On Representing Commonsense Knowledge*

Benjamin Kuipers

1979

Introduction

Commonsense knowledge is knowledge about the world that "everybody knows." Everybody knows that water is wet, that round things roll downhill, that to drive somewhere you have to get in the car first, that if you insult someone he might be unhappy or angry. Everybody knows how to get home from work, where Florida is on the map, that the French Revolution was much longer ago than the Chicago riots, why the sea is boiling hot, and whether pigs have wings (apologies to Lewis Carroll).

Much of this knowledge is never taught: people just seem to pick it up without any effort. Some is learned in school, like facts about the French Revolution, but assuming the commonsense notion of "longer ago than ...." Commonsense knowledge is frequently wrong, but almost always "good enough" for commonsense purposes. Some people lack common sense for practical subjects like electricity or motors, but most people have it for active threats like open manholes or raging fires.

How can we account for the power and usefulness of commonsense knowledge? More technically, how can we represent commonsense knowledge in a computer? In this chapter, I explore some of the characteristics of representations that make them suitable for commonsense knowledge. The chapter arose from an attempt to capture some of the general features of representations that I used in constructing a computer model of commonsense knowledge of large-scale space [Kuipers, 1977, 1978]. In Sections 1 and 2, I propose a definition for commonsense knowledge, list some of the performance constraints it must satisfy, and discuss what it means to represent

*In N. V. Findler (Ed.), *Associative Networks: The Representation and Use of Knowledge by Computers.* NY: Academic Press, 1979, pages 393–408.

knowledge on a computer. Section 3 presents a collection of design features that appear to be important for representations of commonsense knowledge. Sections 4 and 5 discuss two representations in more detail: the partial order, and a route description. They discuss the performance characteristics of these representations and show how they tie into the larger structure of the TOUR model of commonsense knowledge of large-scale space.

1 What Is Commonsense Knowledge?

It would be desirable to have a definition of "commonsense knowledge" that characterized this concept precisely. This is not yet possible, and may never be, but the following definition is an attempt to distinguish commonsense knowledge from basic cognitive skills such as manipulation, vision, and language on the one hand, and expert knowledge of electronics, mathematics, or history on the other. The abundant qualifiers in the definition are necessary because of the fuzziness of the phenomenon itself:

> *Commonsense knowledge is knowledge about the structure of the external world that is acquired and applied without concentrated effort by any normal human that allows him or her to meet the everyday demands of the physical, spatial, temporal, and social environment with a reasonable degree of success.*

Commonsense knowledge is useful exactly because it is a description of the environment that is maintained at very low cost. Planning and action can then take place within a rich description of the context that need not be constructed for the particular occasion.

The above definition is superficially quite different from McCarthy's, "a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows" [McCarthy, l968, p. 403]. McCarthy's definition is much more inclusive than mine, applying to any deductive process that does a significant portion of its work in response to new information. In both cases, however, the emphasis is on structuring a description of some environment from observations. The primary goal of McCarthy's Advice Taker research is to show how "interesting changes in behavior can be expressed in a simple way" [p. 404]. My observations here can be seen, in part, as an elaboration on one of the other features he ascribes to commonsense knowledge: "the machine must have or evolve concepts of partial success because on difficult problems decisive successes or failures come too infrequently" [p. 405].

The fact that commonsense knowledge is reasonably useful under real-world performance constraints without concentrated effort suggests an opportunistic way of operating. A system is called opportunistic if it performs computations only when information and processing resources are available and inexpensive, and functions adequately the rest of the time on partially processed information. Since commonsense knowledge is acquired and used under relatively unfavorable conditions, generally as a background to other more pressing activities, it seems forced to adopt an opportunistic mode of operation. Many kinds of knowledge do not lend themselves to opportunistic computation, and so cannot be commonsense knowledge. Vision and motor abilities cannot be opportunistic, for example; the importance to survival of quick recognition and physical activity is too high, implying that these processes must be able to demand resources when necessary. To make a different contrast, expert problem-solving often requires concentrated effort, and is very vulnerable to the destructive effects of interruptions.[1]

There are certain performance constraints (PCs) that are observably satisfied by human commonsense knowledge.[2] We must evaluate potential representations for commonsense knowledge against such performance constraints.

PC1. Learning must be possible, i.e., it must be computationally feasible to combine new information with what is already known and store the

[1] It is not always possible to use the need for concentration to distinguish expert from commonsense knowledge. Many kinds of expertise apparently involve simply a more appropriate way of describing the world, for the purpose at hand. A representation for such expert knowledge would have exactly the same properties as a commonsense representation. Indeed, some experts consider their expertise to be simply "educated common sense." A doctor, for example, may make a preliminary diagnosis on the basis of evidence that appears obvious and compelling to him even at first glance, but might be missed entirely by the layman. On the other hand, a doctor's expertise also includes processes for considering and excluding possible alternative diagnoses by judging the interactions among multiple diseases and treatments, a process that can require sustained concentration.

[2] "Performance constraints" as used in this chapter have little to do with Chomsky's [1965] distinction between competence and performance. This chapter examines the implications of certain practical limitations of the real-world computing environment for the structure of useful representations of commonsense knowledge. Aside from inspinng the selection of constraints, the remarks here are independent of the properties of people. A "performance theory of common sense" (in Chomsky's sense) would be an enterprise of almost unbounded scope, attempting to explain what makes people use knowledge in practice the way they do.

result [Newell and Simon, 1972, p. 8].

PC2. Performance must degrade gracefully under limitations of information and processing resources, rather than failing catastrophically [Norman and Bobrow, 1975; Marr, 1975].

PC3. The amount of working memory (short-term memory) available to any process is limited and probably quite small [Miller, 1956].

PC4. Processing time is subject to frequent interruptions, destroying some or all of the contents of working memory [Norman and Bobrow, 1975].

PC5. Observational input is generally local and fragmented with respect to the scale of the overall phenomenon being described [Minsky and Papert, 1969].

PC6. The contents of long-term memory are occasionally destroyed or lost.

There are two other constraints that seem very plausible to apply to knowledge representations, but their implications remain quite unclear at this time:

PC7. The same basic representational framework should accommodate the range of individual variation that is observable among human performances on the task [Hunt et al., 1973].

PC8. Development should be possible. There should be a sequence of representations by which adult performance can be reached without disastrous lapses of performance on the way as one representation is replaced by another [Howe and Young, 1976].

When discussing "graceful degradation of performance," we must outline the different kinds of potential disasters that we are trying to avoid. Listed below, in decreasing order of severity, are types of disasters that can befall a computational system as a consequence of an internal error. People and computer programs are both subject to such disasters (Ds) but computer programs incline toward the more severe ones, while people are able to restrict their computational disasters to the milder end of the spectrum.

D1. The computational system attains an illegal state from which no further action is possible, and halts or "crashes." For example, a computer will halt if it is instructed to examine the contents of a nonexistent memory location.

D2. Previously acquired and stored information is rendered worthless and must be discarded. For example, if an interruption occurs while an array is being shifted in place, and the working memory of the shifting process is lost, then the entire array has become inconsistent and must be discarded.

D3. The current action cannot be continued and must be abandoned. For example, in evaluating an arithmetic expression, if a variable returns a nonnumerical value, or a function cannot return a value at all, evaluation halts. This is actually a version of D1, taking place within a larger context which can continue functioning when the current activity fails.

D4. Available observations of interest cannot be represented and stored, and so are discarded. For example, in compiling a map of important features of an area, an observation of the heading of a remote landmark, without its ground position, cannot be represented and stored. When the map is used subsequently, information about that landmark is not represented.

D5. A question cannot be answered, perhaps delaying the reorganization of a body of information. For example, the local coordinate frames of two places will not be combined if the curvature of the road connecting them is unknown.

2 What Is Representing Knowledge?

By Palmer's [1977] definition, a representation consists of two sets of objects and relations, and a correspondence between them. The two sets are known as the represented domain and the representing domain. The correspondence between them allows questions about the represented domain to be answered by examining the representing domain.

In normal usage, the objects and relations in the representing domain are often referred to as the "representation," with an implicit reference to the represented domain and the correspondence. The assumption in artificial intelligence is that computational objects and relations make a particularly good representing domain. In this case, the processes that examine the representing objects and relations to answer questions are also of interest, and so are referred to as part of the representation as well. Bobrow [1975] discusses a number of dimensions along which computational representations can be compared.

Notice that there are two different kinds of enterprises involved in representing commonsense knowledge in the computer. First, there is research attempting to find computational representations for objects and relations in the world that satisfy a given set of performance constraints like those in Section I. This research takes place in the fields of mathematics, computer science, and artificial intelligence. Second, there is empirical research in psychology that attempts to characterize the human phenomenon of commonsense knowledge by formulating and testing constraints like those above. The research presented in this chapter addresses the first of these questions. (Naturally, the two kinds of research cannot be kept as isolated as this simple dichotomy suggests since the concepts used in each endeavor can and do influence the investigations in the other.)

The mapping between a state of the world and its computational representation or description can fail to be unique, in both directions. The most commonplace version of this is the observation that there are many possible worlds that can satisfy any (consistent) description. Since a description captures only certain facets of the world, the same description can hold of situations that vary in the undescribed facets. Conversely, there can also be many different descriptions of the same world, capturing different facets of it.

It is also true, however, that a given facet of the world can have many different descriptions, corresponding to the states of knowledge an observer goes through while learning about its global structure from a series of local observations. This set of states of partial knowledge can have quite different properties from the set of states that the world itself can take on. That this point is nontrivial can be seen through a simple example.

Let the "real world" consist of a one-dimensional space S (e.g., a street) and a set P of points in S related by an order relation (e.g., places on the street):

$$\cdots A \cdots B \cdots C \cdots D \cdots$$

The "real world" satisfies the condition

$$x, y \in P \Rightarrow (x = y \text{ or } x > y \text{ or } y > x).$$

On the other hand, commonsense knowledge of one-dimensional order need not satisfy this condition. A state of knowledge may correspond to:

$$\cdots A \cdots \{B, C\} \cdots D \cdots$$

with no order on $\{B, C\}$. (It is easy to duplicate this example by asking people about the order relations of places on streets.) Thus, the relation in the "real world" is a total order, while the relation that makes up the corresponding commonsense knowledge is a partial order.

This example shows that a representation for commonsense knowledge (in this case 1-D order), if it is to support learning in a natural and efficient way, must be capable of expressing the intermediate states of knowledge that a person passes through as he learns the order from observations. Along with the other performance constraints, to be discussed below, this suggests that the properties of a representation may be quite different from the properties of the world being described.

3 Design Features of Commonsense Representations

In this section, I shall try to draw some conclusions about properties of commonsense representations from the performance constraints outlined above. Most of these conclusions are motivated by comparisons between examples that differ in whether or not they satisfy the various constraints rather than being deduced as logical consequences of the constraints. Ideally, the properties that are suggested by these comparisons will later be formalizable through more careful mathematical analysis. Extracting some of the relevant characteristics of the performance constraints, we shall look at the ability of representations to function adequately in the face of:

1. Partial destruction of the represented information,
2. Premature demand for the result of an operation,
3. Resource limitations in (a) working memory, (b) computing time, (c) additional information.

The first property we want in a representation for commonsense knowledge is the ability to express intermediate states of the learning process. This property is motivated in part by resource limitations in additional information. A good illustration is the comparison between partial and total orders as representations for knowledge of one-dimensional order. If order relations are observed as sequences that are subsets of the order in the environment, then two possible representations for the description are the partial order and the total order. Suppose, for the sake of concreteness, that the current state of the description is $(ABCD)$, and the observation is (AEC). If the representation can express any partial order, then $(A\{BE\}CD)$ can be stored, perhaps with some cost in working memory and computing time; there is no cost in additional information. However, if only total orders can be stored, then an additional piece of information is required before (AEC) can be assimilated: either (BE) or (EB).

This requirement for additional information, gotten either from further observation or retrieved from elsewhere in memory, can be seen as a cost of the assimilation process, just like working memory or computing time. Furthermore, if the additional information must be gotten from further observation, or is not conveniently indexed in memory, it is a cost that can be quite hard to pay. Such a cost, and the corresponding delay in updating the long-term description of the environment, makes the system more vulnerable to interruptions that could destroy the temporary state and lose even the first observation. Thus, it is a valuable feature for a representation to be able to express the intermediate states of the learning process, corresponding to an arbitrary sequence of relevant observations.

The second property we want in a representation for commonsense knowledge is resilience in the face of partial destruction of the represented information. The performance constraints that motivate this property are the fact that commonsense operations are subject to frequent interruptions that destroy part of the state of working memory, and that long-term memory seems subject to some sort of destructive process. If part of the information in a description is destroyed, we would like the representation to degrade gracefully, suffering only from the actual information lost, rather than being utterly destroyed by violation of some global consistency condition.

As an example of loss of working memory under an interruption, compare three different kinds of tree search: breadth-first, depth-first, and "best-first." The transient state of a breadth-first search is a set of nodes that span the width of the tree, in that every path from root to leaf must intersect the set. Loss of a single element destroys this property and renders the rest of the set useless. (In fact, it is possible for a separate recovery procedure to "back up" the breadth-first search from the remaining fragments, to reach a previous state from which the search can be restarted.) The transient state of a depth-first search is a linked list of nodes connecting the current node with the root of the tree. Loss of a single link can carry with it an arbitrarily long tail of the list, depending on which link was lost, but the initial segment of the list remains a meaningful state for a depth-first search. Thus the depth-first search can be restarted immediately, having to repeat some work, but neither having to start over nor having to reconstruct its state from the remaining fragments. The transient state of a best-first search is a set of nodes judged to have the "best chance" of dominating the desired target. Loss of a single node from this set decreases the chance of success, but is still a meaningful state of a best-first search, and can be resumed without recovery procedures.

A similar example can be shown in the case of long-term memory by looking at the two familiar order representations: represent an order as a set of ordered pairs (which could be links in a semantic net or 2-tuples in a list). A partial order obeys only the global condition that no loops occur, while a total order also states that any two elements of the ordered set must be comparable. Loss of a single pair clearly leaves the condition on a partial order unaffected, but can render the total order meaningless. Notice that in both kinds of memory, the more resilient representations not only continue to function adequately after partial destruction of information but need not even explicitly detect the fact that an error has taken place. The remainder after destruction is a meaningful state of the representation that can be treated exactly like the complete version.

The third property we want in a representation for commonsense knowledge is the ability to express a meaningful response when a process is prematurely asked for the result of a computation. This property is motivated by the frequency of external interruptions, and by limitations in memory and computing resources that may stop a process before it has run to completion. This property refers at least as much to the ability of the surrounding context to use the response as to the ability of an interrupted process to return one.

The simplest response to a premature request for a result is "don't know." Many representations are not able to represent and store "don't know" as part of a description, and so a "don't know" response triggers an error and the system halts or the current activity is abandoned. This passes the problem onward in the hopes of finding a supervisory process resilient enough to treat "don't know" as a value. A simple case of using "don't know" effectively is the structure of a database that responds to an unsuccessful search by computing and storing the requested vaiue, ensuring that searches appear never to be unsuccessful, and ensuring that computations are done as seldom as possible, at the expense of a larger database.

A more complex example of resilience in the face of premature demands for results comes from the taxonomic (or IS-A) hierarchy, used by a process attempting to identify some specimen. A classification tree for birds or plants is one familiar case and the human recognition system may be another [Quillian, 1968; Anderson and Bower, 1974]. A premature request for the identity of a specimen might produce "marsh bird" or "oak," rather than the actual species, but the larger category is as useful as the smaller in most contexts. This example suggests that the taxonomic hierarchy is useful not just as a convenient abbreviation technique that lets shared properties be associated with shared categories, but that it is essential to providing commonsense resilience to interruptions in the identification process.[3]

It has been widely observed that severe time and working memory limitations must be met by processes that make up commonsense knowledge. The mathematical analysis of algorithms (e.g., Knuth [1968]) provides a rich collection of techniques for analyzing representations and procedures. Certainly, many kinds of tree search make excessive resource demands of this kind and cannot be considered serious candidates for operations on representations for large amounts of commonsense knowledge.

There are typically trade-offs of various kinds in the design of representations and algorithms: time vs. space and storage vs. retrieval are two prominant examples. The kinds of disasters outlined in Section I, as well as the observations of McCarthy [1968], suggest that a commonsense knowledge representation should optimize storage of information-the maintenance of an adequate description of the environment-even at the expense of retrieval or the ability to solve each particular problem successfully. Since either kind of process will be vulnerable to interruptions, it is important that the initial storage of observations be as efficient as possible, and that the subsequent assimilation process be made up of small steps so as not to suffer excessively from interruptions.

4 Partial Order

The partial order is a mathematical structure that appears frequently in attempts to devise a computational representation for knowledge. A partial

[3] Thus we see a taxonomic hierarchy as a set of states of partial knowledge of identity useful in the recognition process. If this is true of people, we should find that its structure is not determined by the number of shared characteristics among individuals (as would be required by economy of storage) but by the shared states passed through in the process of recognition. Thus, "whale" fits into the hierarchy under "fish" not because our ancestors thought whales were fishes or even because they have many common characteristics but because the states of partial knowledge encountered in recognizing a whale have more in common with "fish" than with "mammal."

order $(S, <)$ is a set S and a binary relation $<$ on the elements of S that satisfies:

(1) for no $x \in S$, $x < x$,
(2) for any $x, y, z \in S$, if $x < y$ and $y < z$, then $x < z$.

A total order is a partial order that also satisfies:

(3) for any $x, y \in S$, $x < y$ or $x = y$ or $y < x$.

A partial order can arise when representing a relationship, such as containment among subsets of a set, that is inherently partially ordered in the world. Any set with more than one element has subsets that are not related by containment. On the other hand, a partial order can also arise when representing states of partial knowledge about a total order, such as the order of places on a street (as in Section 2).

A partial order can be represented in the computer either as a distributed collection of pointers, or as a single complex data structure. The containment relation might be represented in a semantic network, for example, as a SUBSET-OF link between the nodes representing two regions. On the other hand, the node for a street may refer to a data structure representing, in one place, the current state of knowledge about the order of places on that street.

Two of the design features we want in commonsense representations are satisfied directly, because a partial order is the transitive closure of an arbitrary set of order relations. First, if an observation is a sequence of elements of S, then any set of observations can be represented by some partial order. This is clearly false of total orders, since there are some sets of observations that do not determine a total order uniquely. Second, if some of the order relations are lost, due to a destructive process acting on memory, the remaining information still constitutes a partial order, which furthermore is simply a less-specified version of the original.

Now let us examine a representation of a partial order in the computer. Let $(P, <)$ be a finite set with a partial order. Assume that we already have a representation for the elements of P, so that for each x in P, $f(x)$ is its computer representation. Now let S be a set of sequences of elements from $f(P)$. Define a new relation $\ll$ on $f(P)$ by saying that $a \ll b$ for a and b in $f(P)$ if there is a sequence in S in which a precedes b, and let $\ll^*$ be the transitive closure of $\ll$. Then $(f(P), \ll^*)$ is a partially ordered set, and S can be chosen so that

$$x < y \text{ iff } f(x) \ll^* f(y).$$

Thus, for an arbitrary, finite, partially ordered set, we can represent the order relation in the computer as a set of sequences.

Given the correspondence above, we can define access procedures that operate on the set-of-sequences data structure to answer questions about the represented partial order, and to change it as the partial order changes:

$$
\begin{aligned}
ORDER(a,b,S) &= \begin{cases} +1, & \text{if } a \ll^* b \\ -1, & \text{if } b \ll^* a \\ \text{``don't know''} & \text{otherwise} \end{cases} \\
MERGE(seq,S) &= S \text{ with the side-effect that } seq \text{ has been added to the partial order } S. \\
REORGANIZE(S) &= S \text{ rearranges and combines the sequences in } S \text{ to make } ORDER \text{ more efficient without changing its logical properties.}
\end{aligned}
$$

The retrieval function ORDER can in general be quite inefficient, especially for pathological, highly branching partial orders. However, since the surrounding context must already be capable of dealing with a "don't know" response, an interruption or exhaustion of resources simply results in the least painful class of disasters D5.

MERGE needs only to append its sequence to S and do no extended computation, and so it is not very vulnerable to destructive interruptions and unlikely to lose observations (D4). Extended processing of input is left to REORGANIZE, which examines the set S of sequences, rearranging and combining them to achieve the minimum of branching. However, since the value returned by a call to ORDER is unaffected by REORGANIZE, it can operate completely in the background, freely interruptable by more urgent needs, but when resources are available making S suitable for more efficient access by ORDER.

In the TOUR model of commonsense knowledge of large-scale space [Kuipers, 1977, 1978], such a partial order data structure is part of the description of a street, representing the current state of partial knowledge about the order of places on the street. The observations provided are sequences of places observed when travelling along the street, and they are merged into the street's partial order. However, there is an ambiguity left in the correspondence between an observed sequence (ABC) of places, and the partial order data structure S associated with the street description. Does the observation of (ABC) mean $A \ll^* B \ll^* C$ or $C \ll^* B \ll^* A$?

Information must be provided by the global context to resolve this ambiguity. In the TOUR model, one element of the global context is the current one-dimensional orientation of the traveller on the current street. The 1-D orientation takes on the values $+1$, -1, and "don't know," exactly as provided by ORDER, and represents the correspondence between the dynamic order of observations and the static order associated with a street description. In case the 1-D orientation is "don't know," (ABC) is treated as $\{(A), (B), (C)\}$, and the order information is lost until it is observed again. Inference rules in the TOUR model specify how the 1-D orientation is obtained and updated, and how it guides the interpretation of observations such as these.

5 Route Description

Sequential behavior is an important part of the life of any intelligent human or computer, and the learning of these sequences is an important aspect of commonsense knowledge. I shall present in this section an example of a representation for descriptions of routes through a large-scale space, intending to show how the gradual accretion of local observations produces several qualitatively different phases of behavior. (There are certainly attractive parallels with other kinds of learned sequential behavior, which bear further investigation in other contexts. However, it should also be observed that a representation such as this one is certainly inadequate for many other kinds of sequential behavior, such as language production.)

Let a view V be a description of an observable piece of the environment from one position, and let an action A be some motion that can be performed in the context of that piece of environment. To model the role of the environment in travel, let us define the function $RESULT(V, A) = V'$, which is interpreted as "if performing action A in the context of the piece of environment described by V, the result will be the observation V'." Notice that $RESULT$ does not model a mental operation; it models the physical fact that if you see one thing (V), then turn a corner (A), you will see something different (V').

I am thinking of a view as a visual image and an action as walking a distance or turning, although other interpretations are possible and reasonable. The internal structure of these descriptions is assumed to be extremely complex and data-rich, and fortunately is not of concern to us here. Assume that your eyes are capable of delivering a view description V that can be compared for identity with stored views V', V'', etc., and that an action A can be recorded if performed, and performed if recalled in the appropriate context.

With these preliminaries taken care of, we can define a route description as a set of triples (V, A, V'), where a triple may allow either the A or V' position to be unspecified. [For notation, $(V, A, ?)$ leaves open the question of whether V' was specified, and $(V, A, \_)$ states that it was not.] As part of a route description, (V, A, V') represents the imperative of doing A when in the context described by V, and the assertion that $RESULT(V, A) = V'$.[4]

Three operations are permissible on a route description. If RD is a route description:

$$INSERT(triple, RD) = RD \text{ and adds } triple \text{ to the set } RD$$

$$ASSOC(V, RD) = \begin{cases} (V, ?, ?) & \text{if such a triple is in } RD \\ \text{“don't know”} & \text{otherwise} \end{cases}$$

A or V' can be added to a $(V, \_, \_)$ or $(V, A, \_)$ triple.

In particular, the route description itself is a set with no sequential structure. That will arise as a property of the particular set of triples.

As information is gradually added to a route description, the behavior it supports goes through three phases:

1. Considered simply as a collection of triples, a route description RD supports recognition of a previously observed view $(V, \_, \_)$, knowledge of the action to take at that point $(V, A, \_)$, and even anticipation of the result of the action (V, A, V'). In the early stages of learning a route description, the set contains only a few partially filled triples and so it will support only occasional recognition and further acquisition of information but not selfguided travel.

2. The first threshhold is passed when RD contains enough $(V, A, ?)$ triples to complete a sequence

$$V_0, A_0, V_1, A_1, \cdots V_n$$

connecting the beginning and end points of the route, satisfying

[4] The similarity of (V, A, V') triples to stimulus-response associations has not gone unnoticed but arose from the computational requirements of the problem, rather than being a theoretical principle that I brought to this design. However, a (V, A, V') triple is quite different from a stimulus-response pair in that it also acts as a symbolic assertion that can be the input or output of deductive procedures.

(a) $ASSOC(V_i, RD) = (V_i, A_i, ?)$
(b) $RESULT(V_i, A_i) = V_{i+1}$

This permits the route description to be followed as an imperative set of instructions, but only in conjunction with physical observation of the environment. The association between view and action is contained in the route description, while order information is contained in the physical structure of the environment.

3. The second threshhold is passed when RD contains enough triples of the form (V, A, V') so that the sequence:

$$V_0, A_0, V_1, A_1, \cdots V_n$$

can be constructed solely from

(c) $ASSOC(V_i, RD) = (V_i, A_i, V_{i+l})$

In this case, the order information is now represented explicitly in the route description, and need not be acquired through physical interaction with the world. This level of organization supports planning and mental rehearsal of the route, which is valuable for constructing descriptions of larger spatial structures, as well as physical travel over that particular route.

This representation is interesting for a number of reasons. First, its performance undergoes two qualitative changes over a range of very simple information acquisition steps. Second, adding a new piece of information consists of either inserting a new triple into the set, or of filling an empty slot in a triple. Neither constitutes a heavy computational load or is likely to suffer much from destructive interruptions. Third, the only search involved is the associative retrieval from the set of triples.

This route description is also reasonably resilient in the face of various kinds of degraded performance:

1. Extra information is no burden at all: associative retrieval simply overlooks information that is not needed.

2. Loss of V' from a triple is not a barrier to physical travel (D3), but only to mental rehearsal of the route description, and therefore to some kinds of structure-building (D5).

3. Loss of either A from a triple, or of an entire crucial triple, is a barrier to self-directed travel (D3), but not to recognition, and it does not render the remaining route description worthless (D2). Later observations can still fill the gap and restore full performance without an error ever having to be explicitly detected and recognized.

In the TOUR model, this route description bridges a gap between a plausible level of observational experience and the sequence of GO-TO and TURN instructions that formed the basis of the early formulations [Kuipers, 1977, 1978]. The further process of assimilation described there takes spatial information from the context of particular route descriptions and associates it with fixed features of the environment. Thus the egocentric view and action descriptions are assimilated into a nonegocentric place description whose local geometry summarizes the possible actions that can take place there. Similarly, order information that is implicit in the sequence of places encountered on a route is associated with path descriptions of the streets involved.

6 Conclusion

This chapter has essentially been about graceful degradation: how the performance of a representation for commonsense knowledge can survive the unpleasant computational environment of the real world. In particular, how it can minimize the level of disaster resulting from interruptions, destruction of information, and limitations in resources. Important properties of commonsense representations include (1) the ability to express intermediate states of the learning process, (2) resilience in the face of partial destruction of the represented information (preferably without even having to notice the destruction), and (3) the ability to express a meaningful response when a process is prematurely asked for the result of a computation. A number of examples were presented to illustrate different aspects of these properties.

The examples and properties give a qualitative description of the characteristics that are important in a commonsense representation, but they need to be supplemented by a quantitative, mathematical analysis.

> *How* resilient is a representation under *what kinds* of destruction?
> *Which* sequences of observations can and cannot be represented completely?
> *What kinds* of partial results of procedures are usable by the process that made a premature demand for an answer?

How bad is a given class of disaster, for a given request in a given context?

For some of the questions ("how resilient," "how bad"), it is not even clear how to state the question mathematically or what form the answer would take. For other questions, the answer requires a precise description of the set of all possible observations, outputs, or items vulnerable to destruction .

Another intriguing class of mathematical questions that arises from these examples concerns the characterization of commonsense knowledge. How many computational representations are there? Infinitely many, of course, but can they be generated in some interesting way from a small basis, the way we describe cyclic groups? Of all these representations, how many satisfy the performance constraints we would like to apply to representations of commonsense knowledge? Even if we cannot generate all representations, perhaps we can generate all commonsense representations?

An alternative possibility is that there really are not very many commonsense representations at all, perhaps a dozen or two. Then we could catalog them in the way we catalog the "simple machines" out of which other mechanical devices are constructed. The wide use of metaphor from concrete physical and spatial knowledge to more abstract domains ends some plausibility to this curious notion. Perhaps one reason for the use of metaphor is that there are only a few suitable representations for commonsense knowledge, we learn them initially when learning about the physical world, then apply them metaphorically to get access to powerful representational devices in new domains.

Acknowledgments

I would like to thank the reviewers for their helpful comments, and the inhabitants of the MIT Al Lab for their general inspiration. This chapter is based on research that was performed while the author was at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Project Agency of the Department of Defense under Office of Naval Research contract N00014-75-C-0643.

References

[Anderson, J. R., and Bower, G., 1974]. *Human Associative Memory*, 2nd ed. Washington, DC: Hemisphere Publishing Corp., Washington, D.C.

[Bobrow, D. G., 1975]. Dimensions of representation. In *Representation and Understanding*. D. G. Bobrow and A. Collins (eds.). Academic Press, New York, pp. 1-34.

[Chomsky, N., 1965]. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, Massachusetts.

[Howe, J. A. M., and Young, R. M., 1976]. "Progress in Cognitive Development," Research Report No. 17, Department of Artificial Intelligence, University of Edinburgh, Edinburgh, Scotland.

[Hunt, E., Frost, N., and Lunneborg, C., 1973]. Individual differences in cognition: A new approach to intelligence. *Psychology of Learning and Motivation* 7, 87-122.

[Knuth, D. E., 1968]. *The Art of Computer Programming*. Vol. 1. Addison-Wesley, Reading, Massachusetts .

[Kuipers, B. J., 1977]. "Representing Knowledge of Large-Scale Space," TR-418. Artificial Intelligence Laboratory, MIT, Cambridge, Massachusetts. (Doctoral thesis, MIT Mathematics Department.)

[Kuipers, B. J., 1978]. Modelling spatial knowledge. *Cognitive Science* 2, 129-153.

[McCarthy, J., 1968]. Programs with common sense. In *Semantic Information Processing*. M. Minsky (ed.). MIT Press, Cambridge, Massachusetts, pp. 403-418.

[Marr, D., 1975]. Early Processing of Visual Information. *Phil. Trans. Roy. Soc. B.* 275, 483-524.

[Miller, G. A., 1956]. The magical number seven, plus or minus two. *Psychology Review* 63, 81-97.

[Minsky, M. L., and Paper, S. A., 1969]. *Perceptrons*. MIT Press, Cambridge, Massachusetts.

[Newell, A., and Simon, H. A., 1972]. *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, New Jersey.

[Norman, D. A., and Bobrow, D. G., 1975]. On data-limited and resource-limited processes. *Cognitive Psychology* 7, 44-64.

[Palmer, S. E., 1977]. Fundamental aspects of cognitive representation. In *Cognition and Categorization*. E. H. Rosch and B. B. Lloyd (eds.). Erlbaum Press, Potomac, Maryland.

[Quillian, M. R., 1968]. Semantic memory. In *Semantic Information Processing*. M. Minsky (ed.). MIT Press, Cambridge, Massachusetts, pp. 227-270.