

PROPOSITIONAL ATTITUDES: FREGEAN REPRESENTATION AND SIMULATIVE REASONING

Lewis G. Creary

Artificial Intelligence Laboratory, Computer Science Department
Stanford University; Stanford, California 94305

ABSTRACT: This paper describes briefly a general notation that has been constructed for representing information about propositional attitudes, and a method that is being developed for using this notation to reason about such information. The notation is Fregean in spirit, and is based on McCarthy's first-order theory of individual concepts and propositions, extended to permit adequate representation of iterated propositional attitudes, and with a semantically simpler representation of quantified propositions and "definite-description" concepts of individuals. The notation provides a systematic and intuitively motivated way of dealing with apparent failures of extensionality and quantification into intentional constructions. The reasoning method is knowledge-based, and uses a program's own ordinary inferential apparatus to "simulate" the reasoning of other intelligent organisms. The method thus avoids any need for redundant re-representation of information about propositions at higher levels in the hierarchy of concepts, and provides a natural approach to making realistic inferences about the propositional attitudes of others.

Topics and Key Words: Representation of knowledge, formal representations, propositional attitudes, intentional contexts, reasoning about knowledge and belief.

Human beings are able (unconsciously) to represent and use effectively a wide range of information concerning people's beliefs, knowledge, preferences, and goals -- both their own, and those of others. Just how they manage to do this, however, is not at all well understood. Investigators in artificial intelligence, for example, have just begun to confront seriously some of the more complex and subtle problems involved in explicitly representing and reasoning about such information. These problems include those of i) dealing with apparent failures of extensionality, ii) quantifying into intentional constructions, iii) representing iterated propositional attitudes, and iv) making realistic inferences about the reasoning of intelligent organisms. This paper discusses an approach to these problems that is being developed by the author.* We shall indicate the main desiderata that motivate our approach, describe briefly a general notation we have constructed for representing information concerning propositional attitudes, sketch a method that we are developing for the use of this notation in reasoning about such information, compare our approach with some alternatives, and indicate the present status and future direction of our research along these lines.

1 Desiderata

Largely because we desire a representation and reasoning system of potentially great generality, we adopt the following, somewhat idealized, criteria to guide the system's development:

a. The primary representation employed should be declarative, should have at least the logical power of a first order predicate calculus, and (with suitable choice of vocabulary) should be capable of representing almost any information concerning propositional attitudes that a person could understand or express.

b. Reasoning about propositional attitudes should be based on separately represented knowledge concerning the organisms holding the attitudes; empirical assumptions should not be built into either the top-level reasoning mechanism or the basic representation language.

2 Representing Propositional Attitudes

Our representational scheme has its roots in the semantic theory of Gottlob Frege [6], and stems more directly from John McCarthy's recent use of Fregean ideas in the construction of a first-order theory of individual concepts and propositions [McCarthy 1977, 1979]. Our language is a modification and extension of McCarthy's. In order to use most efficiently our limited space here, in this section we shall first describe the philosophical framework within which our own representation language is developed, then explain the main features of our language, and finally indicate its major points of difference from McCarthy's language and the reasons for them.

2.1 The Philosophical Framework

The basic idea of our representational scheme is that propositional attitudes are relations between people (or robots) and propositions, with both terms of the relation being taken as members of the domain of discourse. We regard propositions (and also concepts) as abstract entities, in much the same way as we do numbers, sets, etc. Unlike the latter, however, propositions (and concepts) are for us abstractions of things psychological; they are language-independent (except for the language of thought) and person-independent components of situations involving belief, knowledge, desire, and the like. Because it is possible for a person to discover that one of his beliefs is logically equivalent to a proposition that he has never believed, we are led to reject logical equivalence as a criterion of identity for propositions. Though this criterion is often adopted for logical purposes, our acceptance of such an unrealistic idealization would violate the second desideratum mentioned above.

* This research was supported by the Advanced Research Projects Agency of the Department of Defense under ARPA Order No. 2494, Contract MDA903-76-C-0206.

In general, we view the objective concepts that figure in our semantic theory as abstractions of possible subjective concepts, which are in turn possible configurations of information and associated information-processing procedures in the memories of individual organisms. Thus, organism O's having formed an objective concept C is the same as O's having computed for itself a configuration of information and associated procedures that is an instance of the abstraction C. Needless to say, details of the structures and contents of concepts, both objective and subjective, are very much in need of investigation. This viewpoint is somewhat more psychologicistic than anything that either Frege or even McCarthy has seemed willing to embrace. Nevertheless, we believe that this framework is entirely appropriate as a basis for work in artificial intelligence and information-processing psychology, and expect it to be a fruitful source of heuristic insights for the development of technical details and the resolution of difficulties.

Since we take propositions to be composed of concepts in a structured way (and also to be themselves concepts of truth values), our representation scheme reflects this philosophical perspective: it is only through the mediation of concepts that intentional attitudes bring those holding them into relation with the (frequently non-conceptual) situations and objects with which those attitudes are concerned.

2.2 The Representation Language

With Frege, we distinguish the *sense* (intensional meaning, connotation) of an expression from its *denotation* (extension). Thus, the expressions "7+5" and "4x3" differ in sense but have the same denotation, since they both denote the number 12. In general, we take the sense of an expression to be a concept of appropriate type,* and in particular we follow Frege in taking the sense of a sentence to be the proposition (for him, *Gedanke*) that the sentence is used to express. For example, consider the formula:

1. **wants(mike, Meet(Mike, Wife Jim))**

[i.e., mike wants to meet Jim's wife *as such*.]

Here, the subformula of the form "**Meet(—)**" is a complex name of the proposition that mike meets Jim's wife, compounded of simple names of the relation-concept Meet, the function-concept Wife, and individual concepts of mike and Jim. The proposition named is the sense of the formula "**meet(mike, wife jim)**", while the denotation of this latter formula is its truth value. In general, capitalized names denote concepts of the things denoted by the corresponding uncapitalized names. (In order to avoid confusion in the present discussion, we shall follow this rule even in the English sentences commenting on our formalism.)

It is well known that linguistic constructions describing propositional attitudes often create contexts within which substitution of equals for equals, a standard rule of inference for identity, appears to break down. Thus, the sentences

2. mike wants to meet Jim's wife.

3. Jim's wife is Sally's mother.

might both be true, yet

4. mike wants to meet Sally's mother.

might under the same interpretation be false because of Mike's unawareness that Jim's wife is a mother. However, on a Fregean analysis, such apparent failures of extensionality turn out to be illusory; the rule of substitution in question simply does not apply in such cases, and the inferences it seemed to sanction turn out to be fallacies of equivocation. Thus, in our Fregean formalism we have (in addition to 1) the following formulas:

5. **wife jim = mother sally**

[i.e., Jim's wife is Sally's mother.]

6. **wants(mike, Meet(Mike, Mother Sally))**

[i.e., mike wants to meet Sally's mother *as such*.]

Note that 6 does not follow from 1 and 5, even though the contexts involved are all extensional. The reason is that "**Wife Jim**" denotes, not the person denoted by "**wife jim**", but only a concept of her, so there is no way here to substitute equals for equals. This is as it should be, given our initial interpretation of the sentences 2 - 4. However, it is worth noting that there is a second interpretation of 4 under which it *can* reasonably be inferred from 2, 3, and appropriate background information, though the logical structure of the inference may not immediately be clear. This inference is explicated in our present formalism as follows: From 1, 5, and the presumably true

7. **conceptof(Wife Jim, wife jim)**

[i.e., Wife Jim is a concept of Jim's wife.]

we may easily deduce the conclusion

8. **∃P.wants(mike, Meet(Mike, P)) ∧**

conceptof(P, mother sally)

[i.e., mike wants to meet a specific person who just happens to be Sally's mother.]

In three steps by conjunction, substitution of equals for equals, and existential generalization. This conclusion is our formalization of the second interpretation of 4; among other things, it shows the way in which our notation permits quantification into intentional contexts, and illustrates the value of this kind of "quantifying in" as a representational technique.*

So far, our departures from [13] are relatively minor; the most significant differences arise only in connection with the representation of iterated propositional attitudes, a matter to which we now turn. As indications of the way in which our notation handles this phenomenon, we offer three quite different interpretations of the ambiguous sentence:

9. **pat believes that mike wants to meet Jim's wife.**

* One can imagine circumstances in which the formula 8 would be clearly true, but in which the truth of the indicated English interpretation of it might nevertheless be questionable. The issue involved is whether the truth of 4 (on the second interpretation) requires that Mike be represented as having a concept of a *certain kind* of the person he wants to meet (for discussion of such issues, see [9]). However, the refinements of 8 that might thus be suggested would not take us outside the basic representational framework that we have indicated, and could best be developed in the context of an application of the framework to a specific class of reasoning problems.

* Our employment of the term "concept" in this context conforms with a fairly standard modern usage, but it should be noted that Frege's use of "Begriff", the German equivalent, is quite different [Frege 1891, 1892a]. An excellent overview of Frege's semantics is provided by the editor's introduction to [8].

(In the formulas to follow, "Mike\$" denotes a concept of Mike, while Mike is in turn a concept of the person mike. Also, Mike\$ is the *sense* of the concept-name "Mike", just as Mike is the *sense* of the personal name "mike". Analogous comments apply in the case of Meet\$, Wife\$, etc. The symbol "Exist" is an operator that binds concept-variables to form names of propositions.)

- 9.1 **believes(pat, Wants(Mike, Meet\$(Mike\$, Wife\$ Jim\$)))**
[here, pat is represented as attributing to mike a desire to meet Jim's wife *as such*.]
- 9.2 **believes(pat, Exist P\$.Wants(Mike, Meet\$(Mike\$, P\$)) And Conceptof(P\$, Wife Jim))**
[here, pat is represented as believing that mike wants to meet a specific person (e.g., his daughter's piano teacher) who just happens to be Jim's wife.]
- 9.3 **3P\$ P.believes(pat, Wants(Mike, Meet\$(Mike\$, P\$))) A conceptof(P\$, P) A conceptof(P, wife jim)**
[here, it is represented that the specific person whom pat believes that mike wants to meet just happens to be Jim's wife.]

Still other readings of 9 are expressible in our notation, but we omit them here for the sake of brevity. The reader should note the relationship between the formulas 9.1 and 1, and between 9.2 and the result of substituting "wife jim" for "mother sally" in 8.

2.3 What Is New to AI Here, and Why

The three readings just formalized provide a fairly stringent test of a notation's expressive power; the representational scheme of [13], for example, is incapable of capturing all three. Since the contributions of our notation to representation theory lie mainly in certain improvements to McCarthy's scheme that are responsible for the expressive power just illustrated in 9.1 - 9.3, we shall briefly outline the two main improvements, and indicate why they are needed.*

2.3.1 A Hierarchy of Concepts

The addition of a notational system for a potentially infinite, multibranched hierarchy of concepts is our major improvement; it is needed to enable McCarthy's language to adequately represent information concerning iterated propositional attitudes. The language of [13] provides for propositions and for concepts of individuals, but not for concepts of propositions or for concepts of concepts of individuals. However, it turns out that concepts of the latter two sorts are in fact required for the adequate treatment of constructions containing propositional attitudes nested to a depth of 2, as illustrated above in 9.1 - 9.3. Moreover, the obvious inductive generalization holds for propositional attitudes nested to depths of 3, 4, etc. Though the number of levels of concepts required is thus in principle infinite, in the vast majority of applications

only a few of the lowest levels of the hierarchy will actually be used. The nature of our notational hierarchy is illustrated by the following initial sequence of formulas:

- 10a. **teacher wife jim** (a sentence)
10b. **Teacher Wife Jim** (a name of a proposition)
10c. **Teacher\$ Wife\$ Jim\$** (a name of a concept)
10d. **Teacher\$3 Wife\$3 Jim\$3** (")
10e. **Teacher\$4 Wife\$4 Jim\$4** (")

Our language, like McCarthy's, is embedded in a multi-sorted first-order logic with functions and identity. However, because of the hierarchy of concepts, we make a more extensive use of sorts than McCarthy does. Thus, we have different sorts not only for individuals, truth values, concepts of individuals, and propositions (as McCarthy does), but also for concepts of concepts of individuals, concepts of propositions, and so on up the hierarchy. This aspect of our language has much in common with Alonzo Church's logic of sense and denotation [Church 1951, 1973, 1974], though in other respects there are major differences from Church's development. Church's logic is presently one of the most fully investigated formal treatments of concepts and propositions in the literature, and his papers are a valuable source of insights and technical ideas for those concerned with the application or development of Fregean intensional logic.

One aspect of our language development that has not yet received detailed attention is the matter of avoiding antinomies, especially semantic ones. The main reason for this is our desire to maintain flexibility in the formulation of axioms, rules of inference, and other features, as our language is developed through a series of applications to specific reasoning problems in a computational setting. Unlike formal developments in pure logic, a language designed for use in artificial intelligence must be convenient to apply and lend itself to efficient computer implementation. Yet, it is not possible to anticipate *a priori* what the optimal language organization will be for a given class of reasoning problems. When the lessons of developmental application have been learned, studies can then be undertaken to determine what technical means of avoiding antinomies is most appropriate.

2.3.2 Semantically Simpler Concept Names

In order to provide for names denoting quantified propositions and "definite-description" concepts of individuals, our language must contain some special-purpose name-forming machinery. McCarthy ([1979], pp. 13-16) has introduced such machinery, but in order to avoid any extension of the underlying logic, he has resorted to a notational system whose semantics is rather obscure and cumbersome, requiring that both possible worlds and "propositions with inner variables in them" be included in the domain of discourse. We believe that the needed conceptual names can be supplied more simply and naturally through a small expansion of the underlying logic that is completely extensional, preserves its essentially "first-order" character, and is semantically and syntactically transparent. What we have added for this purpose are three variable-binding, term-forming operators "All", "Exist", and "The", together with associated axioms. The operators "All" and "Exist" form names of quantified propositions and their higher-level relatives in the conceptual hierarchy, as illustrated above for "Exist" in the formula 9.2. The operator "The" forms names of "definite-description" concepts of individuals and their higher-level relatives. Thus, "The P.Childof(P, Mike)"

* A paper giving a much more complete and detailed account of our improvements to McCarthy's representational scheme is currently in preparation. It will include precise statements of the syntax and semantics of the notations introduced, and formulations of some associated axioms. It will also include discussion of some improvements not mentioned here.

denotes the concept expressed by the phrase "the child of mike".

3 Reasoning about Propositional Attitudes

Our approach to reasoning about propositional attitudes has its source in a familiar common-sense heuristic that people often use when thinking about the mental states of others: they ask themselves "What would I believe (or infer, or feel) in that person's situation if I were psychologically very much like that person?" The idea behind this heuristic plays a central role in our current efforts to construct an effective and natural computational method for reasoning about knowledge, belief, desire, and the like. In this section we shall indicate the lines along which our approach is being developed, discuss its relation to what has been called the "data base approach," and explain the sense in which it is *simulative*.

3.1 Characterization of Our Approach

In accordance with the second desideratum noted earlier, our approach to reasoning is knowledge-based, with some of the required knowledge being represented as formulas, some as production rules, and some as LISP programs. The approach has these salient features:

a. A program's reasoning about the propositional attitudes of other organisms is facilitated by using the program's own "mental machinery" to "simulate" the thinking of these organisms. This avoids any need for redundant re-representation of information about propositions at the second and higher levels in the hierarchy of concepts, and, for example, permits logical relations among beliefs to be determined directly by the program's ordinary inferential apparatus.

b. We require an access/maintenance system for an associatively indexed data base of logical formulas that has a CONNIVER-like context mechanism [15], and the ability to associate with each formula an indicator of its epistemic status, together with the rules and other formulas on which this status is based. The context mechanism is to be used both as an efficient way to keep track, for specific inferential purposes, of what is (or is not) believed by a given person, and as an aid in presumptively including common knowledge and opinion in the corpus of each person's beliefs.

c. As a supplement to deductive rules of the usual sort, we make use of "inference by evaluation," a procedure whereby a sub-expression of the form "f(a1, a2, ..., an)" is replaced by the result of evaluating it as a LISP expression.* Thus, "believes(mike, Uncle(Jim, Sally))" may be an immediate consequence of "believes(mike, sense "uncle(jim, sally)")". This use of content-dependent inference rules can lead to substantial efficiencies in the reasoning process.

To illustrate the intended style of reasoning, let us consider how a system S would use it to determine whether or not mike believes that jim is sally's uncle. Interest is thus focussed on the formulas

11. believes(mike, Uncle(Jim, Sally)), and

12. believes(mike, Not Uncle(Jim, Sally)).

Having determined (we suppose) that neither 11 nor 12 nor their negations are either explicitly stored in or trivially

inferred from its knowledge corpus, S then "pretends" that it is mike, and tries to figure out what its state of belief is concerning the proposition that jim is sally's uncle.* It does this as follows: S computes from 11 the sentential formula

11.b uncle(jim, sally),

and then constructs two contexts: C_{bm} , containing (sentential) formulas expressing the propositions that it initially thinks are believed by mike and relevant to 11.b, and $C_{\neg bm}$, containing formulas expressing the propositions that it initially thinks are not believed by mike and relevant to 11.b. In constructing these contexts, S is guided not only by specific knowledge about mike's beliefs, but also by general principles concerning what people commonly do and do not know and believe, and by an explicitly represented corpus of such common beliefs. Wherever needed, sentential formulas are computed directly from the names of the propositions they express, as in the case of 11.b.

At this point S sets up two tasks, T1 and T2. In T1, 11.b is added to mike's belief-context C_{bm} and the resulting premise-set is taken as the starting point for an attempt to infer either: i) a contradiction (in which case it might be concluded that mike believes that jim is not sally's uncle), or ii) members of mike's non-belief context $C_{\neg bm}$ (in which case it might be concluded merely that mike does not believe that jim is sally's uncle). In the companion task T2, the negation of 11.b is added to C_{bm} and a similar attempt is made to infer either: i) a contradiction (in which case it might be concluded that mike believes that jim is sally's uncle), or ii) members of mike's non-belief context $C_{\neg bm}$ (in which case it might be concluded merely that mike does not believe that jim is not sally's uncle). Measures of the resources expended by the inference process are computed for use in conjunction with knowledge about the level of mike's reasoning abilities; the indicated conclusions about mike's belief-state will actually be drawn if mike can reasonably be expected to see the logical relationships involved. At certain points in the reasoning, S may add a formula to $C_{\neg bm}$ after determining that there is no reason (for mike) to believe the proposition expressed by the formula.

The tasks T1 and T2 are to be pursued in parallel until either: i) a complete characterization of mike's state of belief with respect to the proposition in question has been inferred, or ii) the resources allotted to the tasks are exhausted. In many cases, a good initial heuristic for S will be to try to attribute some minor variant of S's own state of belief concerning the proposition in question (and the reasons for it) to mike. The combinatorics inherent in these tasks are to be controlled by using a knowledge-based theorem prover of the general sort envisioned in [19], and by keeping the contexts involved fairly small, allowing them to grow larger only when the theorem prover indicates a need for it, and then only by judicious choice of relevant propositions. The overall *simulative* process must be callable recursively, in order to permit reasoning about iterated propositional attitudes.

* Essentially this mode of inference has been implemented in the "semantic attachment" feature of the FOL proof checker; see [20].

* We assume here that an organism believes that P if and only if it either has P explicitly stored as a belief that is readily accessible, or can easily infer P from explicitly stored beliefs by a relatively routine inference procedure.

3.2 Relation to the "Data Base Approach"

Our use of contexts, illustrated above, to represent subsets of beliefs and non-beliefs for inferential purposes is related to what Robert Moore has called the "data base approach" to representation and reasoning about knowledge, which he characterizes as follows:

[An] ... idea which initially seems very appealing is to use the multiple data-base capabilities of advanced AI languages to set up a separate data base for each person whose knowledge we have some information about. We then can record what we know about his knowledge in that data base, and simulate his reasoning by running our standard inference routines in that data base. This idea seems to have wide currency in AI circles, and I advocated it myself in an earlier paper [16]. Unfortunately, it doesn't work very well. ([17], p. 224.)

Moore goes on to point out some difficulties with the "data base approach" that he evidently regards as casting doubt on its viability. We shall discuss this matter here, in order both to clarify the degree of our own involvement with the data base approach, and to rescue the idea behind its second part from any presumption of guilt by association with the quite different idea behind its first part.

Within the data base approach as characterized by Moore, sets of formulas (data bases) are assigned two distinct functions: they are to serve both as a general means of representing facts about knowledge, and also as premise sets for *simulative* inferential processes. However, it appears that the difficulties raised by Moore for this approach cause serious trouble only in relation to the first function. The two main difficulties concern the representation, using multiple data bases, of logical compounds (disjunctions and negations) of knowledge attributions [17], p. 224; [18], sec. 2.2). These problems do not arise within our own approach in connection with the first of the two functions mentioned above (that of a general representation medium), since we use the Fregean formalism described in Section 2, *not* multiple data bases, to perform that function. We do make use of what are in effect small temporary data bases to perform the second of the two functions mentioned (that of premise sets), in the way illustrated in Section 3.1. For this use of "data bases," however, the problems raised by Moore appear to be quite manageable.

One such problem concerns the use of separate data bases to represent what a given person does and does not believe. Moore (*ibid.*) implies, without further explanation, that the problem of utilizing the information in a person's belief and non-belief data bases in a single *simulative* reasoning process is not easily solved. The illustration in Section 3.1 is directly relevant to this problem, and shows that a solution lies in running the *simulative* theorem prover in the belief data base (context), while using the non-belief data base in an appropriate way as a source of goal statements for the theorem prover.

A second problem concerns the combinatorially large number of possibilities for mike's knowledge state compatible with a description of it containing multiple disjunctions of knowledge attributions. The resulting uncertainty can require that many different knowledge contexts be constructed and investigated in order to draw useful conclusions about what mike has inferred from his knowledge. However, this difficulty is not peculiar to our approach to reasoning, but rather is inherent in the given

type of reasoning problem itself. Our approach helps to mitigate such problems by using specific knowledge of the inference task to minimize both the number and size of the contexts constructed.

3.3 Indirect Simulation

While the type of reasoning about belief illustrated in Section 3.1 may properly be called "*simulative*," it is not *directly* *simulative*. That is, conclusions about mike's beliefs are not drawn simply by running a supposedly accurate simulator of mike and directly reading off the conclusion. Rather, an admittedly approximate simulator of mike is run, and then conclusions about his beliefs are *inferred* from from observed characteristics of the simulation, together with knowledge of the simulator and of mike. In this way the raw results of the simulation can be corrected on the basis of known differences between the simulator and mike. This sort of *indirect simulation* (as we shall call it) has the advantage that a single simulator can be used to learn about the beliefs of many different organisms. It has the obvious limitation that great differences between the processes of the simulator and those of the organism simulated can make it difficult to draw accurate conclusions.

4 Comparison with Some Alternative Approaches

As indicated earlier, our Fregean approach to the representation of propositional attitudes is closely related to that of McCarthy [1977, 1979]. With the improvements to McCarthy's scheme discussed above in Section 2, our representational method goes a long way toward satisfying the first desideratum of Section 1, though much more experience with it is needed to gain a clear idea of its limitations. Unlike straightforward modal representations of propositional attitudes, a Fregean scheme permits the use of a fully extensional logic, and allows quantification over concepts and propositions, as in 9.3 and in the formalization of "All of mike's beliefs are well-founded."

When it comes to reasoning about propositional attitudes, our approach is quite different from the one currently taken by McCarthy, which relies heavily on principles relating statements about propositional attitudes to the truth values of the propositional objects of the attitudes in various possible worlds. While this "possible worlds" approach provides a powerful tool for systematizing reasoning about knowledge [McCarthy 1975; Moore 1977, 1979], in its simplest form it is committed to some quite strong idealizations that would preclude its applicability in many practical situations, and that violate the second desideratum adopted above in Section 1. For example, the following principle plays a fundamental role in this approach:

PW1: If a proposition P is true in every possible world compatible with what a given person knows, then that person knows P.

This entails (most implausibly) that a person knows all the logical consequences of his knowledge. Moore [1977, 1979] acknowledges this difficulty in principle, and proposes to solve it by regarding the conclusions sanctioned by PW1 as only presumptive in nature, to be overridden by better supported information to the contrary. However, his technical developments thus far take no account of this caveat, and instead assume PW1 in its simple, unqualified form.

A similar assumption can serve to simplify our *simulative* approach to reasoning. In its crudest form, the assumption

is that an intelligent program's own reasoning mechanism, used to investigate the beliefs of another organism, is sufficiently accurate as a simulation that it requires no correction. In the terminology of Section 3.3, this amounts to assuming that such a program can use a very convenient form of *direct* simulation to investigate the beliefs of others. This simplified simulative approach would be on a par with existing possible-worlds reasoning methods with respect to our second desideratum. However, we suspect that a simulative system could be more readily and naturally extended to take realistic account of the reasoning abilities of other organisms than could a possible-worlds system incorporating the principle PW1. The reason is that, since PW1 makes no direct reference to logical implication or inference, the most efficient revision might be to replace PW1 outright rather than to reinterpret it as a merely presumptive rule. Of course, the correctness of our suspicion can be decided only when extensions of the sort it envisions have been seriously attempted.

5 Further Research

Development of our approach is presently proceeding on three fronts:

1. Epistemological investigation to determine what knowledge and inferential procedures are appropriate for the solution of particular reasoning problems within the framework we have established.

2. Investigation of alternative designs for an initial implementation of our ideas in which all reasoning is deductive.

3. Studies leading to the design of a computational framework that integrates deductive and non-deductive reasoning, and deals in a principled way with degrees of belief, and uncertainty of evidence and inference.

An intermediate goal of our research is the construction of a problem solver of the "advice taker" type [McCarthy 1959; McCarthy & Hayes 1969], to serve as a tool for the detailed study of knowledge, beliefs, goals, and normative precepts, as they function in planning and decision-making.

Acknowledgments

At various stages in the work reported here, I have benefitted from interactions with Robert Filman, Reed Letsinger, John McCarthy, Robert Moore, Richard Weyhrauch, and an anonymous referee, but of course, they are not responsible for the result.

References

- [1] Church, A. "A Formulation of the Logic of Sense and Denotation." *Structure Method and Meaning: Essays in Honor of Henry M. Sheffer*. Edited by P. Henle et al. New York: Liberal Arts Press, 1951.
- [2] Church, A. "Outline of a Revised Formulation of the Logic of Sense and Denotation (Part I)." *Nous*. 7 (1973) 24-33.
- [3] Church, A. "Outline of a Revised Formulation of the Logic of Sense and Denotation (Part II)." *Nous*. 8 (1974) 135-156.
- [4] Frege, G. "Funktion und Begriff." Address given to the *Jenaische Gesellschaft für Medizin und Naturwissenschaft*. January 9, 1891. Translated by P. T. Geach under the title "Function and Concept." [7].
- [5] Frege, G. "Über Begriff und Gegenstand." *Vierteljahrsschrift für wissenschaftliche Philosophie*. 16 (1892) 192-205. Translated by P. T. Geach under the title "On Concept and Object." [7].
- [6] Frege, G. "Über Sinn und Bedeutung." *Zeitschrift für Philosophie und philosophische Kritik*. 100 (1892) 25-50. Translated by H. Feigl under the title "On Sense and Nominatum." *Readings in Philosophical Analysis*. Edited by H. Feigl and W. Sellars. New York: Appleton-Century-Crofts, 1949. Also translated by M. Black under the title "On Sense and Reference." [7].
- [7] Frege, G. *Translations from the Philosophical Writings of Gottlob Frege*. Edited by P. Geach and M. Black. Oxford: Basil Blackwell, 1952.
- [8] Frege, G. *The Basic Laws of Arithmetic. Exposition of the System*. Translated and edited by M. Furth. Berkeley and Los Angeles: Univ. of California Press, 1967.
- [9] Kaplan, D. "Quantifying In." *Words and Objections: Essays on the Work of W. V. Quine*. Edited by D. Davidson and J. Hintikka. Dordrecht, Holland: D. Reidel Publishing Co., 1969. Reprinted in *Reference and Modality*. Edited by L. Linsky. London: Oxford University Press, 1971.
- [10] McCarthy, J. "Programs with Common Sense." *Proceedings of the Symposium on Mechanisation of Thought Processes*. Teddington, England: National Physical Laboratory, 1959. Reprinted in *Semantic Information Processing*. Edited by M. Minsky. Cambridge, Mass.: The MIT Press, 1968.
- [11] McCarthy, J. "An Axiomatization of Knowledge and the Example of the Wise Man Puzzle." Unpublished memorandum. 1975.
- [12] McCarthy, J. "Epistemological Problems of Artificial Intelligence." *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*. Cambridge, Mass.: 1977. 1038-1044.
- [13] McCarthy, J. "First Order Theories of Individual Concepts and Propositions." Artificial Intelligence Laboratory Memo AIM-325. Stanford University, March 1979. Also forthcoming in *Machine Intelligence 9*. Edited by D. Michie. 1979.
- [14] McCarthy, J., and Hayes, P. "Some Philosophical Problems from the Standpoint of Artificial Intelligence." Artificial Intelligence Laboratory Memo AIM-73. Stanford University, November 1968. Also in *Machine Intelligence 4*. Edited by D. Michie. New York: American Elsevier, 1969.
- [15] McDermott, D., and Sussman, G. *The Conniver Reference Manual*. AI Lab Memo 269. Massachusetts Institute of Technology, May 1972.
- [16] Moore, R. C. "D-SCRIPT: A Computational Theory of Descriptions." *Advance Papers of the Third International Joint Conference on Artificial Intelligence*. Stanford, Calif., 1973. 223-229. Also in *IEEE Transactions on Computers*. C-25:4 (1976) 366-373.
- [17] Moore, R. C. "Reasoning About Knowledge and Action." *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*. Cambridge, Mass., 1977. 223-227.
- [18] Moore, R. C. *Reasoning About Knowledge and Action*. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, (forthcoming).
- [19] Nilsson, N. *A Production System for Automatic Deduction*. Computer Science Department Report No. STAN-CS-77-618. Stanford University, July 1977.
- [20] Weyhrauch, R. *A Users Manual for FOL*. Artificial Intelligence Laboratory Memo AIM-235.1. Stanford University, July 1977.