# Chapter 2: Knowledge Representation

## Defining the Knowledge Representation Enterprise

The standard reference for a "simple" characterization of knowledge representation is the "Knowledge Representation Hypothesis" given by Brian Smith:

> Any mechanically embodied intelligent process will be comprised of structural ingredients that
> a) we as external observers naturally take to represent a propositional account of the knowledge that the overall process exhibits, and
> b) independent of such external semantical attribution, play a formal but causal and essential role in engendering the behavior that manifests that knowledge.[1]

This hypothesis is used in defining knowledge representation (KR) by both [Barr&Feigenbaum 1981] and [Brachman&Levesque 1985]. The hypothesis as stated by Smith is verbose[2] and can be restated in a shorter form (already presented in the previous chapter):

> Any artificial intelligent process consists of structural ingredients that: represent a propositional account of the knowledge that the overall process exhibits; and, play a formal, causal, and essential role in its behavior.

This hypothesis is meaningful in terms of a particular model of intelligent agents. A version of this model of intelligent agents is given by Jon Barwise in the course of a multi-paper debate with Jerry Fodor as some common ground between them. He states that an agent can be characterized:

> (1) as having receptors, to pick up information about the agent's environment;
> (2) as having a "computational domain" consisting of, or containing, representational "mental states";
> (3) as having a "processor" which is the locus of change in the computational field;
> (4) as having effectors, to act in and on the environment.[3]

---

[1] p. 33 in [Smith 1982].

[2] "...the behavior that manifests that knowledge." This use of manifests is odd. If the behavior manifests the knowledge, then the knowledge must be essential (or causal, or both) to that behavior, otherwise the behavior couldn't be considered to "manifest" it. Thus, this use of "manifests" is superfluous.

[3] p.161 in [Barwise 1988].

He then identifies the “Strong Modularity” view of intelligent agents as one where the information processor (of point (3) in the characterization of an intelligent agent) is divided into three parts:

> “...one ‘transducer’ responsible for getting expressions into the field, another transducer responsible for translating expressions in the field into action, and then a subagent insulated from the environment that is responsible for changes in the field.”[4]

Barwise objects to this characterization of the processor since it ignores the possibility of the situation in which the processor is embedded as having a direct effect on the behavior of the processor. He points out Strong Modularity does fit well with the “methodological solipsism”[5] of most of cognitive science and AI, however.

This question about Strong Modularity aside, this model of intelligent agents is consonant with the KRH of Smith. Essentially, the model presented by Barwise provides some of the details of the “structural ingredients” mentioned by Smith.

A central concept here is the *decision* an agent makes. What is a decision? A decision is an *action*. It is not a physical action, perhaps. A decision is *not* an inference, although it’s based on the results of inference (at least some times). As a result of making a decision, an agent acquires an *intention*. Any agent’s action is representable as a decision, followed by an effector (in the earlier model of a intelligent agent) noticing the resulting intention in the computational field and taking that intention as a command. There could be many effectors continuously checking for enabling intentions. Thus, the basic step in making an action happen is the primitive action of a decision.

A related problem is what is an *inference*. An inference is also an action. By the fore-

---

[4] pp. 168-169 in [Barwise 1988].

[5] Solipsism is defined as “a theory holding that the self can know nothing but its own modifications and that the self is the only existent thing”, p.1106 in [Woolf et. al. 1977]. Jerry Fodor has an essay on the desirability of using methodological solipsism in cognitive psychology, “Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology”[Fodor 1980]. He presents the idea that the “formality” condition “...is tantamount to a sort of methodological solipsism. If mental processes are formal, they have access only to the formal properties of such representations of the environment as the senses provide. Hence, they have no access to the *semantic properties* of such representations, including the property of being true, of having referents, or, indeed, the property of being representations *of the environment*.” (p. 314 in [Fodor 1980]).

going, an inference can be analyzed as a decision followed by one or more "inference-type" effector's response to that decision. This is a more fine-grained analysis than is used in this thesis.

One shortcoming of this model is that it provides no element which explicitly corresponds to "memory". One must fit memory into the model as a part of the computational field. This implies that memories are acquired either by the "input" transducer (i.e. perceptions being placed directly in memory) or by the reasoning "subagent" adding information (inferred or copied from information in either the memory or non-memory part of the field) to the memory part of the computational field. It also implies that memories are forgotten *only* by the reasoning subagent modifying the memory part of the field - that forgetting is somehow a result of reasoning (by the subagent).

An alternative to the above handling of memory is to propose an extension to the basic model, a fifth element of an intelligent agent - its memory. This memory can be viewed as a special part of the environment of the agent (in the terms of the model), so that the agent has environment effectors (in element 4 of the model) which act on the agents memory (adding memories, requesting retrieval of a memory), and the agent has memory "receptors" (in element 1 of the model) which pick up memories (at least as requested by the memory retrieval effector, perhaps "spontaneously" as well - i.e. for reasons separate from the intentions or conscious requests of the agent).[6]

This latter approach to modeling memory seems more in line with the nature of intelligent agents, and does not preclude the first approach (there could be both kinds of "memory"). The latter approach allows for arbitrary models of memory. Such a model might well include some kinds of reasoning/inference in the memory, entirely distinct from the operation of the computation field and its processor. If this approach is followed far enough, the computational field and its processor become only

[6] This approach could be made consonant with the memory model used in [Elgot-Drapkin et. al. 1987]. They propose a memory model of several parts, long term memory, short term memory, relevance memory, queue waiting memory, and history memory. The short term, relevance, and history memories, together, might be considered as the computational field.

minor elements in an intelligent agent, with nearly all of the interesting activity in the operation of the memory. The knowledge representation hypothesis insists that this is not the case, the computational field and its processor must be central to and account for most of the behavior of an intelligent agent, with the *processing* of memory as distinct from the processing of the computational field being a minor element. This thesis, in assuming the KRH, takes this latter position.

By the above discussion, the form of the contents of the computational field, the relationship between a particular situation of a particular agent and the contents of the computational field (a particular "domain"), and the nature of the processor (rules and manner of operation) which operates on the computational field, are central problems in modeling an intelligent agent. Solving these problems is broadly the knowledge representation enterprise.

This thesis concentrates on the part of the enterprise having to do with the form of the contents of the computational field and the rules and manner of operation of the processor. Particular problem domains are only investigated as exemplars of particular issues, no attempt is made at an analysis of a "real" problem domain. The work in this thesis does not shed any light on the computational implications of "situated inference". For this thesis, the Strong Modularity principle is assumed. A future improvement on this work is to pursue situated inference.

## Approaches to Knowledge Representation

There are many aspects to specifying the form of the contents of the computational field. There are several types of information which can be present, and which the form must allow. Barr and Feigenbaum[7] describe these types as: objects, events, performance and meta-knowledge. They identify three classes of uses of knowledge - acquisition, retrieval, reasoning. There are several divisions of the reasoning class:

formal reasoning,
procedural reasoning,
reasoning by analogy,
generalization and abstraction, and
meta-level reasoning.

They identify several conflicting attributes of knowledge representation languages:

scope and granularity,
indeterminacy and semantic primitives,
modularity and understandability,
explicit knowledge and flexibility, and
declarative vs procedural representations.

They provide the following categorization of existing KR languages:

logic,
procedural representations,
semantic networks,
production systems,
direct (analogical) representations,
semantic primitives, and
frames and scripts.

Finally, they identify certain areas that are known to be hard problems:

quantification, the ability to specify properties of arbitrarily defined sets,
representing belief's (which may or may not be true),
degrees of certainty,
mass nouns,
time and tense information, and

---

[7] [Barr&Feigenbaum 1981]

processes that consist of sequenced actions taking place over time.

A KR language/system can be analyzed in terms of the issues provided by Brachman and Levesque[8] as follows:

- Overall adequacy
  - Expressive adequacy
  - Reasoning efficiency
- Basic epistemology
  - Primitives
  - Meta-Representation
  - Incompleteness
- Aspects of KR not covered by standard logic
  - Definitions vs. Facts
  - Universals vs. Defaults
  - Non-Deductive Reasoning
  - Nonmonotonic Reasoning
  - Non-propositional (yes-no) representations
  - Procedural
  - Analogical
  - Probabilistic
- Domain of the KR
  - Substances
  - Causality and Time
  - Knowledge

It's interesting to compare and contrast these two approaches to characterizing KR and KR systems. In the following discussion Barr and Feigenbaum's work is referred to as BF and Brachman and Levesque's work is referred to as BL. The BL knowledge subitem of the domain item encompasses three of the BF types of knowledge, objects, events, performance, and the BF representing beliefs problem. The BL substance subitem of the domain item is equivalent to the BF mass nouns problem. The BL causality-and-time subitem of the domain item is equivalent to the BF time-and-tense information problem. The BL meta-representation subitem of the basic episte-

[8] [Brachman&Levesque 1985]

mology item is equivalent to the BF meta-knowledge knowledge type together with the meta-level kind of reasoning.

The BL “aspects...” item implies the logic KRL category. The BL procedural subitem of the “aspects...” item is equivalent to the BF procedural representations KRL category and the procedural reasoning type. The BL analogical subitem of the “aspects...” item is equivalent to the BF analogical (direct) representations KRL category and the reasoning by analogy reasoning type.

The BL expressive adequacy is related to BF modularity and understandability. The BL probabilistic and nonmonotonic reasoning subitems of the aspects-not-covered-by-standard-logic item and the incompleteness subitem of the basic epistemology item are related to the BF indeterminacy part of the indeterminacy-and-semantic-primitives dichotomy. The BL primitives subitem of the basic epistemology item is equivalent to the BF semantic primitives part of the indeterminacy-and-semantic-primitives dichotomy and the BF semantic primitives language categorization. The BF degrees of certainty problem may be related to the BL non-propositional (yes-no) representations subitem of the aspects-not-covered-by-standard-logic item. It may also be part of the BL probabilistic, nonmonotonic reasoning, or incompleteness subitems.

This rough comparison leaves several points in each approach without a corresponding point in the other approach. The items unique to BF are: the generalization and abstraction reasoning class, the scope and granularity and explicit knowledge and flexibility KRL dichotomies, the semantic networks, production systems and frames and scripts KR language types, the quantification and the processes problems. The items unique to BL are: reasoning efficiency of the overall adequacy item, the definitions vs. facts, universals vs. defaults, and non-deductive reasoning of the aspects-of-KR-not-covered-by-standard-logic item.

## Situated Approaches to Knowledge Representation

It is possible to consider a situated approach to any of the KR issues mentioned above, or a situated version of any of the KRs. The approach investigated in this thesis is to develop a KRL which is of the “logic representation” kind in BF analysis. Of the BF knowledge types, this thesis discusses three types, objects, events, and meta-knowledge, leaving out performance. Of the BF knowledge uses, this thesis investigates only the reasoning use, and has nothing to say about acquisition or retrieval. The kind of reasoning investigated in this thesis, in BF terms, is formal reasoning. The other BF kinds of reasoning (procedural reasoning, reasoning by analogy, generalization and abstraction, and meta-level reasoning) could be addressed within the ST KRL herein developed. In the BF dichotomies, the approach of this thesis emphasizes granularity instead of scope, understandability instead of modularity, declarative instead of procedural representations. The ST KRL is indeterminate with respect to semantic primitives. The other dichotomy is unclearly applicable to the ST KRL (explicit knowledge and flexibility). Of the KR problems BF list, the ST KRL addresses representing beliefs (which may or may not be true), and time and tense information. There are some interesting works in ST which may present a useful handling of mass nouns.[9] The “degrees of certainty” issue is not addressed directly by ST, or this thesis.

The ST KRL work of this thesis can also be analyzed with respect to the BL criteria. In terms of overall adequacy, only expressive adequacy is dealt with. Reasoning efficiency is ignored. In terms of the basic epistemology, only incompleteness is addressed. ST offers no explicit guidance on the choice of primitives. Full ST does deal with meta-representation, but this is *not* dealt with in this thesis. This thesis does not address any of the aspects of KR not covered by standard logic. In terms of the domain of the KR, this thesis deals only with knowledge. Time is handled in full ST, but is not explored in this thesis. Causality and substances (mass nouns) are not dealt with by ST.

[9] The perspectival approach presented by Jerry Seligman in [Seligman 1990] provides a definition of “object” which looks equally applicable to discrete objects (e.g. table, pear) and mass objects (e.g. water, air).

## Logic

The logic-based approach to KR is clearly the most similar to ST as it is commonly developed. This approach can be divided into monotonic versus nonmonotonic approaches. ST is not equivalent to FOL or any modal logic based on the classical logical tradition. The following discussion examines the relationship between ST and some standard logics.

### Monotonic logic

The most basic of logic-based approaches is simply First Order Classical Logic. Full ST is *not* a first-order logic. In its most extreme form, every statement in full ST can be used as an argument (an object, a term) to a relation. However, ST as developed in this thesis *is* a first order logic.

### Modal logic

Representation in logic can be made more convenient (simpler and readily computable) for KR purposes by using a *modal* first order logic. Generally, a modal first order logic is defined using a (first-order) possible worlds semantics. Alternatively, the Z modal logic is defined using a (nearly first-order) *intensional* semantics.

ST is *not* definable directly in terms of possible worlds semantics, so a ST-based logic for KR is not simply another use of modal logic, nor is there a direct semantic equivalent of ST in Z modal logic. That is, a *situation* is *not* some kind of possible world. It is not even a partial or small possible world.[10] A *real* possible world may be considered a kind of situation, in some versions of ST. However, ST and possible-

[10] The Z intensional semantics may be consonant with this attitude toward declarative sentences, perhaps by defining a SITUATION analogous to the definition of WORLD. However, the Z modal logic incorporates all of a traditional first order logic's axioms. In a direct representation of ST, this logic must serve as the "infon logic" analog. The "infon logic" of ST requires that the disjunction of an infon and its dual is *not* a tautology. But, this *is* a tautology in FOL and in Z. See [Brown 1987] for a presentation of Z. A more complex mapping into Z involves placing a necessitation operator ('[S]' where S is a conjunction of positive and negative literals (basic infons) which defines a situation) in front of every term and/or operator of the infon logic formulae. Thus, $S \models (A \vee \sim A)$ maps into $[S]([S]A \vee [S]\sim A)$. Gödel used an approach like this to model intuitionistic logic in an S4 modal logic, so such an approach might work for modeling much of infon logic/ST in Z modal logic (which is a kind of S5 logic).

world semantics agree that : *a declarative sentence can be used to make an assertion, that is, a claim that the actual world (or some portion thereof) is some particular way or other*.[11] One of the bases of Barwise's rejection of small worlds as situations is that situations need not be fully "closed" - if a situation "settles" property P for object *a* and property Q for object *b*, it need not settle property P for object *b* or property Q for object *a*.[12] In this work, situations *are* closed with respect to a given set of logical constraints, but this still allows for the indeterminacy of the foregoing as long as these constraints do not derive that P(*a*) or Q(*b*) implies (P(*b*) or not P(*b*)), and similarly for Q(*a*).[13]

Since "infon" logic for ST does not support the law of the excluded middle (i.e. for P an arbitrary infon, it is *not* necessarily that case that an arbitrary situation supports the disjunction of P and its dual (or "negation") ($P \vee \sim P$)), there is a relationship between the infon logic of ST and intuitionistic logic[14].

An additional problem for modal models of ST is the use of situations as terms or objects in infon formulae. A first-order infon may have a situation as the value of one or more of its arguments. An example where this is useful is in the model of perception discussed in Chapter 5 of this thesis. This defines the 'sees' relationship between and agent who is doing the seeing and the situation which is seen. The idea of possible worlds provides no useful analogy with situations in this context. A person certainly doesn't see an entire world (even if limited to a particular possibility). But, this is what a reference to a possible world in a sees relation would imply. In the Z modal logic approach, where a situation is modeled by a proposition which is a conjunction of positive and negative literals, including this representation of a situation as a term requires a higher order logic, since a variable which is quantified over terms may now be quantified over propositions (albeit of a special form). Also, since a situation in this propositional form is represented by the literals it supports and not by a name, recursive references are more difficult to reason about.

---

[11] [Barwise 1988], p. 79. This is part of a multi-paper discussion on possible world semantics and situation theory between Jon Barwise and John Perry on one side and Cresswell on the other.

[12] p. 85 in [Barwise 1988].

[13] This idea of closure is discussed in several ST papers. For instance, p. 47 in [Barwise&Etchemendy 1990].

[14] There is a sketch of modal *intuitionistic* logic presented in [Plotkin&Sirling 1986].

The *metatheory* for full ST can be given as a classical First Order Logic, as is done by [Westerståhl 1990]. Thus, the metatheory could have a possible-world or intensional semantics (since these both encompass FOL).

Barr&Feigenbaum 1981 Chapter III, “Knowledge Representation”, pp. 141-221 in *The Handbook of Artificial Intelligence*, Avron Barr & Edward A. Feigenbaum (eds.), Los Altos: William Kaufmann, 1981.

Barwise 1988 *The Situation in Logic* by Jon Barwise, Center for the Study of Language and Information: Stanford University, 1988.

Barwise&Etchemendy 1990 “Information, Infons, and Inference” by Jon Barwise and John Etchemendy, pp. 33-78, in *Situation Theory and Its Applications* by Robin Cooper, Kuniaki Mukai, and John Perry, eds., Stanford: Center for the Study of Language and Information, 1990.

Brachman&Levesque 1985 *Readings in Knowledge Representation* by Ronald J. Brachman and Hector Levesque (eds.), Los Altos: Morgan Kaufmann, 1985.

Brown 1987 “A Modal Logic for the Representation of Knowledge” by Frank M. Brown, pp. 135-156 in *The Frame Problem in Artificial Intelligence* edited by Frank M. Brown. Los Altos, CA: Morgan Kaufmann Publishers, 1987.

Elgot-Drapkin et. al. 1987 “Life on a Desert Island: Ongoing Work on Real-Time Reasoning” by Jennifer Elgot-Drapkin, Michael Miller, and Donald Perlis, pp. 349-357, in *The Frame Problem in Artificial Intelligence* edited by Frank M. Brown. Los Altos, CA: Morgan Kaufmann Publishers, 1987.

Fodor 1980 “Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology”, by Jerry A. Fodor, in *The Behavioral and Brain Sciences*, **3** (1980), 63-73. Reprinted on pp. 307-338 in *Mind Design: Philosophy, Psychology, Artificial Intelligence*, edited by John Haugeland (A Bradford Book). Cambridge, MA: The MIT Press, 1981.

Plotkin&Sirling 1986 “A Framework for Intuitionistic Modal Logics (Extended Abstract)” by Gordon Plotkin and Colin Sirling, pp. 399-406, in *Theoretical Aspects of Reasoning About Knowledge* edited by Joseph Y. Halpern, proceedings of the 1986 conference, March 19-22, 1986, Monterey, California. Los Altos, CA: Morgan Kaufmann Publishers, 1986.

Seligman 1990 “Perspectives in Situation Theory” by Jerry Seligman, pp. 147-192, in *Situation Theory and Its Applications* by Robin Cooper, Kuniaki Mukai, and John Perry, eds., Stanford: Center for the Study of Language and Information, 1990.

Smith 1982 “Prologue to ‘Reflection and Semantics in a Procedural Language’” by Brian Cantwell Smith, Technical Report MIT/LCS/TR-272 M.I.T., Cambridge, MA, 1982. Reprinted on pp. 32-40 in *Readings in Knowledge Representation* by Ronald J. Brachman and Hector Levesque (eds.), Los Altos: Morgan Kaufmann, 1985.

Westerståhl 1990 “Parametric Types and Propositions in First-Order Situation Theory” by Dag Westerståhl, pp. 193-230, in *Situation Theory and Its Applications* by Robin Cooper, Kuniaki Mukai, and John Perry, eds., Stanford: Center for the Study of Language and Information, 1990.

Woolf et. al. 1977 *Webster’s New Collegiate Dictionary*, Editor-in-Chief Henry Bosley Woolf, Springfield, MA: G.&C. Merriam Company, 1977.