.1) arises from the reasoning to (3′) and (I.2) from the reasoning to (4′).

Now, part of ATT-Meta's understanding of the MIND AS PHYSICAL SPACE metaphor is:

(TR.1b) *X's performing an inference process yielding Q from P1, ..., Pn corresponds metaphorically to P1, ...,* ***Pn physically interacting*** *within X's mind space to produce Q.*

(If n is 1 then Q arises just out of P1, without an interaction with something else.) This principle is couched in a set of transfer rules analogous in form to TR.1c. In addition, ATT-Meta has a rule *purely about physical interactions* that says

(R.2) IF *some things are spatially separated from each other [rather than being close together]* THEN, presumably, *they do not interact.*

Another purely physical rule is

(R.3) IF *P1 ,..., Pn physically interact to produce Q and the Pi are all within a particular region R,* THEN, presumably, *Q is in R.*

Other parts of ATT-Meta's understanding of the metaphor are the following transfer principles:

(R.4) *X believing P corresponds to the thought-that-P being at some position in X's mind-space;*

(R.5) *X consciously believing P corresponds to X's mind having a front region and the thought-that-P being in that region;*

(R.6) IF *a thought is in the recesses of X's mind THEN, presumably, it is not conscious.*

ATT-Meta sets up a metaphorical pretence cocoon for Veronica's-MIND AS PHYSICAL SPACE. ATT-Meta takes (1b) at face value and adopts the within-cocoon premise that in the recesses of this space there was the thought that the recipe was wrong. Because of hypothesis (I.1) and transfer principle (TR.1b), ATT-Meta establishes the following hypothesis inside the metaphorical pretence cocoon:

(8) *the thought that the recipe was wrong* **physically produced** *the thought that it was not good to follow it.*

By R.3, it follows that the latter thought was also in the *recesses* of Veronica's mind.

However, ATT-Meta infers as in (1–1a) that presumably Veronica consciously believed that she was following the recipe. Hence, by R.5, the thought that Veronica follows the recipe was in the *front* of her mind. ATT-Meta takes the front and the recesses to be separated from each other. Therefore, ATT-Meta uses R.2 to infer within the metaphorical pretence cocoon that the thought that Veronica follows the recipe did *not* physically interact with the thought that it is not good to follow the recipe. Via TR.1b, this undermines I.2. But I.2 was the only support for hypothesis (4). Hence, (4) is defeated. As a result, (4c) and (5) are not inferred.

On the other hand, (I.1) is not undermined because it reflects a within-simulation-cocoon inference from a *single* premise (that the recipe is wrong); there is no issue arising here of two premises being being physically separated within Veronica's mind-SPACE.

We have seen that ATT-Meta does not conclude (4c). In fact, for reasons explained at the end of the next section, ATT-Meta is able to come to the stronger conclusion that Veronica actually failed to believe the conjunction (4′) (even though she does believe each conjunct separately). This then allows the following rule to proceed:

(R.7) IF *agent X does action A, believes that it's not good to do A, but fails to believe that [X does A AND it's not good to do A]* THEN, presumably, *this failure explains the apparent disparity between X's action and belief.*

Thus, ATT-Meta is able to infer the main result of the example, namely:

(9) *presumably, the explanation for the apparent disparity concerning Veronica is that she failed to believe that [Veronica follows the recipe AND the recipe is wrong].*

It turns out that ATT-Meta does arrive at some *weak* support for the explanation given by (9) in cases (1–1a) and (1–1c), and conversely comes up with some weak support for the explanation given by (5) in case (1–1b). (What weak support amounts to is clarified below.) This reflects our point in section 1 that the metaphors affect the balance of reasonableness of explanations, and do not totally discount particular explanations.

Finally, recall that in (1–1a) ATT-Meta inferred that, presumably, Veronica *consciously* believed the recipe to be wrong. This inference is attempted also in case (1–1b). However, it is defeated indirectly by the given information that the thought that the recipe was wrong was in the *recesses* of her mind, which supports via R.6 the hypothesis that the belief was *not* conscious. The support for this hypothesis is judged to be more specific, and therefore stronger, than the support for the hypothesis that the belief was conscious. This is because of general specificity-comparison principles in ATT-Meta, discussed below.

The Two Types of Pretence Cocoon

ATT-Meta uses simulation pretence cocoons to "get inside people's minds," and metaphorical pretence cocoons to "get inside metaphorical views." Notice carefully, however, that the reasoning within a metaphorical pretence cocoon concerned with someone's mind is nevertheless from a standpoint *outside* the person's mind. We could unify the two types of cocoon by regarding a metaphorical pretence cocoon as a simulation cocoon for simulating a hypothetical observer who takes the metaphor at face value — e.g., actually believes Veronica's mind literally to be a physical space.

Uncertainty and Defeasibility

Propositions (facts in the database, or reasoning goals) are tagged with a confidence rating (CR). The most important ratings are *Suggested*, *Presumed* and *Certain*. *Presumed* means that ATT-Meta presumes that

the proposition holds, but allows that it might not do so. The rating can also be glossed as "by default." *Presumed* corresponds to our use of "presumably" in the previous section. *Suggested* means intuitively that there is reason to suggest the proposition might hold, but the available evidence is not enough to make ATT-Meta presume it. A proposition and its complement (where P and NOT-P are complements) may both be rated as *Suggested*, or one may be *Suggested* and the other *Presumed*. However, other combinations are not allowed.

Suggested is the rating that (9) ends up with in case (1–1a,c) [contrasting with its *Presumed* rating in case (1–1b)] and that (5) ends up with in (1–1b) [contrasting with its *Presumed* rating in case (1–a,c)].

The notion of belief is divided up into qualitative degrees. We can have that "X believes-C P", where C is one of *Suggested*, *Presumed* or *Certain*. The meaning is respectively that: X believes there is some reason to suggest P; X presumes P; and X is certain that P. Notice carefully that ATT-Meta's CR for "X believes-ρ P" is independent of degree ρ. In the previous section, we oversimplified by not mentioning degrees of belief.

In a simulative reasoning cocoon for agent X, any of ATT-Meta's own rules can be used, although in the future we plan to add the capability of of restricting the ascription of rules to agents. Within-cocoon premises arise only from *Presumed* or *Certain* propositions about X's beliefs. However, such propositions can be conclusions from general rules like "if X is an English telephone operator then X believes that New Mexico is not part of the U.S."

ATT-Meta has a rule such that, for each agent X and each proposition P, the system presumes that X *lacks* a belief (of any degree) in P. (Note that this does not mean that X believes not-P.) However, the presumption is overridden by any *Presumed* or *Certain* database fact to the contrary, or by rule that tries to contribute a *Presumed* or *Certain* rating to X believing P. The default that X lacks a belief in P is the reason that in case (1–1b) ATT-Meta can infer that (presumably) Veronica failed to believe the conjunction (4′) mentioned there, because (as it happens) the evidence from the discourse fragment leads merely to a *Suggested* rating for her believing the conjunction.

The CR values are also used as strengths of rules. The proposition concluded by the rule is given a strength equal to the minimum of the rule's own CR together with the CRs of the propositions picked up by the conditions of the rule. When several rules support a conclusion, the maximum of the rules' individual strength-contributions is taken. Special actions are taken when both a proposition and its complement are given a strength of *Presumed* or *Certain*. Most interestingly, in the *Presumed*/*Presumed* case both ratings are reduced to *Suggested*, unless the system's specificity principles (see below) are able to establish that the support for one is more specific than the support for the other. In that case, only the less specific *Presumed* is downgraded to *Suggested*.

Specificity Comparison Heuristic

The current specificity comparison heuristic with which we are currently experimenting is complex and is set out in detail in Barnden *et al.* (1994). Here will just give the flavor of one aspect of it. Recall that at the end of the subsection on the reasoning for Case (1–1b), we stated that when ATT-Meta considers whether or not (C) Veronica *consciously* believes the recipe to be wrong, it finds the support for NOT-C stronger (i.e.. more specific) than the support for C. Initially, C gets a *Presumed* rating by R.1 being applied (in part) to the *Certain* proposition (F1) that Veronica believes the recipe to be wrong. NOT-C gets an initial *Presumed* rating by R.6 being applied to the *Certain* proposition (F2) that the thought that the recipe was wrong was in the recesses of Veronica's mind. Now, F1 can be inferred from F2 via R.4, but F2 cannot be inferred from F1. As a result, F1 is regarded as being less specific than F2. (F2 describes a more special situation than F1 does.) Because of this, it turns out that the overall support for C is less specific than the overall support for NOT-C.

We also stated that the specificity comparison principles could also be used to judge (5) to be more strongly supported in the (1–1c) case than in the (1–1a) case. (However, the ATT-Meta implementation currently does not try to compare conclusions from different examples.) (5) arises from R.2, whose condition part requires a *conscious* belief. The belief in question is that Veronica follows the recipe and it's not good to follow it. The inference process in case (1–1a) that supplies the conscious quality of this belief is similar to the process that occurs in case (1–1c), and uses the *Presumed* proposition C (that Veronica consciously believes the recipe to be wrong). However, a crucial difference is as follows. For (1–1a), it is only R.1 acting (partly) on proposition F1 that supports the consciousness in C. By contrast, for (1–1c) the consciousness in C is also supported by TR.1c acting on the proposition (F3) that there was an utterance of "The recipe is wrong" in Veronica's mind. F3 is more specific than F1, much as F2 is.

Nesting of SR and M4BR

An important though relatively neglected advantage of SR is that it allows *any* type of reasoning, not just deductive reasoning, to be ascribed to agents. Without SR, the system would need an elaborate, explicit *theory of* each type of reasoning — e.g., abduction or analogy-based reasoning — to enable it to reason about other agents' hypothesized abduction, analogy-based reasoning, etc. In the ATT-Meta case, the system's own defeasible reasoning is routinely used within simulation cocoons. As a special case, the system's M4BR capability could be used within simulation cocoons. This would allow the simulation of agents' hypothesized M4BR. Equally, simulative reasoning could be used within metaphorical pretence cocoons, to simulate people that appear in the source domain of a metaphor. For example, if an inanimate entity were metaphorically viewed as a person, that metaphorical person could be subjected to simulative reasoning.

The nesting of SR within itself is already used in the

ATT-Meta implementation to reason about agents' reasoning about other agents' beliefs. The final case is nesting of M4BR within itself. This is useful for dealing with chained metaphor. For instance, a personification of an inanimate object could be combined with a MIND AS PHYSICAL SPACE view of the metaphorical person.

Novel Uses of Familiar Metaphor

Although ATT-Meta currently only deals with metaphors familiar to it, it can respond to novel uses of them. For instance, suppose an input sentence is "One part of Veronica was vociferously insisting that the recipe was wrong." This manifests a familiar metaphor called MIND PARTS AS PERSONS. Assume nevertheless that ATT-Meta has no transfer rules (cf. TR.1b,c) for mapping the "insisting" to some mental quality of Veronica's. This does not prevent ATT-Meta from inferring at least as much as it could from the sentence "One part of Veronica believed that the sentence was wrong." From its general knowledge about real natural language communication, it knows that if someone insists something they normally believe it. This rule can be used within the metaphorical pretence cocoon to infer that the insisting part of Veronica believed that the recipe was wrong. This then allows the inference that Veronica to some extent believed that the recipe was wrong, by a transfer rule. Similarly, since real people usually only *insist* things when faced by objections, ATT-Meta can presume that there's another metaphorical person in Veronica's mind who believed that the recipe was *correct*. Therefore, ATT-Meta can infer that Veronica also to some extent believes that the recipe was correct.

Here we have a simple illustration of the point that, as long as metaphorical elements in the input sentence lead to within-metaphorical-cocoon conclusions that do link up with transfer rules, much can be done.

Concluding Remarks

To our knowledge the use of metaphor-based reasoning to aid belief reasoning is unique, whether or not belief reasoning can be done simulatively as in ATT-Meta or otherwise. (Not all of ATT-Meta's belief reason is simulative: for instance, conclusions that an agent does not believe something are arrived at non-simulatively.) ATT-Meta is also one of the few detailed computational schemes for performing significant metaphor-based *reasoning* in discourse understanding (others are the schemes of Hobbs, 1990, and Martin, 1990), as opposed to doing semantic analysis of metaphorical sentences.

Finally, we are starting on a psychological experimentation program to assess the extent to which people come up with explanations concerning mind-metaphorical discourse fragments in broadly the way we have designed ATT-Meta to. At the time of writing we are conducting the first set of experiments.

Acknowledgements

This work was supported in part by grants IRI-9101354 and CDA-8914670 from the National Science Foundation. We thank Kim Gor, Kanghong Li and Jan Wiebe for suggestions and help with system development, and Lori Markson for conducting experiments.

References

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

Barnden, J.A., Helmreich, S., Iverson, E. & Stein, G.C. (1994). An integrated implementation of simulative, uncertain and metaphorical reasoning about mental states. In J. Doyle, E. Sandewall & P. Torasso (Eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the Fourth International Conference*. San Mateo, CA: Morgan Kaufmann.

Chalupsky, H. (1993). Using hypothetical reasoning as a method for belief ascription. *J. Experimental and Theoretical Artificial Intelligence*, 5(2&3), 119–133.

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

Creary, L. G. (1979). Propositional attitudes: Fregean representation and simulative reasoning. *Proceedings of the 6th. Int. Joint Conf. on Artificial Intelligence* (pp. 176–181). Los Altos, CA: Morgan Kaufmann.

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

Dinsmore, J. (1991). *Partitioned Representations: A Study in Mental Representation, Language Processing and Linguistic Structure*. Dordrecht: Kluwer Academic Publishers.

Haas, A.R. (1986). A syntactic theory of belief and action. *Artificial Intelligence*, 28, 245–292.

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

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

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and Thought*, 2nd edition (pp.202–251). New York: Cambridge University Press.

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

Martin, J.H. (1990). *A Computational Model of Metaphor Interpretation*. Academic Press.

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

Nayak, N.P & Gibbs, R.W., Jr. (1990). Conceptual knowledge in the interpretation of idioms. *J. Experimental Psychology: General*, 119(3), 315–330.

Sweetser, E.E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge, U.K.: Cambridge University Press.