Epistemic Logics for Multiple Agent Nonmonotonic Reasoning I

Leora Morgenstern
IBM T.J. Watson Research
P.O. Box 704, Mail Stop H1L05
Yorktown Heights, N.Y. 10598
(914)784-7151
leora@watson.ibm.com

Ramiro Guerreiro
IBM Rio Scientific Center
Rio de Janeiro, 20071 Brazil
Ramiro@riosc.bitnet

Abstract

Although most research in plausible reasoning has focused on the single agent case, multiple agent nonmonotonic reasoning is a prerequisite for many commonsense reasoning problems. This paper explores the features that are essential to any multiple agent nonmonotonic logic, such as explicit belief, inter-agent reasoning rules, and arrogance. We present MANML, an extension of Autoepistemic Logic to multiple agents and C_EMAT, an extension of a circumscriptive theory of temporal reasoning to multiple agents. Next, we outline a method of extending C_EMAT to a general multiple agent theory of circumscription. We discuss the connections between MANML and the extended version of C_EMAT, and demonstrate that both theories can handle a wide variety of commonsense reasoning problems.

1 Introduction

Much past research has focused upon the construction and development of nonmonotonic logics since it has long been recognized that standard monotonic logics are inadequate for much of commonsense reasoning. Virtually all of this research has focused upon the single agent case. [1]

In fact, however, multiple-agent nonmonotonic reasoning is prevalent in commonsense domains. We use it, for example, when we evaluate expert advice. Consider, for example, my thought processes when I listen to the meteorologist on my local radio station in the morning. I reason that he makes his predictions based on partial knowledge about the current weather situation and default rules of meteorology. Thus, while on the whole I will accept what the weatherman says, I view his predictions with a dose of healthy skepticism. So, if the weatherman predicts sunshine all day, but I see ominous storm clouds gathering, I will probably disagree with his conclusions.

Similarly, the sophisticated patient knows that his doctor is fallible and uses some kind of default reasoning (often, a type of abduction). He may thus question his doctor's orders, ask for a second opinion, etc. In general, the more closely he can model the doctor's line of reasoning, the more directly he will be able to question his doctor, the more likely he will be to choose an accurate specialty for a second opinion, etc.

Multiple agent nonmonotonic reasoning is thus essential for planning. My plan to get dressed in the morning – in particular, my decision to carry or not carry an umbrella – rests upon my ability to evaluate the meteorologist's forecast. Jim's plan to get healthy rests

[1] But see the Previous Work Section (section 4) for a discussion of steps towards a theory of plausible reasoning among multiple agents.

upon his ability to reason about his physician's diagnosis and suggested course of treatment and the reasoning process that the doctor uses to reach these conclusions.

It is interesting to note that certain classical reasoning problems, such as the Prisoner's Dilemma ([Luce and Raiffa, 1957], [Rapoport, 1974]), are much more naturally modeled in a multiple agent nonmonotonic logic. Each prisoner must reason about what he believes the other prisoner is reasoning about what he is reasoning ... etc. Standard approaches which model the problem in a monotonic logic of belief are overly simplistic, since it is almost never the case that each prisoner knows exactly what the payoff for each action is, and/or that this information is common knowledge.

Finally, we note that virtually every problem that calls for standard nonmonotonic reasoning has associated with it a problem that calls for multiple agent nonmonotonic reasoning - specifically, what we will call the *nested* version of the single agent problem.

For example, associated with the well-known Tweety problem is the *nested* Tweety problem. Suppose that Bill believes that Carol believes that birds typically fly and that Bill believes that Carol believes that Tweety is a bird. What can we say about Bill's beliefs about Carol's beliefs vis a vis Tweety's flying ability? It takes a multiple agent nonmonotonic logic to handle the problem, and surprisingly, the seemingly obvious conclusion that Bill believes that Carol believes that Tweety can fly is true only if we add some fairly strong assumptions. Similar remarks can be made about the nested Yale Shooting Problem, the Nixon diamond, and other well-known classical problems of non-monotonic reasoning.

A theory of multiple agent nonmonotonic reasoning is thus crucial for general commonsense reasoning. Below, we will develop such a theory, aiming towards the solution of a set of benchmark problems that spans a wide variety of commonsense reasoning issues: the Tweety, Nixon Diamond [Reiter and Criscuolo] (this is one case where we *want* multiple extensions), and the Yale shooting [Hanks and McDermott, 1986] problems, and their nested versions. The YSP has associated with it a number of important variations such as the Bloodless YSP (a.k.a the Stolen Car problem [Kautz, 1986]), the Stanford Murder Mystery [Baker, 1989], and the Message Passing Problem [Morgenstern and Stein, 1988], and the theory should solve these as well. In addition, the theory should handle the Third Agent Frame Problem [Morgenstern, 1991], a temporal reasoning problem involving nested planning and chains of requests (see section 3. for a description of this problem).

The move from a single agent non-monotonic logic to a multiple agent non-monotonic logic is not trivial. In particular, the following features, not present in a single agent non-monotonic logic, are a necessary part of a multiple agent logic:

[1] *explicit mention of agents:* In standard non-monotonic logics, it is assumed that there is only one reasoning agent. It is always the case, therefore, that this agent is left implicit. Once multiple agents are introduced, however, these agents must be explicitly mentioned in the theory.

[2] *explicit mention of knowledge and/or belief:* In single agent non-monotonic logics it is implicitly assumed that the sentences of the theory and the conclusions which this theory nonmonotonically entails are believed by some agent. Many nonmonotonic logics ([Reiter, 1980], [McDermott and Doyle, 1980], [McCarthy, 1980]) have no mention of knowledge or belief at all. Moore's Autoepistemic Logic [1985] is notable for its explicit inclusion of a belief operator L; nonetheless, even in AEL, a so-called "objective formula" (one with no occurrences of the L operator) is still considered to be believed by the implicit, shadowy single agent hovering in the background of any traditional non-monotonic logic.

[3] *inclusion of inter-agent reasoning rules:* In standard nonmonotonic logics, an agent reasons only from and about his beliefs. In a multiple agent non-monotonic logic, agents will have to reason about other agents' reasoning abilities. For example, an agent should be

able to reason that another agent can perform modus ponens and positive introspection.

[4] *Inclusion of arrogance:* Traditional non-monotonic logics are typically characterized by either optimism (e.g, Reiter's Default Logic [1980]) or introspection (e.g., Moore's Autoepistemic Logic [1985]). In contrast, multiple agent non-monotonic logics are characterized by their *arrogance*. This arrogance is needed in in all but the most trivial cases. Suppose, e.g., that I believe that Susan believes that birds typically fly and that I believe that Susan believes that Tweety is a bird. Can I conclude that Susan believes that Tweety flies? At first glance, the answer is yes. But upon further reflection, it becomes clear that the situation is not that simple. Susan may (and very likely does) believe more than I know. It might be that Susan has an additional belief; that Tweety is an ostrich. In that case (assuming that Susan believes that ostriches do not fly), Susan would certainly *not* conclude that Tweety can fly.

In order to reasonably conclude that Susan believes that Tweety can fly, I must (somehow) assume that Susan *doesn't* have any reason to believe that Tweety cannot fly. That is I must assume that Susan does not have certain sorts of beliefs. Now, it is all very well for me to contemplate the sum total of my beliefs and conclude that I do not have a certain belief – this is called introspection, and it is generally considered to be a good thing. My assumption, however, tht someone *else* does not have a particular believe – based on my partial knowledge of her beliefs – is not introspection; it is arrogance, which is not generally considered to be an attractive quality at all.

Indeed, in too large quantities, arrogance often leads to incorrect and absurd results. For example, suppose I know that nearly everyone believes that if he has an older brother, he believes he has an older brother, or equivalently, that if he doesn't believe he has an older brother, he doesn't have one. Now, presumably, when I see people on line in front of me at the supermarket, I don't know anything about what they believe about their brothers - except for the fact that they'd know they had one if they had one. If I were sufficiently arrogant, I would thus assume that each stranger does *not* have the belief that he has an older brother, and I could therefore conclude that all of these people do not have older brothers! Obviously, this is absurd. This is a case where partial knowledge of someone else's beliefs will not sanction any conclusions.

Developing a multiple agent non-monotonic logic will thus require capturing this quality of arrogance in just the right amount. For example, restricting arrogance to sentences believed by agents one knows well, as suggested in Section 2, would avoid the ludicrous conclusions of unbridled arrogance.

Below, we will explore several extensions to single agent non-monotonic logics. We begin by extending Moore's Autoepistemic Logic to multiple agents. We next discuss an extension of a model-based theory of temporal reasoning to multiple agents, and discuss the ways in which it could be extended to a general theory of multiple agent circumscription.

2 Extending Autoepistemic Logic to a Multiple Agent NML

In this section, we present MANML (Multiple Agent NonMonotonic Logic), an extension of Moore's Autoepistemic Logic (AEL)[1985] to multiple agents. [2] We give several versions in order of increasing power and expressiveness.

We choose to extend Autoepistemic Logic since AEL already has an explicit belief operator L. [We briefly review AEL: Sentences of AEL are defined as in any modal logic, with L

[2] Although the term MANML was used to describe the account of the account of a multiple agent non-monotonic logic in [Morgenstern, 1990], the non-monotonic logics discussed in this paper are substantially different from the MANML of [Morgenstern, 1990]. MANML1, the first, and least powerful, pass of the current version of MANML, discussed below, is essentially equivalent in power to the old version of MANML. However, MANML1 repairs the omissions (in particular, the lack of the concept of a stable expansion) of the old version of MANML.

the standard belief operator. A theory T is a stable set if it obeys the following rules: [1] T is closed under logical consequence ; [2] if $P \in T$, then $LP \in T$; [3] if $P \notin T$, then $\neg LP \in T$. T is grounded in a set of premises A if every formula of T is a tautological consequence of $A \cup \{LP \mid P \in T\} \cup \{\neg LP \in P \notin T\}$. Stable expansions of A are those extensions that are stable and grounded in A.]

To extend AEL to MANML, we choose as our underlying logic the modal logic $\mathcal{L}$. Since we wish to model multiple agents, we index the belief operator by agents. L_aP is to be read as: "a believes P." The formation rules of $\mathcal{L}$ are:
a) if ϕ is a sentence of FOPC, ϕ is a sentence of $\mathcal{L}$
b) if ϕ is a sentence of $\mathcal{L}$ and a is a constant of $\mathcal{L}$ denoting an agent, $L_a\phi$ is a sentence of $\mathcal{L}$
c) if ϕ and ψ are sentences of $\mathcal{L}$, so are $\phi \vee \psi$ and $\neg\phi$.
Quantifiers do not include modal operators in their scope.

We present the stable-set formation rules for a theory T. As a first step, we would like agents in MANML to be able to perform autoepistemic reasoning. Thus, we suggest a set of inference rules that mimic the stable-set formation rules of AEL, modulo an explicit level of indexing for agents. Specifically, we have:

Definition 1 *A set T is a MANML1 stable set if it obey conditions (1) -(4), below:*
(1) T is closed under first order consequence
(2) If $L_aP_1 \cdots L_aP_n \in T, and P_1 \cdots P_n \vdash Q$, then $L_aQ \in T$ (where $\vdash$ is the logical derivation operator of FOPC)
(3) if $L_aP \in T$, then $L_aL_aP \in T$
(4) if $L_aP \notin T$, then $L_a\neg L_aP \in T$

Conditions (1) through (4) may be thought of, loosely, as inference rules: we call $\mathcal{L}$ with rules (1) - (4) MANML1.

The definition of a MANML1 stable set motivates the definition of a MANML1 stable expansion. We begin with some preliminaries. Let:
$CONS$ (mnemonic for consequences) = $\bigcup_a \{L_aQ \mid L_aP_1 \cdots \mathrm{L}_aP_n \in T, P_1 \cdots P_n \vdash Q\}$
PI (mnemonic for pos. introspection) = $\bigcup_a \{L_aL_aP \mid L_aP \in T\}$
NI (mnemonic for neg. introspection) = $\bigcup_a \{L_a\neg L_aP \mid L_aP \notin T\}$
Then we have

Definition 2 *T is a MANML1 stable set expansion of a set of premises A iff T is the set of first order consequences of $A \cup CONS \cup PI \cup NI$.*

We introduce a notion of groundedness analogous to Moore's [1985] definition:

Definition 3
T is MANML1 (weakly) grounded in A iff $T = \{\phi \mid A \cup CONS \cup PI \cup NI \vdash \phi\}$

The following theorem follows immediately:

Theorem 1 *A set T is a MANML1 stable set expansion of A iff T is MANML1 grounded in A.*

Agents in MANML1 can perform autoepistemic reasoning. For example, if $A = \{L_a(B \Rightarrow L_aB)\}$, then every MANML1 expansion of A contains the sentence $L_a\neg B$. Likewise agents can perform simple default reasoning. Defaults are represented as $L_a\alpha \wedge \neg L_a\neg\beta \Rightarrow \gamma$. MANML1 can handle both Tweety and the Nixon Diamond.

Thus far, it seems that MANML1 is identical to AEL, except that MANML1 wraps an extra level of belief operator about sentences. In fact, it can be shown that MANML1 subsumes AEL [Morgenstern, 1990]. The real question, of course, is: does MANML1 have any advantages over AEL? We can point to two advantages: 1. It allows the representation of genuine *objective* formulas, formulas that are not necessarily believed by any agent, but are true of the world. 2. It allows the representation of many agents' beliefs within one theory.

Nevertheless, MANML1 doesn't allow any sort of multiple agent reasoning at all; agents reason only about their own beliefs. To allow multiple agent beliefs, we will have to add *inter-agent reasoning rules.* We thus present our second pass at a multi-agent NML, MANML2:

Definition 4 *A set T is a MANML2 stable set if it obeys the following conditions:*
(1) through (4), above
(5) if $L_{a1}\cdots L_{am}P_1, L_{a1}\cdots L_{am}P_n \in T, P_1\cdots P_n \vdash Q$, then $L_{a1}\cdots L_{am}Q \in T$. (multiple agent consequential closure)
(6) If $L_{a1}\cdots L_{am}P \in T$, then $L_{a1}\cdots L_{am}L_{am}P \in T$ (multiple agent positive introspection).

Thus, agents in MANML2 know that (a) other agents reason using first order logic and (b) other agents use positive introspection. We discuss the motivation for these rules below.

We extend the definitions of stable expansion and groundedness from MANML1. Let:
$MULTI\_CONS = \bigcup_{ai}\{L_{a1}\cdots L_{am}Q \mid L_{a1}\cdots L_{am}P_1 \in T, \cdots, L_{a1}\cdots L_{am}P_n \in T, P_1\cdots P_n \vdash Q\}$
$MULTI\_PI = \bigcup_{ai}\{L_{a1}\cdots L_{am}L_{am}P \mid L_{a1}\cdots L_{am}P \in T\}$

Then we have the following definitions:

Definition 5 *T is a MANML2 stable set expansion of A if T is the set of first order consequences of $A \cup CONS \cup PI \cup NI \cup MULTI\_CONS \cup MULTI\_PI$*

Definition 6 *T is MANML2 grounded in a set of premises A iff $T = \{\phi \mid A \cup CONS \cup PI \cup NI \cup MULTI\_CONS \cup MULTI\_PI \vdash \phi\}$*

Again, we immediately have:

Theorem 2 *T is a MANML2 stable set expansion of A iff T is MANML2 grounded in A.*

MANML2 allows general *monotonic* multiple agent reasoning. However its nonmonotonic power is limited to its autoepistemic reasoning power. This is chiefly due to the fact that in MANML, agents are not arrogant. In this vein, it is interesting to note how MANML2 extends MANML1. We note that rule (5) is the multiple agent analogue of rule (2) (consequential closure) and rule (6) is the multiple agent analogue of rule (3) (positive introspection). Interestingly, there is no analogue of rule (4) (negative introspection) ! Such an analogue would look something like this:

[**Principle of Supreme Arrogance**] if $L_{a1}\cdots L_{am}P \notin T$, then $L_{a1}\cdots L_{am}\neg L_{am}P \in T$
(i.e., if A1 doesn't know that .... that AM knows P, then he knows that ... that AM knows that AM doesn't knows P.)

It is clear that the Principle of Supreme Arrogance would be much too strong; in particular it runs into the older-brother arrogance problem discussed in Section 1. A theory with only the premise $L_a(L_bP \Rightarrow L_bL_bP)$ would have grounded extensions containing $L_aL_b\neg P$.

As argued before, some principle of arrogance will be necessary for MANML. We can restrict arrogance by 1) Positing that arrogance is a pairwise relation on (groups of) agents or 2) Positing that arrogance applies only to certain types of statements. Combinations of these restrictions may be used; e.g., a social studies teacher may be arrogant with respect to his students' knowledge of geography.

Here, we will explore restriction 2. In particular, we will show that a specific restriction of the principle of arrogance will permit the solution of the nested temporal reasoning problems cited in Section 1.

We begin by extending MANML2 so that it can handle temporal reasoning. We replace $\mathcal{L}$, the modal language underlying MANML1 and MANML2 by $\mathcal{L}_{\mathcal{T}}$, a temporal logic as in [McDermott, 1982]. $\mathcal{L}_{\mathcal{T}}$ contains both atemporal statements (e.g. theorems of arithmetic) and temporal statements such as $(t, IsPresident(GeorgeBush)$. If an action occurs at time t, we write *(t,Occurs(act))*. A statement of the form (t, L_aP) is usually written as $L_{a,t}P$.

We define MANML3 to be the theory that we obtain by replacing $\mathcal{L}_{\mathcal{T}}$ as the underlying language of MANML2; the definitions of stable set, the mnemonic sets, stable set expansion, and groundedness, as well as the theorem, look exactly the same, except for the extra temporal index.

It has been verified that MANML3 can handle the family of Yale shooting problems. [3] Casual rules are represented as monotonic rules; e.g., we have $L_{a,tj}((t, Occurs(load)) \Rightarrow (t+1, loaded))$. Persistence rules, on the other hand, are represented as default rules, e.g., we have $L_{a,tj}(t, alive) \wedge \neg L_{a,tj}[((t, Occurs(shoot)) \wedge (t, loaded)] \Rightarrow (t+1, alive)$. A complete formalization of the YSP problems and the underlying causal theory, as well as the proofs that these problems can be handled, appears in the expanded version of this paper.

It is also possible to represent persistence rules as monotonic, e.g., $L_{a,tj}(((t, alive) \wedge \neg((t, Occurs(shoot)) \wedge (t, loaded))) \Rightarrow (t+1, alive))$, as long as one adds to the theory the general rule: $L_{a,tj}((t, Occurs(act)) \Rightarrow L_{a,tj}(t, Occurs(act)))$ That is, every agent believes that if an action occurred, he would know that that the action occurred. This formalization also solves the YSP family. [4] We note, as an aside, that we have thus demonstrated the solution of the YSP in an autoepistemic-style language. As this solution can easily be transferred to AEL, this represents the first solution, as far as we know, of the YSP in AEL.

We now present MANML4, a multiple agent nonmonotonic logic that allows for temporal reasoning and that has restricted arrogance. MANML4 adds only rule (7), the Restricted Event Arrogance Rule to MANML3. We have:

Definition 7 *T is a MANML4 stable set if it obeys the following conditions:*
(1) - (6) of MANML3.
(7) if P is of the form $(t, Occurs(act))$*, and* $L_{a1,tj_1}\cdots L_{am,tj_m} P \notin T$*, then* $L_{a1,tj_1}\cdots L_{am,tj_m} \neg L_{am,tj_m} P \in T$*. (Restricted Event Arrogance Rule)*

We then have the following definitions and theorem: Let
$MULTI\_NI\_TEMP = \bigcup_{ai}\{L_{a1,tj_1}\cdots L_{am,tj_m} \neg L_{am,tj_m} P \mid L_{a1,tj_1}\cdots L_{am,tj_m} P \notin T$ where P is of the form $(t, Occurs(act))\}$

Definition 8 *T is a MANML-4 stable set expansion of A if T is the set of first order consequences of* $A \cup CONS \cup PI \cup NI \cup MULTI\_CONS \cup MULTI\_PI \cup MULTI\_NI\_TEMP$.

Definition 9 *T is MANML-4 grounded in a set of premises A iff* $T = \{\phi \mid A \cup CONS \cup PI \cup NI \cup MULTI\_CONS \cup MULTI\_PI \cup MULTI\_NI\_TEMP \vdash \phi\}$

Theorem 3 *T is a MANML-4 expansion of A iff T is MANML-4 grounded in A*

The Restricted Event Arrogance Rule is justified by the observation that if A reasons about B's temporal reasoning ability, it makes sense for A to assume that he is aware of all the relevant actions that B is aware of. This assumption is implicit, e.g., in story comprehension.

MANML4 can handled the nested Yale Shooting Problems. With additional restricted instances of the principle of arrogance, MANML can also handle the nested Tweety and Nixon Diamond problems. Thus, MANML can handle all of the benchmark problems in Section 1.

[3] The Yale shooting problem can briefly be described as follows: At time 1, Fred is alive. A gun is loaded at time 1. At time 3 the gun is fired at Fred. It is known that firing a loaded gun at someone causes that person to die. Moreover, guns that are loaded typically stay loaded; people who are alive typically stay alive. The question: Is Fred dead or alive at time 4? While we would expect that Fred is dead, most early non-monotonic temporal logics ran into a version of the multiple extension problem and could not predict that Fred would be dead. The original statement of the problem can be found is [Hanks and McDermott, 1986]. Numerous solutions and refutations of these solutions

[4] Note that while we have a solution to the YSP problems and in a more general sense to the temporal projection problem, we haven't strictly speaking solved the frame problem since we still need an exhaustive set of persistence rules.

3 Extending Circumscription to Multiple Agents: EMAT

We aim to eventually extend the major categories of nonmonotonic formalisms to their associated multiple agent nonmonotonic formalisms. In this section, we lay the groundwork for the extension of circumscription [McCarthy, 1980] to multiple agents. We first present EMAT, Epistemic Motivated Action Theory, a nonmonotonic model-based theory that permits multiple agent nonmonotonic reasoning in temporal contexts. We suggest a way of recasting EMAT as a circumscriptive theory, C_EMAT. C_EMAT is thus a restricted multiple agent version of circumscription. We then suggest a broad framework for extending C_EMAT to a more general theory of multiple agent circumscription.

EMAT was originally developed to solve several temporal reasoning problems that arise in multiple agent planning contexts, [5] such as the Third Agent Frame Problem [Morgenstern, 1991]): Suppose Susan would like to open a safe and needs to know the combination. She doesn't know the combination herself, but knows that her friend Bill knows someone who knows the combination. Susan doesn't know who this third agent is, but knows that he is a friend of hers. Moreover, it is known that friendly agents cooperate. Presumably, to open the safe, Susan can execute the following plan:
(1) Susan asks Bill for the name of the third agent
(2) Bill tells Susan the name
(3) Susan asks that agent for the combination
(4) The agent tells Susan the combination
(5) Susan opens the safe. [6] The problem is: the combination may (though it usually doesn't) change. How does Susan know that the third agent will still know the combination when she gets around to asking him?
It is demonstrated in [Morgenstern, 1991] first, that this problem is immune to an attack from frame axioms. No amount of frame axioms can solve the problem. Even if the third agent knows all relevant frame axioms, and knows how to reason with them, he may not know all that has happened during the course of the plan, and thus may not be able to apply the frame axioms. Second, it turns out that most current nonmonotonic temporal formalisms ([Lifschitz, 1987], [Baker, 1989]) cannot be extended in any way to handle these problems. The problem is that such formalisms are *dense* and *complete*; that is, they are based on the assumptions that one and only one action happens at a time, and that the system (or agent) knows about all the actions that occur.

In order to solve this problem, we must instead model the way in which Susan reasons, presumably as follows: she doesn't know of anything that would happen to change the combination, and doesn't know that the third agent knows of any such occurrence. Thus, she assumes that the third agent does not in fact know of such an occurrence, and will thus know the combination. Thus, she can ask the third agent for the combination, and he will tell it to her.

Susan's reasoning is based on the intuition that actions typically occur only when there is a *reason* for them to happen. Such an intuition is formalized in MAT, Motivated Action Theory [Stein and Morgenstern, 1993], a model-based theory of nonmonotonic reasoning in which we prefer models in which the fewest number of *unmotivated* actions take place. MAT is ideal for solving temporal reasoning problems such as the Yale shooting problem. Nonetheless, it cannot as it stands solve the Third Agent Frame Problem. MAT is a single-agent non-monotonic logic, and we need here the power of a multi-agent nonmonotonic logic. That is, we need to develop a formalism in which agents can reason about

[5] At the time EMAT was developed (1989), there was no general effort underway to develop multiple agent NML's. In fact, the research into EMAT first sparked this effort.

[6] A theory in which one can construct and reason about such plans and their execution, is described in [Morgenstern, 1988].

how other agents reason using the principles underlying MAT.

EMAT extends MAT in precisely this manner. EMAT integrates MAT with an epistemic logic, adding a belief operator, *Bel* to the temporal language underlying MAT. Crucial to EMAT are concepts such as "what A believes" and "what A thinks B thinks." To formalize such concepts, we extend the concept of a theory instantiation from MAT. In MAT, a theory instantiation is a collection of sentences - general rules and particular facts - about which one reasons. The extension of this concept in EMAT is a *relativized theory instantiation* – that is, a collection of sentences which one agent believes that another agent believes. More precisely, $TI(a,t)$, the relativized theory instantiation with respect to a and t, is the set of sentences that a believes at time t. For example if TI contains the sentence $(1, Bel(Bob, (1, On(X,Y))))$, then $TI(Bob, 1)$ would contain the sentence $(1, On(X,Y))$. A relativized

theory instantiation is itself a theory instantiation, so we can express an arbitrarily long relativization. We define TI(a,t1,b,t2) to be TI(a,t1)(b,t2). So, if TI contains the sentence $(2, Bel(Sue, (1, Bel(Bob, (3, At(X, Loc7))))))$, then

$TI(Susan, 2, Bob, 1) = TI(Susan, 2)(Bob, 1)$ would contain the sentence $(3, At(X, Loc7))$. The preference relation among models of a TI, introduced in MAT, is extended to a preference relation among models of a relativized TI. We also add a *Nested Projections* inference rule that allows agents to reason about the projections that other agents make. EMAT can solve nested YSP problems as well as general nested temporal projection and planning problems.

We recast EMAT as a circumscriptive theory. Using techniques described in the proof theory of MAT [Stein and Morgenstern, 1991], we define a predicate *Unmotivated* on action occurrences. We claim that circumscribing *Unmotivated* within a particular TI gives results that are, in essence, identical to MAT. [7] A parallel recasting as a circumscriptive theory is possible for EMAT as well. Specifically, we circumscribe *Unmotivated* within a particular relativized theory instantiation. We must then add the following two rules of inference:

E1:If $Circum(TI(a,t), Unmotivated) \vdash P$, then $Bel_{a,t} P$

E2:If $Circum(TI(a,t1,b,t2), Unmotivated) \vdash P$, then $Bel_{a,t1} Bel_{b,t2} P$ [8]

E1 says that agents perform circumscription on unmotivated action occurrences (and believe the results); E2 says that agents believe that other agents perform such circumscription.

This circumscriptive version of EMAT, C_EMAT, gives identical results for the variant frame problems as the original model-based version. EMAT and C_EMAT are thus examples of restricted multi-agent non-monotonic logics. It is worthwhile to analyze these multi-agent NML's using the list of requirements for such logics outlined in Section 1. We can clearly see that features (1) and (2) of a multi-agent non-monotonic logic, namely, explicit mention of agents and belief, are present here. Moreover, inter-agent reasoning rules (feature (3)) are captured by E2. However, we know that some concept of arrogance is essential for a multiple agent NML. At first glance, it seems that no such concept is present in this logic.

A careful examination of C_EMAT, however, reveals that arrogance is indeed present; it is just well disguised. Specifically, the concept of arrogance is implicit in the concept of a relativized theory instantiation being used as the base theory for the circumscription. If A reasons about the results of circumscribing *Unmotivated* in B's relativized theory instantiation, A is being arrogant with respect to B. He is implicitly assuming that his partial model of B's beliefs is sufficient for nonmonotonic temporal reasoning; that there is nothing important that B believes about actions that have occurred that A doesn't believe B believes.

[7] We have not yet proven this claim; but have demonstrated identical results for the YSP family of problems.

[8] notation taken from [Davis, 1990].

Now, since in C_EMAT, only one predicate - *Unmotivated* - is circumscribed, an agent's arrogance is not supreme. In fact, agents are arrogant with respect to statements of the form $(t, occurs(act))$ - just as in rule (7), the Restricted Event Arrogance Rule, of MANML.

This analysis suggests a natural way of extending C_EMAT to a more general theory of multiple agent circumscription. In particular, we can extend E1 and E2, above, replacing *Unmotivated* by a set of predicates. Unfortunately, if this set is large enough, this would result in supreme arrogance. Instead, it would be wiser to extend C_EMAT by only allowing the circumscription of other restricted predicates in selected relativized theory instantiations, thus restricting the principle of arrogance as suggested in section 2. Specifically, we extend E1 to a set of selected predicates G, thus allowing agents to perform circumscription with a variety of predicates. This allows agents to perform general nonmonotonic reasoning (in the single agent case) as opposed to only temporal nonmonotonic reasoning. To allow general *multiple* agent nonmonotonic reasoning, we assume a three-place predicate $Arrogant(A, B, \tilde{Q})$ where $\tilde{Q}$ is the term representing the predicate Q, (to be read: "A is arrogant with respect to B's knowledge of Q"), and restate E2 as:
if $Arrogant(a, b, \tilde{Q})$ and $Circum(TI(a, t1, b, t2), Q) \vdash P$, then $Bel_{a,t1} Bel_{a,t2} P$

We are currently working on showing that with suitable axioms on the predicate *Arrogant*, C_EMAT can handle the nested Tweety and nested Nixon Diamond problems.

4 Related Work

Virtually all research in nonmonotonic logic has focused upon the single agent case. There have been some attempts, however, to fuse aspects of multiple agent reasoning with nonmonotonic logics. Perrault [1987] and [Appelt and Konolige, 1988] have applied nonmonotonic logic to speech acts. The emphasis, there, however, is on the default assumptions that the speaker [resp. hearer] of a speech act must make about the hearer's [resp. speaker's] beliefs. Konolige [1988] has looked at an indexed L operator in his Hierarchical Autoepistemic Logic. These indices refer to different subtheories that a single agent might have. Halpern and Moses [1984] explore the formalization of the concept of "knowing only p." While this concept is well developed for the single agent case, and there is some discussion on the possibility of extending the concept to the multiple agent case, that extension was never done. Levesque [1990] has likewise formalized the concept of "all that I know", but has not extended this to multiple agents. Parikh[1984,1991] has also looked at similar topics.

There has been no work on the way in which agents reason about other agents' abilities to reason nonmonotonically. Currently, Halpern [1992] is looking at similar issues.

5 Conclusions and Future Work

We have motivated the need for a multiple agent nonmonotonic logic, and have demonstrated that any such logic must have explicit mention of agents and belief, inter-agent reasoning rules, and a good dose of arrogance. We have presented two forms of multiple agent non-monotonic logics: an extension of Moore's Autoepistemic Logic, and a multiple agent version of a circumscriptive theory of temporal reasoning.

We are currently working on two other extensions of single agent non-monotonic logics. The first is an alternative extension to AEL. In this work, we use two epistemic operators, E, a set membership modality, denoting a statement in an agent's knowledge base, and L, which more closely resembles ordinary belief. The second is an extension of Reiter's [1980] Default Logic to multiple agents. We introduce a belief operator L into Default Logic, and give an interpretation to L using default rules.

Our ultimate goal is to extend the major categories of single-agent non-monotonic logics to multi-agent non-monotonic logics. We plan to investigate the properties of and relations between these different non-monotonic logics. In addition, we plan to look at additional domain theories of commonsense reasoning and suggest further restrictions on the principle of arrogance.

Acknowledgements Thanks to Ernie Davis, Hector Geffner, Benjamin Grosof, and Wlodek Zadrozny for comments and criticism.

References

[Appelt and Konolige, 1988] Appelt, Douglas and Kurt Konolige: "A Practical Nonmonotonic Theory for Reasoning About Speech Acts," *Proceedings of the 26th Conference of the ACL*, 1988

[Baker, 1989] Baker, Andrew: "A Simple Solution to the Yale Shooting Problem," *Proceedings, KR 1989*

[Davis, 1990] Davis, Ernest: *Representations of Knowledge for Commonsense Reasoning,* Morgan Kaufmann, San Mateo, 1991

[Halpern, 1992] unpublished work on multiple agents.

[Halpern and Moses, 1984] Halpern, Joseph and Yoram Moses: "Towards A Theory of Knowledge and Ignorance," *Proceedings, AAAI Workshop on Non-monotonic Logic,* pp. 125-143, 1984

[Hanks and McDermott, 1986] Hanks, Steven and Drew McDermott: "Default Reasoning, Nonmonotonic Logics, and the Frame Problem," *Proceedings, AAAI 1986*

[Kautz, 1986] Kautz, Henry: "The Logic of Persistence," *Proceedings, AAAI 1986*

[Konolige, 1988] Konolige, Kurt: "Hierarchical Autoepistemic Theories for Nonmonotonic Reasoning," *Proceedings, AAAI 1988*

[Konolige, 1987] Konolige, Kurt: "On the Relation between Default Theories and Autoepistemic Logic," *Proceedings, IJCAI 1987*

[Levesque, 1990] Levesque, Hector: "All I Know: A Study in Autoepistemic Logic," *Artificial Intelligence 42*, 1990

[Luce and Raiffa, 1957] Luce, R. Duncan and Howard Raiffa: *Games and Decisions, Introduction and Critical Survey*, John Wiley, New York, 1957

[McCarthy, 1980] McCarthy, John: "Circumscription - A Form of Non-monotonic Reasoning," *Artificial Intelligence 13,* 1980

[McDermott, 1982] McDermott, Drew: "A Temporal Logic for Reasoning About Processes and Plans," *Cognitive Science,* 1982

[McDermott and Doyle, 1890] McDermott, Drew and Jon Doyle: "Nonmonotonic Logic I," *Artificial Intelligence 13*, 1980

[Moore, 1985] Moore, Robert: "Semantical Considerations on Nonmonotonic Logic," *Artificial Intelligence 25*, 1985

[Morgenstern, 1990] Morgenstern, Leora: A Formal Theory of Multiple Agent Non-Monotonic Logics, *Proceedings, AAAI 1990.*

[Morgenstern, 1988] Morgenstern, Leora: "Foundations of a Logic of Knowledge, Action, and Communication," NYU Ph.D. thesis, Dept. of Computer Science, 1988

[Morgenstern, 1991] Morgenstern, Leora: "Knowledge and the Frame Problem," *International Journal of Expert Systems,* 1991. Also published in Ken Ford and Pat Hayes, eds: *AI and the Frame Problem,* JAI Press, Greenwich, 1991

[Morgenstern and Stein, 1988]
Morgenstern, Leora and Lynn Andrea Stein: "Why Things Go Wrong: A Formal Theory of Causal Reasoning," *Proceedings AAAI 1988.* Expanded and revised version in Stein and Morgenstern "Motivated Action Theory,", 1993, to appear in AIJ.

[Parikh, 1984] Parikh, Rohit: "Logic of Knowledge, Games, and Dynamic Logic," *FST-TCS, Lecture Notes in Computer Science,* Vol. 181, pp. 202-222, Springer Verlag, 1984. Cited in Joseph Halpern and Moshe Vardi: "Model Checking vs. Theorem Proving: A Manifesto," *Proceedings, KR 1991*

[Parikh, 1991] Parikh, Rohit: unpublished paper

[Perrault, 1987] Perrault, Ray: "An Application of Default Logic to Speech Act Theory," *Proceedings, Symposium on Intentions and Plans in Communication and Discourse*, Monterey, 1987

[Rapoport, 1974] Rapoport, Anatol: *Game Theory as a Theory of Conflict Resolution*, D. Reidel, Boston, 1974

[Reiter, 1980] Reiter, Raymond: "A Logic for Default Reasoning," *Artificial Intelligence 13*, 1980

[Reiter and Criscuolo, 1981] Reiter, Raymond and G. Criscuolo: "On Interacting Defaults," *Proceedings, IJCAI 1981*,

[Stein and Morgenstern, 1993]
"Motivated Action Theory," to appear in *Artificial Intelligence*, 1993