Macro and Micro Attributions of Mental Attitudes to Agents[1]

Afzal Ballim
ISSCO, University of Geneva
54 Route des Acacias, CH-1227
Geneva, Switzerland

Abstract

Of the various mental states or attitudes, *belief* and *knowledge* are the two that have received most attention in artificial intelligence due to their having been considered the "core" attitudes upon which reasoning is built. In recent years the development of systems that perform complex tasks involving interaction with other entities has led to the investigation of many other mental attitudes. However, these investigations have not always benefited from the results of investigation into belief and knowledge. This paper has two goals: first, to increase the awareness that problems associated with "belief" often apply to *any* mental attitude and that solutions for belief can be adapted for those other attitudes; and, second, to show that this is the case for a particular problem, namely the attribution of "belief" to agents.

1 Introduction

Work in artificial intelligence (AI) on the modeling of mental states or attitudes of reasoning entities (agents) has traditionally concentrated on *belief* and *knowledge*, and the problems of these two mental attitudes have received much intensive investigation as a result. In recent years, however, researchers have begun to approach the problem of the construction of more complete reasoning systems that interact with other agents (human or non-human) and have found that this cannot be accomplished with just those two attitudes. Rather, a panoply of mental attitude and states must often be modeled, including *desire, intention, obligation, goals, plans, ability, choice, commitment, hope,* and *perception.*

While research into these other mentalistic notions has benefited to a certain degree from research into knowledge and belief, there may be yet greater benefit possible than has so far been achieved. To this end, this paper makes two contributions:

(i) It is not often realised that many of the problems associated with "belief" apply, in general, to any notion that demarcates a "mental universe". Similarly, the solutions to these problems can be applied to those other notions. A number of such problems are demonstrated, and their solutions applied to various other mental attitudes.

(ii) As a special case of (i), the problem of attributing mental attitudes is considered, and experience gained from investigating belief attribution is applied to the question of ascribing other mental attitudes.

2 Problems of Belief = Problems of Mental Boundary Markers

The emphasis of the our previous work has been on properties of the mental state of belief (Ballim (1987), Ballim (1992b), Ballim (1993), Ballim & Wilks (1990), Ballim & Wilks (1991a), Ballim & Wilks (1991b), Wilks & Ballim (1987)). In particular, we have been concerned with the practical problem of determining the contents of belief states of different agents from the perspective of a system that is interacting with them via dialogue. In other words, the formation of nested models of the beliefs of others, and recursively their beliefs of each other and of the system. This also entails investigating questions of belief representation and maintenance, and the problems that lie therein.

The objective of this section is to highlight that various problems that are classically considered problems of belief are problems for any mentalistic notion which can be considered to define a boundary between *mental universes*. This

[1]This work was supported by the Fonds National Suisse de la Recherche Scientifique, grants no. 21-30156.90 and 20-33903.92

term, "mental universe" needs some explanation. The intuition is that a mental universe is a collection of "terms" (or, what one might loosely call "concepts") that are the objects of "thought" of an agent.

Of course, a topic that is of concern for any system the reasons about/with the mental states of agents concerns the objects about which the system expresses propositions. More precisely, presuming that the representation is built of terms and functions and relations between the terms, what then is the status of those terms?

One approach might be to take a straightforward model-theoretic method and say that these terms form a domain such that each object in the domain has a *denotation* in the world. This has a certain appeal (one can make claims about the system having a model-theoretic semantics of the world), but with respect to attitudes of agents towards propositions, it has serious problems. Not the least of these is that terms that appear to mean the same thing (by which it is meant that they refer to the same thing in the world) do not necessarily have the same meaning when used in belief contexts, nor (as will be shown further down) in contexts of other mental attitudes.

The prototypical example of this comes from Frege (Frege (1892)). The terms "Venus", "the Evening Star", and "the Morning Star" all refer to the same object, Venus. However, each is slightly different in the *sense* in which it refers to Venus. "Venus" is used generically to indicate the second planet from the sun. "The Evening Star" is used to identify the first bright star seen in the early evening sky. This "star" happens to be, in reality, the planet Venus. Similarly, "the Morning Star" identifies the last bright star seen in the early *morning* sky, which also happens to be, in reality, Venus. However, at one time they were believed by astronomers to be different objects, with the "Evening Star" and the "Morning Star" believed to actually be stars.

Thus, it is quite possible that someone could have beliefs about all these terms, without knowing that they are in fact the same object. In trying to tackle this sort of problem, Frege distinguished between the *sense* of a term (what he called the *Sinn*) and its *referent* (what he called the *Bedeutung*). However, for various technical reasons this distinction proved difficult to formulate, so Russell (1905) developed the notion of *intensional*[1] sentences. If replacing terms in a sentence by other terms with the same reference does not preserve the truth or meaning of the sentence, then that sentence is said to be an intensional sentence and the terms are said to be intensional terms.

As has been pointed out by many in AI (Brachman (1977), Brachman (1979), Maida & Shapiro (1982), McCarthy (1979), Woods (1975)) for a system that represents beliefs it is necessary that intensionality be accounted for, because it leads to problems of *referential opacity* in belief sentences. That is where substitution of equal terms in logical sentences that involve beliefs is not valid. From this, if, for example, Silvia believes that Venus is a planet, and that the evening star is Venus, it does not logically follow that she believes that the evening star is a planet.

2.1 Skolemisation and Mental Attitudes

Referential opacity leads to a number of problems. One of these is the skolemisation of propositions involving belief operators and existential quantifiers. Consider the following example, adapted from Konolige (1986, pp. 38), where in the first proposition (1) we wish to express that the detective Hercule Poirot knows who the murderer is in some case he is investigating.

$$\exists x.Bel(HerculePoirot, Murderer(x)) \tag{1}$$

In the second proposition (2) we wish to express that Hercule Poirot knows that *someone* is the murderer but we do not make any commitment towards him knowing who that person is.

$$Bel(HerculePoirot, \exists x.Murderer(x)) \tag{2}$$

In a normal logical language we can replace existentially quantified variables (such as the "x" in equations 1 and 2) by a constant that represents an object of which the existentially quantified statement is true. This replacement is known as *skolemisation* and the resulting constant is called a *skolem constant*. Doing this in proposition 1 results in 3 where c is the skolem constant.

$$Bel(HelculePoirot, Murderer(c)) \tag{3}$$

Similarly, skolemising proposition 2 results in proposition 4.

$$Bel(HelculePoirot, Murderer(c)) \tag{4}$$

As can be seen, skolemising the two different proposition has transformed them into the same proposition. Thus, the status of existentially quantified variables, that have an intervening belief operator between them and their quantifier,

[1] To be distinguished from intentional, which is concerned with the intent of something.

must be different from those that have no such intervening belief operator. In other words, quantification is sensitive to intensional contexts.

A number of solutions have been proposed to this problem (see Ballim (1992b), Konolige (1986)). However, it appears to have gone largely unnoticed that this problem applies equally to other mental attitudes. Consider, for example, propositions 5 and 6, the first intended to convey that Hercule Poirot wants to capture someone in particular, and the second that he wants to capture someone (though not some particular individual).

$$\exists x.Want(HerculePoirot, Capture(x)) \tag{5}$$

$$Want(HerculePoirot, \exists x.Capture(x)) \tag{6}$$

Skolemising either 5 or 6 would result in a proposition like 7, which looses the difference between them.

$$Want(HerculePoirot, Capture(c2)) \tag{7}$$

In fact, replacing "Want" or "Bel" by other attitude operators illustrates that the problem applies to them as well (see 8–11).

$$\exists x.Goal(HerculePoirot, Capture(x)) \tag{8}$$

$$Goal(HerculePoirot, \exists x.Capture(x)) \tag{9}$$

$$\exists x.Intend(HerculePoirot, Capture(x)) \tag{10}$$

$$Intend(HerculePoirot, \exists x.Capture(x)) \tag{11}$$

As with the "belief" case, skolemisation would map the two forms to an single form, again losing the difference. Luckily, the solutions (whichever solution one accepts) for this problem applied to belief can work equally well for the other attitudes.

2.2 *De Dicto* and *De Re* Attitude Reports

In the philosophical literature on belief there is much discussion of the distinction between *de dicto* belief, and *de re* belief and the ability to distinguish these two is taken as a fundamental problem for representing belief.

To understand the problem, consider the following situation (following Rapaport (1986)): let us suppose that Ralph while working late at night sees someone, who he knows to be the janitor, stealing government papers. We could then say:

$$\textit{Ralph believes that the janitor is a spy} \tag{12}$$

Further, suppose that unknown to Ralph the janitor has just had the fortune to win the lottery. In that case as the janitor is the lottery winner we could also report:

$$\textit{Ralph believes that the lottery winner is a spy} \tag{13}$$

In an obvious sense this is incorrect. Ralph does not believe that the lottery winner is a spy because he has nothing to connect the lottery winner to the janitor, who he does believe is a spy. However, in another sense it is true because the person who Ralph believes to be a spy is also the lottery winner.

In the traditional view, the first is a referentially opaque context and this type of belief report is called *de dicto*. The second is considered to be a referentially transparent context (so equivalent terms can be substituted) and this type of belief report is called *de re*. The interpretation of the two sentences must therefore be different. It would be incorrect to simply represent sentence 12 by proposition 14 if sentence 13 is represented by proposition 15.

$$Bel(Ralph, is\_spy(janitor)) \tag{14}$$

$$Bel(Ralph, is\_spy(lottery\_winner)) \tag{15}$$

One solution to this problem is to explicitly have different *de dicto* and *de re* belief operators (e.g., *BelDD* and *BelDR*), although other solutions exist which do not require this (see Ballim (1992b)). As with the skolemisation problem, it is easy to create a scenario from which it can be seen that the interpretation of other mental attitudes can also have *de dicto* and *de re* interpretations. Consider, for example in our story about Ralph and the janitor, interpreting the following sentences:

Ralph wants to capture the janitor. (16)

Ralph wants to capture the lottery winner. (17)

Ralph saw the janitor steal the documents. (18)

Ralph saw the lottery winner steal the documents. (19)

Again, the solutions for the *de dicto/de re* problem for belief can be applied to the other mentalistic notions. The point being made is that there are a large number of problems of belief which arise from the fact that belief demarcates the propositions and mental universe of an individual. However belief is not unique in doing this. Other mentalistic notions also mark the boundaries of an individuals mental universe, thus the same problems apply to them, and (happily) the same solutions may be used.

3 Attributing Mental Attitudes to Agents

The general process of assigning content to the beliefs of others we have termed *belief attribution* (Ballim (1992a), Ballim (1992b)). We have further divided this into two types:

- *belief interpretation*: which is the process of attributing beliefs to agents based on interpreting their utterances or actions;
- *belief ascription*: which is the process of making general attributions to an agent based on principles of commonality, and on general classification information concerning the agent.

We refer to instances of the former of these as *micro-attributions*, because they depend on the application of particular rules to the actual utterances and actions, producing a small number of attributions as a result. In effect, they act at a micro-level, making individually argued attributions, which are then (in general) only revoked by stronger counter-arguments.

Instances of the latter process are referred to as *macro-attributions*, because they use general principles (such as assuming commonality of belief amongst agents, unless there is evidence to contradict instances of this commonality) to attribute groups of belief to an agent. In effect, they may be said to act at a macro-level, making collections of attributions that are argued for as a collection, and hence individual attributions may easily be revoked by either more specific collective arguments, or by individual counter-arguments.

Note that the terms "micro" and "macro" are not chosen to belittle the notion of micro-attribution, but to indicate the scale at which each works. It is our contention that both are necessary: macro-attributions to provide a quick, large model of the beliefs of the other agent; and, micro-attributions to provide a well honed model. Thus individual attributions made by macro-attribution are generally more open to refutation than those made by micro-attribution.

3.1 Belief Ascription

Our work to now has concentrated on macro belief attributions with, more recently (Ballim (1992b)), investigation of its interaction with micro belief attribution. Belief interpretation is a process which has received attention in philosophy (Bach, 1982; Dennett, 1982) as well as in AI (Kass & Finin, 1988; Maida, 1986; Shadbolt, 1983; Wahlster & Kobsa, 1985).

Belief ascription is the attribution of groups of beliefs to an agent based on various principles of commonality. There are two predominant methods for dynamically making such ascriptions of beliefs, etc., to other agents. One method is to have stereotypical models, i.e., pre-existent models that fit stereotypical groups of people. Then, by

determining which stereotypes fit an individual we can ascribe the beliefs of those stereotypes to the agent. This method is wide-spread within user-modeling systems (see, for example, Chin (1989), Rich (1989)).

A second method is to take the system's beliefs as a starting point, and perturb them in some fashion to generate the beliefs of other agents. The basis of this method is the assumption that for the most part our beliefs are consensual, i.e., that most of the information necessary for communication is assumed to be mutually shared between agents. This notion has been the basis for proposals by Wilks & Bien (1979) and Wilks & Bien (1983), whereby beliefs can be ascribed to an agent unless there is explicit evidence against the ascription.

The simplest form of this ascription algorithm, first described in Wilks & Bien (1979) and Wilks & Bien (1983), is that a nested model (called a *viewpoint,* because it represents someone's view of someone else) should be generated using a default rule for ascription of beliefs that reflects the assumption of commonality. The default ascriptional rule is to assume that another person's view is the same as one's own *except where there is explicit evidence to the contrary.*

Before continuing, it will be useful to discuss briefly the important types of environments used in the basic perturbation ascription algorithm. An environment is a small, explicit grouping of beliefs, collected under some form of indexing (so, when necessary, may be considered as typed, labelled sets). The particular form of indexing is the initial basis for determining the *type* of any given environment.

For the basic perturbation ascription algorithm, there are two important types of environment:

(i) *viewpoints,* environments that represent a particular agent's point of view;

(ii) *topics,* environments that contain beliefs that are relevant to a given subject (the topic).

A belief environment will be denoted by the symbol $\mathcal{B}$, with a superscripted argument denoting the agent whose viewpoint it is. Thus, the system's belief environment could be indicated by $\mathcal{B}^{system}$. A group of objects (propositions or environments) which are within a viewpoint can be referred to by enclosing them in square parentheses, and juxtaposing with the symbol for the viewpoint. So, referring to two viewpoints (say, John's and Jim's) that are within the system's viewpoint can be done as in Equation 20.

$$\mathcal{B}^{system}\left[\begin{array}{c}\mathcal{B}^{John}\\ \mathcal{B}^{Jim}\end{array}\right] \tag{20}$$

A topic environment will be denoted by the symbol $\mathcal{A}$ with a superscripted argument indicating the topic, and using the same mechanism as for viewpoints to indicate individual contents or groups of them. So, a topic environment about atoms could be indicated by $\mathcal{A}^{atom}$. The systems' beliefs about atoms could then be represented as in Equation 21.

$$\mathcal{B}^{system}\,\mathcal{A}^{atom}\left[\begin{array}{c}Light(atom)\\ Small(atom)\end{array}\right] \tag{21}$$

Belief ascription operators can then be defined based on different criteria for projection the contents of one environment into another. The belief ascription program *ViewGen* which is described in Ballim & Wilks (1991b) uses an ascription process which is based on the following:

Definition 1 (Ungrounded Environment) *An ungrounded environment is a (potentially sorted) set of (possibly) sentences in the internal representation, and (again, possibly) grounded environments. An ungrounded environment will generally be written as* $\mathcal{E}_i$*, for* $i \geq 0$.

Definition 2 (Grounded Environment) *A grounded environment is a tuple* $\langle T, L, \mathcal{E}\rangle$*, where* T *is an environment type,* L *is a label appropriate for the type* T*, and* $\mathcal{E}$ *is an ungrounded environment.*

Definition 3 (Ungrounded Environment Proper Subset) *Let* $\mathcal{E}_0$ *and* $\mathcal{E}_1$ *be two ungrounded environments.* $\mathcal{E}_0$ *is said to be an ungrounded environment proper subset of* $\mathcal{E}_1$ *if the contents of* $\mathcal{E}_0$ *are a proper subset of the contents of* $\mathcal{E}_1$*. This is written as* $\mathcal{E}_0 \overset{\mathcal{E}}{\subset} \mathcal{E}_1$.

Definition 4 (Ungrounded Environment Improper Subset) *Let* $\mathcal{E}_0$ *and* $\mathcal{E}_1$ *be two ungrounded environments.* $\mathcal{E}_0$ *is said to be an ungrounded environment proper subset of* $\mathcal{E}_1$ *if the contents of* $\mathcal{E}_0$ *are an improper subset of the contents of* $\mathcal{E}_1$*. This is written as* $\mathcal{E}_0 \overset{\mathcal{E}}{\subseteq} \mathcal{E}_1$.

Definition 5 (Ungrounded Environment Union) *Let $\mathcal{E}_0$ and $\mathcal{E}_1$ be two ungrounded environments. The ungrounded environment union of the two is defined to be the ungrounded environment whose contents is the union of the contents of $\mathcal{E}_0$ and $\mathcal{E}_1$. If $\mathcal{E}_2$ is this ungrounded environment, then it is written as $\mathcal{E}_0 \overset{\mathcal{E}}{\cup} \mathcal{E}_1 \overset{\mathcal{E}}{=} \mathcal{E}_2$.*

Definition 6 (Permissive Ascription) *Let $\mathcal{B}^{source}$ be an environment from which a ascription is being made, and let $\mathcal{B}^{target}$ be the target environment of that ascription.*

Define the permissive ascriptions candidates (PA) such that

$$PA = \left\{ \mathcal{E} \;\middle|\; \begin{array}{l} \mathcal{E} \overset{env}{\subseteq} \mathcal{B}^{source} \\ \mathcal{E} \overset{env}{\cup} \mathcal{B}^{target} \text{ is consistent} \\ \neg\exists \mathcal{E}_i \;\middle|\; \begin{array}{l} \mathcal{E} \overset{env}{\subset} \mathcal{E}_i \\ \text{and } \mathcal{E}_i \overset{env}{\cup} \mathcal{B}^{target} \text{ is consistent} \end{array} \end{array} \right\}$$

The permissive ascription environment (PAE) is then defined to be the environment $\mathcal{E}^{max}$, which is the intersection over the elements of PA.

$$\mathcal{E}^{max} = (\overset{env}{\bigcap} PA) \overset{env}{\cup} \mathcal{B}^{target}$$

This operation must then be coupled with a process that determines the set of environments relevant to the nesting under construction and orders them so that the ascription is performed iteratively over the ordered relevant environments (Ballim & Wilks, 1990).

Our work has fused the two types of belief ascription (stereotypes and perturbation) via the more general notions of *typicality* and *atypicality* — dual notions that allow us to express what sort of agents typically hold particular beliefs, and which beliefs are atypical to particular agents. This is done by considering typicality and atypicality as an expression of the *competency* that some agent (or group of agents) has in holding a belief. In other words, if we can say that only particular groups have the competency to hold a particular belief (say P) then this is equivalent to saying that the belief P is atypically held by that group. This notion of competency is represented via lambda expressions whose evaluation depends on the environment in which they are being evaluated. The exact details of this are too complex to discuss here, and the interested reader is referred to Ballim & Wilks (1991b) and Ballim (1992b) for further information.

3.2 Adapting Belief Ascription for General Attitude Ascription

Consider now the attribution of mental attitudes besides that of belief. At first sight this has been extensively covered in the literature. In, particular, work dealing with *speech acts* (Austin, 1962; Searle, 1969) has been intensively concerned with the identification of intentions, plans, goals, etc., from utterances (e.g., Allen & Perrault (1980), Cohen & Levesque (1980), Perrault & Allen (1980)) as has much other work on language understanding (e.g., Carberry (1990), Konolige & Pollack (1989), Pollack (1987), Sidner & Israel (1981)).

However, these studies have at least one thing in common, which is that they are all influenced by the notion of analysing utterances (or behaviour) to determine the mental attitudes of the utterer. In other words, to adapt the dichotomy previously employed, they are *micro-attributions* of mental attitudes based on direct arguing from the utterances of the agents to whom the attributions are being made. The question is then raised: if these are micro-attributions, then is there some macro-attribution for other mental attitudes such as has been identified for belief?

In this paper it is claimed that the macro-attribution mechanism for belief can be adapted easily for other attitudes. In other words, the notions of commonality: typicality, and atypicality as applied to belief, can be adopted for other mental attitudes. This has a number of advantages: (i) entire groups of attitudes can be ascribed to agents, reducing the amount of work necessary; (ii) the process of micro-attributions may then be simplified to take into account the attitudes attributed by the macro-attributions.

Two distinct possibilities to which one might consider adapting belief ascription to other attitudes will be considered:

(1) The attitude to which we are adapting is to be used as the final attitude in a nested belief context. For example, "the System believes John believes Mark has goal X," or "the System believes David intends Y."

(2) The attitudes may form an arbitrary nesting. For example, "the System intends that John has goal Z," or "the System has a goal that Graham believes W."

The first of these possibilities is referred to in this paper as *simple adaptation* because it can be seen as a special case of belief ascription where the proposition(s) being ascribed are other attitudes. It requires that for any attitude which we wish to adapt, propositional attitudes (PAs) of that form are stored (whenever possible) with respect to agent groups for whom it is typically held in the same way as beliefs are in the *ViewGen* system, and, similarly, they are partitioned according to topics. Thus, for example, the obligation to go regularly to church could be considered a typical obligation of religious people about religion, the desire to meet their favourite film-star could be a typical desire of movie-fans about film-stars, and plans to move blocks could be typical plans for block-moving computer programs about blocks.

The main change is determining to what extent the commonality principle (that other's have roughly the same beliefs as ourselves) can be applied to other attitudes, i.e., to what extent might we assume that other agents have the same desires, goals, abilities, intentions, hopes, etc. as ourselves. While it is certainly true that not everyone has the same abilities, hopes, desires, etc., we are probably safe in assuming a large degree of commonality for most of the propositional attitudes as long as we have the ability to distinguish between typical and atypical attitudes, hence the belief ascription method should adapt directly to other attitudes.

Consider a simple example where the system has the desire about football hooligans that they be barred from attending matches (Equation 22). Now, using the normal belief ascription mechanism this could be ascribed to someone else, say John, resulting in Equation 23.

$$\mathcal{B}^{system}\mathcal{D}^{system}\mathcal{A}^{hooligans}[barred(hooligans)] \tag{22}$$

$$\mathcal{B}^{system}\mathcal{B}^{John}\mathcal{D}^{John}\mathcal{A}^{hooligans}[barred(hooligans)] \tag{23}$$

On the other hand, hooligans would (atypically) hold the desire that they should not be barred, and so the ascription would fail for them.

What then of the second possibility: adapting belief ascription to be used in arbitrary nested attitudes. Simple adaptation works because we are, in effect, simply treating the other attitudes as special "propositions". We must first determine what iterations make sense with the second possibility (e.g., does it make sense to form the nesting "I desire that John has the ability that Mark has the ability that David intends that Philip hopes that the holligans are barred"?). A start is to take each ordered pair of attitudes and state if they form an acceptable nesting. Table 1 is an initial proposal for which ordered pairs (row followed by column) might be acceptable[2] with "$\surd$" indicating an acceptable pair; "$\surd$?" indicating acceptable although debatable; "$\times$" indicating not acceptable; "$\times$?" indicating not acceptable although debatable; and "?" indicating that it's debatable.

	Belief	Knowl.	Desire	Intent.	Oblig.	Goal	Plan	Ability	Choice	Commit.	Hope	Percept.
Belief	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Knowl.	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Desire	$\surd$	$\surd$	$\surd$	$\surd$?	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Intent.	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Oblig.	$\surd$	$\surd$?	$\surd$?	$\surd$?	$\surd$?	$\surd$	$\surd$	$\surd$?	$\surd$	$\surd$	$\surd$	$\surd$
Goal	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Plan	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Ability	$\surd$	$\surd$	$\surd$	$\surd$?	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Choice	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$?	$\surd$	$\surd$	$\surd$	$\surd$?
Commit.	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$?	$\surd$	$\surd$	$\surd$	$\surd$
Hope	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$
Percept.	$\surd$?	$\surd$?	$\surd$	$\surd$	$\surd$	$\surd$?	$\surd$?	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$

Table 1: Valid Attitude Orderings

[2]Remember that a nested pair generally may be of the form *"I x I y P"* or *"He1 x He2 y P"* where "x" and "y" are attitudes; "I", "He1", "He2" are agents; and "P" is the object of the nesting.

Assuming, then, that the acceptability of a nesting of attitudes depends on the nested pairs within that nesting (an assumption which will need to be validated) it is possible to say if a nesting is valid or not. The possibly surprising result of this reflection is that there seems to be no pair for which one can say outright that it is invalid, although some pairs are more dubious than others. An area to be explore further is to redo this table so that there is one for introspective nestings, and another for non-introspective nestings, to see if this makes a difference.

The next question, then, is to determine for which pairs of attitudes does it make sense to assume commonality along with notions of typicality and atypicality to allow ascription from one to another: e.g., does it make sense to say that if A hopes that P, then A hopes that B plans that P? Tables 2 and 3 are initial proposals along this line, the former indicating commonality for introspective ascription, and the latter for non-introspective ascription. These tables are to be read as follows: look at the entry for attitude at row i, and then find the cell for the attitude at column j. This represents the validity of using commonality to make the ascription from $i_{a0}P$ to $i_{a0}j_{a1}P$ (where $a0$ and $a1$ are different agents for non-introspection, but the same agent for introspection).

	Belief	Knowl.	Desire	Intent.	Oblig.	Goal	Plan	Ability	Choice	Commit.	Hope	Percept.
Belief	√	×?	×	×	×	×	×	×	×	×	×	×
Knowl.	√	√	×	×	×	×	×	×	×	×	×	×
Desire	√	√	√	√?	×	√	√	√	×	√	√	√
Intent.	√	√	√	√	×	√	√	√	√	√	√	√
Oblig.	×	×	×	√	×	×	×	×	×	√	×	×
Goal	×	×	×	×	×	×	√	√	×	×	×	×
Plan	×	×	×	×	×	×	×	√	×	×	×	√
Ability	√	√	×	×	×	√	√	?	√	√	√	√
Choice	×	×	√	√	×	√	×	×	√	√	×	×
Commit.	√?	√?	√	√	√	√	×	×	√	√	×	×
Hope	×	×	√	√?	×	√	√	√	√	√	√	×?
Percept.	√	√	×	×	×	×	×	×	×	×	×	√

Table 2: Introspective Commonality Ascription Validity

	Belief	Knowl.	Desire	Intent.	Oblig.	Goal	Plan	Ability	Choice	Commit.	Hope	Percept.
Belief	√	×	×	×	×	×	×	×	×	×	×	×
Knowl.	√	√	×	×	×	×	×	×	×	×	×	×
Desire	?	?	√	√?	×?	×?	×?	×?	×?	×?	√?	×?
Intent.	√	√?	√?	√?	×?	×?	×?	×?	×?	×?	×?	√?
Oblig.	×?	×	×	×	×	×	×	×	×	×	×	×
Goal	√	√	×?	×	×	×?	×	×	×	×	×	√?
Plan	×	×	×	×	×	×	×	×	×	×	×	×
Ability	?	×	×	×	×	×	×	×	×	×	×	√?
Choice	×	×	×	×	×	×	×	×	×	×	×	×
Commit.	×?	×	×	×	×	×	×	×	×	×	×	×
Hope	×?	×	√?	×?	×?	×?	×?	×?	×?	×?	×?	×?
Percept.	×	×	×	×	×	×	×	×	×	×	×	×

Table 3: Non-Introspective Commonality Ascription Validity

In interpreting these two tables, it must be remembered that their purpose is to allow us to decide what nesting might be made using commonality and the notions of typicality and atypicality, so the nestings generated would be easily defeasible and would allow for the ascription of attitudes to an agent based on atypical properties of the agent. So, where there is "√" in the table it is proposed that ascription is possible using the principle of commonality, and where there is "√?", "?", or "×?" ascription might be possible by use of atypicality properties. It is proposed, therefore, that the belief ascription algorithm can be adapted to form nestings where no pair in the nesting are given as "×" in the above tables.

4 Conclusion

This paper first described how certain problems of belief hold for any mentalistic notion, but that the solutions to these problems of belief can also be used for those other notions. The idea that solutions in "belief" can be applied to the other notions was taken one step further, by postulating that belief ascription (a macro-attribution method for beliefs) could be useful in making macro-attributions of other attitudes, etc. In the simple case where the other notions are the last in a nesting of beliefs, this is simple. In constructing a nesting of arbitrary it is not so simple.

A number of steps where taken towards determining the applicability of macro-attribution to other attitudes: (1) the mental notions were examined to see which pairs formed sensible nestings; (2) it was suggested for which pairs the ascription mechanism might reasonably work. The suggestions here need to be born out by experiment. This is the next phase of this work

5 Bibliography

Allen, J. F. & C. R. Perrault (1980) "Analyzing Intention in Utterances," *Artificial Intelligence* 15, 143–178.

Austin, J. L. (1962) *How to do things with words*, Oxford University Press, Oxford.

Bach, K. (1982) "*De Re* Belief and Methodological Solipsism," in *Thought and Object*, A. Woodfield, ed., Oxford University Press, Oxford, 122–151.

Ballim, A. (1987) "The Subjective Ascription of Belief to Agents," in *Advances in Artificial Intelligence (Proceedings of AISB-87)*, J. Hallam & C. Mellish, eds., John Wiley & Sons, Chichester, England, 267–278.

Ballim, A. (1992a) "What's needed in a framework for nested beliefs and other attitudes," Proceedings of the AAAI Spring Symposium on Propositional Knowledge Representation, AAAI.

Ballim, A. (1992b) "ViewFinder: A Framework for Representing, Ascribing and Maintaining Nested Beliefs of Interacting Agents," Ph.D. Dissertation, Dept. d'Informatique, Université de Genève, Geneva, Switzerland.

Ballim, A. (1993) "Propositional Attitude Framework Requirements," *Journal for Experimental and Theoretical Artificial Intelligence (JETAI)* to appear.

Ballim, A. & Y. Wilks (1990) "Relevant Beliefs," *Proceedings of ECAI-90*, Stockholm., 65–70

Ballim, A. & Y. Wilks (1991a) "Beliefs, Stereotypes and Dynamic Agent Modeling," *User Modeling and User-Adapted Interaction* 1 (1), 33–65.

Ballim, A. & Y. Wilks (1991b) *Artificial Believers*, Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Brachman, R. J. (1977) "What's in a Concept: Structural foundations for semantic networks," *International Journal for Man-Machine Studies* 9, 127–152.

Brachman, R. J. (1979) "On the Epistemological Status of Semantic Networks," in *Associative Networks*, N. Findler, ed., Academic Press, New York, 3–50.

Carberry, S. (1990) *Plan Recognition in Natural Language Dialogue*, MIT Press, Cambridge, MA.

Chin, D. N. (1989) "KNOME: Modeling what the User Knows in UC," in *User Models in Dialog Systems*, A. Kobsa & W. Wahlster, eds., Springer, Berlin, Heidelberg, 74–107.

Cohen, P. R. & H. J. Levesque (1980) "Speech acts and recognition of shared plans," *Proceedings of the 3rd Biennial Conference of the Canadian Society for Computational Studies in Intelligence.*

Dennett, D. (1982) "Beyond Belief," in *Thought and Object*, A. Woodfield, ed., Clarendon Press, Oxford, 1–95.

Frege, G. (1892) "Ueber Sinn und Bedeutung," in *Readings in Philosophical Analysis (English translation)*, H. Feigl & W. Sellars, eds., Appleton-Century-Crofts, New York, NY (1949).

Kass, R. & T. Finin (1988) "Modeling the User in Natural Language Systems," *Computational Linguistics* 13, 5–22.

Konolige, K. (1986) *A Deduction Model of Belief*, Morgan Kaufmann, Los Altos, CA.

Konolige, K. & M. Pollack (1989) "Ascribing Plans to Agents," *Proceedings of the International Joint Conference on Artificial Intelligence.*, 924–930

Maida, A. S. (1986) "Introspection and reasoning about the beliefs of other agents," *Proceedings of the 8th Annual Conference of the Cognitive Science Society*, Hillsdale, NJ., 187–195

Maida, A. S. & S. C. Shapiro (1982) "Intensional concepts in propositional semantic networks," *Cognitive Science* 6, 291–330.

McCarthy, J. (1979) "First Order Theories of Individual Concepts and Propositions," *Machine Intelligence* 9, 120–147.

Perrault, C. R. & J. F. Allen (1980) "A Plan-Based Analysis of Indirect Speech Acts," *American Journal of Computational Linguistics* 6, 167–182.

Pollack, M. E. (1987) "Some Requirements for a Model of the Plan Inference Process in Conversation," in *Communication Failure in Dialogue and Discourse*, R. G. Reilly, ed., North-Holland, Dordrecht, 245–256.

Rapaport, W. J. (1986) "Logical Foundations for Belief Representation," *Cognitive Science* 10, 371–422.

Rich, E. (1989) "Stereotypes and User Modeling," in *User Models in Dialog Systems*, W. Wahlster & A. Kobsa, eds., Springer–Verlag.

Russell, B. (1905) "On Denoting," in *Readings in Philosophical Analysis*, Feigl & Sellers, eds., Appleton-Century-Crofts, New York, NY (1949).

Searle, J. (1969) *Speech acts: an essay in the philosophy of language*, Cambridge University Press, Cambridge.

Shadbolt, N. (1983) "Processing reference," *Journal of Semantics* 2, 63–98.

Sidner, C. L. & D. J. Israel (1981) "Recognizing Intended Meaning and Speakers' Plans," *International Joint Conference on Artificial Intelligence*, Los Altos, CA., 203–208

Wahlster, W. & A. Kobsa (1985) "Dialog–Based User Models.

Wilks, Y. & A. Ballim (1987) "Multiple Agents and the Heuristic Ascription of Belief," *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, 118–124.

Wilks, Y. & J. Bien (1979) "Speech Acts and Multiple Environments," *Proceedings of the IJCAI-79*, Los Altos, CA., 968–970

Wilks, Y. & J. Bien (1983) "Beliefs, points of view and multiple environments," *Cognitive Science* 8, 120–146.

Woods, W. A. (1975) "What's in a Link: Foundations for Semantic Networks," in *Representation and Understanding: Studies in Cognitive Science*, D. G. Bobrow & A. M. Collins, eds., Academic Press, New York, 35–82.