Towards a General Theory of Action and Time

James F. Allen
Computer Science Department, University of Rochester, Rochester, NY 14627, U.S.A.

Recommended by Yorick Wilks

ABSTRACT

A formalism for reasoning about actions is proposed that is based on a temporal logic. It allows a much wider range of actions to be described than with previous approaches such as the situation calculus. This formalism is then used to characterize the different types of events, processes, actions, and properties that can be described in simple English sentences. In addressing this problem, we consider actions that involve non-activity as well as actions that can only be defined in terms of the beliefs and intentions of the actors. Finally, a framework for planning in a dynamic world with external events and multiple agents is suggested.

1. Introduction

The concept of action arises in at least two major subareas of artificial intelligence, namely, natural language processing and problem solving. For the most part, the formalisms that have been suggested in each subarea are independent of each other and difficult to compare, although there is recent work that attempts to merge work from both areas [1]. Even considering such work, however, there is presently no computational theory of action that is sufficiently powerful to capture the range of the meanings and distinctions expressible in English. The primary goal of this paper is to suggest a formalism that is considerably more expressive than current theories of action and to explore its use in defining the meanings of English sentences that describe actions and events.

A secondary, but important, requirement on the formalism is that it should be a useful representation for action reasoning (i.e., problem solving). Some effort will be made to describe how the representation could be used by a planning or plan recognition system. This is essential to the natural language research as well, because problem-solving techniques are being used more and more in our models of language comprehension (e.g., see [2]). While interest in this approach is growing, progress is inhibited by inadequate representations of actions and plans.

There are at least three major difficulties with nearly all existing models of action in AI. Such models cannot represent:
- actions that involve non-activity (e.g., "I stood on the corner for an hour");
- actions that are not easily decomposable into subactions (e.g., "I spent the day hiding from George");
- actions that occur simultaneously and many interact with each other.

The theory outlined below will allow all three of these situations to be represented. Each problem will be examined in detail below when discussing previous work.

1.1. Relevant work

Relevant work on this problem can be divided into three broad categories: representations for natural language systems, problem-solving systems, and work in linguistics and philosophy.

The most common formulation for actions in natural language systems is based on case grammar [3]. Each action is represented by a set of assertions about the semantic roles the noun phrases play with respect to the action denoted by the verb. Such a formalism is useful for interpreting the semantic structure of sentences, but doesn't address the issue of what an action is, or what inferences can be made from the fact that an action occurred. Typically there is only a simple temporal, or causal, ordering on the actions, which is not heavily used. Such representations only work in situations where actions are simply related to each other and no uncertainty exists.

Work in problem solving has used more sophisticated models of action and time. The most influential theory in this work has been the situation calculus [4]. The world is represented as a set of situations, each describing the world at a single instant in time. An action is a function from one situation to another, and can be described by a set of prerequisites on the initial situation and a set of effects that will hold in the final situation. While this model has been extremely useful in modeling physical actions by a single agent in an otherwise static world, it cannot easily be extended to account for simultaneous actions and events. For example, the action described by the sentence, "I walked to the store while juggling three balls," seems to be composed of the action of walking to the store and the action of juggling three balls. It is not clear how such a composite action would be defined if we view an action as a function from one instantaneous world description to another. Furthermore, since an action in the situation calculus is equated with change, actions that involve no activity, or restore the world to its original state (e.g., running around a track), cannot be modeled.

The most common implementation inspired by the situation calculus is the

Artificial Intelligence 23 (1984) 123–154
0004-3702/84/$3.00

state space model. The world is described by a data base containing what is presently true, and actions are simulated by adding and deleting facts from the data base. This model suffers from all the above criticisms of the situation calculus and in addition has no model of the past or the future.

None of these formulations can describe the following simple situation. In the blocks world, assume we have two actions, PUSHR, push to the right, and PUSHL, push to the left. Let us also assume that the effect of each of these actions is that the block moves one unit in the appropriate direction. But if two robots perform a PUSHR and a PUSHL simultaneously on the same block, the block doesn't move. Since we cannot express or reason about the simultaneity of actions in any of the above theories, we cannot express this situation. The best we can do will be to have the block oscillate as the robots push alternately.

McDermott [5] introduces a notion of event that is general enough to address the three problems above. To a first approximation, our two approaches are compatible; however, major differences exist. Some of these differences come about as a result of the difference in intended application, and some as a result of different underlying assumptions in our temporal logics. These issues will be discussed as they become relevant.

Work in linguistics and philosophy has provided many insights for this research, although the theories typically lack a computational viewpoint. The major influences on this paper come from Mourelatos [6], Jackendoff [7], and Goldman [8].

Mourelatos presents a detailed analysis of the different classes of occurrences describable in English, and his terminology will be adopted here. The term *occurrence* is used to describe the class of all events, processes, actions, activities, and accomplishments. This effectively includes all forms of sentence meanings except for assertions of states such as "The building is red" or "I am ten years old". The class of occurrences is further subdivided by Mourelatos, and we will consider this subcategorization later as it becomes relevant.

Goldman presents a detailed theory of human action. At this stage, we need only examine the problem of action individuation, which is demonstrated by the question: if I pull the trigger of a gun, thereby shooting my friend, how many actions have I performed? One view is that only one action is performed, and that it exemplifies two action types [9], pulling the trigger and shooting my friend. Goldman's view is that two distinct—though intimately related—actions are performed. The latter view is adopted here as it provides for a simpler semantics. Thus a physical situation will typically be described as a set of distinct occurrences and actions, and actions may be distinct even if they have the same agent and occur at the same time.

1.2. The proposed theory

The world is described by a set of temporally qualified assertions outlining what is known about the past, present, and future. This includes descriptions of both static and dynamic aspects of the world. The static aspects are captured by *properties* (e.g., Cleo owning a car) that hold over stretches of time. The dynamic aspects are captured by *occurrences* (e.g., Cleo running a race) which describe change of forms of resistance to change over stretches of time. The distinction between properties and occurrences is usually obvious, but situations do arise where it is quite subtle. The most subtle cases arise from situations that can be described from either a static or dynamic perspective. For instance, adopting an example from Jackendoff [7], we could describe the situation in which a light was on in a house all night by the static description, "the light was on all night", or by the dynamic description, "the light remained on all night". The only difference here appears to be that the latter suggests that it might have been otherwise. In most cases, however, the static/dynamic distinction will be straightforward.

It should be clear that a temporal logic is necessary to support this theory. Time plays a crucial role, and cannot be relegated to a secondary consideration as in most natural language systems or hidden in a search process as in most problem-solving systems. The temporal logic described below is based on temporal intervals, and denies the standard view of mapping time to points on the real number line.

Given a temporal logic, we address how *occurrences* can be defined. In particular, what do we know when we know an occurrence has occurred? In problem-solving systems, actions are described by prerequisites (i.e., what must be true to enable the action), effects (what must be true after the action has occurred), and decomposition (how the action is performed, which is typically a sequence of subactions). While such knowledge is crucial for reasoning about what actions to perform in a given situation, it does not define what we know when we know an action has occurred. To clarify this, consider the simple action of turning on a light.

There are few physical activities that are a necessary part of performing the action of turning on a light. Depending on the context, vastly different patterns of behavior can be classified as the same action. For example, turning on a light usually involves flipping a light switch, but in some circumstances it may involve tightening the light bulb (in the basement) or hitting the wall (in an old house). Although we have knowledge about how the action can be performed, this does *not* define what the action is. The key defining characteristic of turning on the light seems to be that the agent is performing some activity which will cause the light, which was off when the action started, to become on when the action ends. An important side effect of this definition is that we could recognize an observed pattern of activity as "turning on the light" even if we had never seen or thought about that pattern previously.

Thus, we want a level of causal explanation that characterizes the consequences one can draw from a sentence describing an action, or, more generally, an occurrence. Such a description would not replace or subsume the

prerequisite/effect/method characterization of actions, although there will be some overlap. For example, the effects of an action should be included, or be derivable from, the causal definition of the action. Some prerequisites would appear in the definition of an action, though others, dealing with an agent's abilities, might not be part of the action definition. Similarly, some parts of a method might be necessary in an action definition, but for the most part, method descriptions do not define what an action is. At a few places in the paper, we shall consider how problem-solving knowledge about actions can be integrated into the proposed framework.

1.3. An outline of the paper

Section 2 introduces an interval-based temporal logic and discusses properties than can hold over intervals. In Section 3, a logic for occurrences is introduced. Occurrences are subdivided into processes and events. Section 4 introduces the notion of action and makes the distinction between definitional knowledge of an action and generational knowledge which is needed for planning actions. Section 5 deals with intentional action and provides a semi-formal notion of plans and of an agent committing to a plan. The paper concludes with an analysis of the meaning of the verb 'to hide', using the formalism developed.

The following conventions will be used throughout. Predicates and constants will be in upper case, and variables will be in lower case. A full range of connectives and quantifiers will be used in their standard interpretation. I use:

- & conjunction,
- v disjunction,
- ~ negation,
- ⇒ implication,
- ⇔ equivalence,
- ∀ universal quantifier,
- ∃ existential quantifier,
- ∃! existence of a unique object.

Any variables that appear with no quantifier are assumed to be universal variables with global scope. I shall often resort to typing variables as in a many-sorted logic. In these cases, the type of the variable will be indicated by its name. Scoping of operators and quantifiers will be indicated by use of parentheses or by indentation of formulas. In general, quantifiers are assumed to be scoped as broadly as possible.

2. A Temporal Logic

Before we can characterize events and actions, we need to specify a temporal logic. The logic described here is based on temporal intervals rather than time points. This approach arises from the observation that the only times we can identify are times of occurrences and properties. For any such time, say the time I was opening the door, it appears to be possible to look more closely at the occurrence and decompose it; hence, times can be decomposed into subtimes. In other words, it seems that there is always a more detailed causal explanation if one cares, and is able, to look for it. A good analogy, then, is that times correspond to intervals on the real line. If we accept this, why not allow instantaneous time points as well? First, they do not appear to be necessary. Second, instantaneous time points will present difficulties with the semantics of our logic. If one allows time points, one must consider whether intervals are open or closed. For example, consider the time of running a race, R, and the time following after the race, AR. Let P be the proposition representing the fact that the race is on; P is true over R, and ~P is true over AR. We want AR and R to meet in some sense. Whether both ends of the intervals are open or closed, AR and R must either share a time point or allow time between them. Thus we have a choice between inconsistency or truth gaps, i.e., either there is a time when both P and ~P are true, or there is a time when neither P nor ~P is true. One solution to this problem is to stipulate by convention that intervals are open at the lower end and closed at the upper end, but then every interval has only a single endpoint. The artificiality of this solution reinforces the argument against allowing points. Events that appear to refer to a point in time (e.g., finishing a race) are considered to be implicitly referring to another event's beginning or ending. Thus time 'points' will be considered to be very small intervals. This will be made more precise below.

The logic is a typed first-order predicate calculus, in which the terms fall into many categories. The following three are needed at present:

- terms of type TIME-INTERVAL denoting time intervals;
- terms of type PROPERTY, denoting propositions that can hold or not hold during a particular time;
- terms corresponding to objects in the domain.

There are a small number of predicates. One of the most important is HOLDS, which asserts that a property holds (i.e., is true) during a time interval. Thus

HOLDS(p, t)

is true if and only if property p holds during t. As a subsequent axiom will state, this is intended to mean that p holds at every subinterval of t as well. Note that if we had introduced HOLDS as a modal operator we would not need to introduce properties into our ontology. We have not followed this route, however, since it seems more complicated in the later development of occurrences.

There is a basic set of mutually exclusive primitive relations that can hold between temporal intervals. Each of these is represented by a predicate in the logic. These relationships are summarized in Fig. 1.

- DURING($t1$, $t2$): time interval $t1$ is fully contained within $t2$;

as we would expect from intuition. Mirroring the normal definition of disjunction to define the function 'or', i.e.,

$$\text{HOLDS}(\text{or}(p, q), t) \equiv \text{HOLDS}(\text{not}(\text{and}(\text{not}(p), \text{not}(q))), t),$$

we can derive

$$\text{HOLDS}(\text{or}(p, q), T) \Leftrightarrow \quad \text{(H.7)}$$
$$\forall t.\text{IN}(t, T) \Rightarrow (\exists s.\text{IN}(s, t) \wedge (\text{HOLDS}(p, s) \vee \text{HOLDS}(q, s))).$$

Many treatments of temporal logics introduce the notion of branching futures into the model. This is used to analyze the notion of possibility of some event (i.e., there is a branch in which it occurs), and necessity of some event (i.e., it occurs on all branches). The model has also been suggested as a computational framework for reasoning about future actions (e.g., [5, 11]).

There is no branching future in the model described here. This is because reasoning about the future is considered to be just one instance of hypothetical reasoning. Other examples include reasoning about the past (i.e., how could the world possibly have arrived at the present state), as well as reasoning independent of time and physical causality (such as mathematics). Since all these forms of reasoning are necessary, it seems arbitrary to put one subclass into the model of time. If there were a good reason to encode such reasoning in a branching time model, then the model should also include a branching past, for the types of mechanisms needed for reasoning about the past and future appear to be identical.

Thus there is a simple single time line (which would correspond to the actual past and actual future in a branching time model). Of course, the reasoner never can totally identify the actual past or future, and reasoning about what actually has occurred or will occur consists of constructing the most plausible hypotheses given what is known about the past, present, and future.

As a final comment, note that this does not mean that the reasoning agent is simply a passive observer of the world. By deciding to do actions, the agent changes his expectations about what the future will actually be. This will be discussed in further detail after we have introduced the notion of events and actions.

3. Defining Occurrences

In order to define the role that events and actions play in the logic, let us consider a possible logical form for sentences asserting that an event occurred and see how it fails. The suggestion is that we define a property for each event class such that the property HOLDS over an interval I just in the cases when an instance of the event class occurred over interval I. But this immediately presents problems, for axiom (H.1) would not hold for such properties. In particular, an event such as turning on the light may occur over an interval T, but not occur over any subinterval of T. In other words, T could be the smallest interval over which "turning on the light" occurred. This cannot be captured by a property, for axiom (H.1) would imply the event occurred over subintervals of T as well.

We introduce a new type of object into our ontology, named an *occurrence*. By representing occurrences as objects in the logic, we are following Davidson's [9] suggestion for representing events. His major argument for this position is that it allows a clean formalism for modeling modifiers and qualifiers of events as predicates acting on the event objects.

Following many others, including Mourelatos [6], we will divide the class of occurrences into two subclasses, *processes* and *events*. Processes refer to some activity not involving a culmination or anticipated result, such as the process denoted by the sentence, "I am running". Events describe an activity that involves a product or outcome, such as the event denoted by the sentence "I walked to the store". A useful test for distinguishing between events and processes is that one can count the number of times an event occurs, but one cannot count the number of times a process is occurring.

Above, we saw that a property could HOLD over many different time intervals. For example, the property that "my wagon is red" might HOLD in the summer of 1981, but not in 1982, and yet again in 1983. We can view a property as defining a set of time intervals over which it holds. We treat occurrences similarly. For example, the occurrence "I walked from home to school" might OCCUR every weekday morning. We cannot specify a particular instance of an occurrence without specifying the unique time over which it occurred.

Properties, processes, and events may be distinguished by considering the characteristics of the set of temporal intervals that they hold or occur over. As we have already seen, the set of intervals over which a property holds is closed under the IN relation. In other words, if interval I is in the set, all intervals J such that $\text{IN}(J, I)$ are also in the set. In contrast, the set of intervals over which an event occurs contains no pair of intervals such that one is IN the other. In other words, an event occurs over the smallest time possible for it to occur. This is the same treatment of events as in McDermott [5]. Processes fall between events and properties. To see this, consider the process "I am walking" over interval I. Unlike events, this process may also be occurring over subintervals of I. Unlike properties, however, it is not the case that the process must be occurring over all subintervals of I. For example, if I am walking over interval I, then I am walking over the first half of I; however, there may be some subintervals of I where I paused for a brief rest.

Let us return to the more formal development of occurrences. We shall start with events, as they are the simplest. The predicate OCCUR takes an event and a time interval and is true only if the event happened over the time interval

t, and there is no subinterval of t over which the event happened. Thus for any event e, and times t and t', we have the axiom

$$\text{OCCUR}(e, t) \,\&\, \text{IN}(t', t) \Rightarrow \sim\text{OCCUR}(e, t') . \qquad \text{(O.1)}$$

Related classes of events can be described using functions. For example, consider the set of events consisting of an object changing location. We can define a function CHANGE-POS with three arguments: the object, the source location, and the goal location. Thus

$$\text{CHANGE-POS}(\text{Ball}, x, y)$$

generates the class of events that consists of a ball moving from x to y. This does not assert that a ball actually did move from x to y. That claim is made by asserting that the event occurred over some time interval. Thus to assert that BALL1 moved from POS1 to POS2 over time $T100$, we say

$$\text{OCCUR}(\text{CHANGE-POS}(\text{BALL1}, \text{POS1}, \text{POS2}), T100) .$$

We can now define necessary conditions for the class of events involving a change of location:

$$\begin{aligned}&\text{OCCUR}(\text{CHANGE-POS}(\text{object}, \text{source}, \text{goal}), t) \Rightarrow\\ &\quad \exists t1, t2 .\\ &\qquad \text{MEETS}(t1, t) \,\&\, \text{MEETS}(t, t2) \,\&\\ &\qquad \text{HOLDS}(\text{at}(\text{object}, \text{source}), t1) \,\&\, \text{HOLDS}(\text{at}(\text{object}, \text{goal}), t2) .\end{aligned}$$

Notice that this definition is of the form

$$\text{OCCUR}(e, t) \Rightarrow P_t$$

where e is an event-defining function and P_t is a set of conditions involving t.

If the P_t are necessary and sufficient conditions for an event's occurrence, we can define the event with an assertion of the form

$$\text{OCCUR}(e, t) \Leftrightarrow P_t \,\&\, \forall t'.\text{IN}(t', t) \supset \sim P_{t'} .$$

This more complicated form is necessary to ensure the validity of axiom (O.1), which insists that an event occurs only over the smallest interval in which it could have. For example, since the conditions above are sufficient to define a CHANGE-POS event, we could have the assertion

$$\begin{aligned}&\text{OCCUR}(\text{CHANGE-POS}(\text{object}, \text{source}, \text{goal}), t) \Leftrightarrow\\ &\quad (\exists t1, t2 .\\ &\qquad \text{MEETS}(t1, t) \,\&\, \text{MEETS}(t, t2) \,\&\\ &\qquad \text{HOLDS}(\text{at}(\text{object}, \text{source}), t1) \,\&\\ &\qquad \text{HOLDS}(\text{at}(\text{object}, \text{goal}), t2)) \,\&\\ &\quad (\forall t'.\text{IN}(t', t) \Rightarrow\\ &\qquad \sim (\exists t3, t4 .\\ &\qquad\quad \text{MEETS}(t3, t') \,\&\, \text{MEETS}(t', t4) \,\&\\ &\qquad\quad \text{HOLDS}(\text{at}(\text{object}, \text{source}), t3) \,\&\\ &\qquad\quad \text{HOLDS}(\text{at}, (\text{object}, \text{goal}), t4))) .\end{aligned}$$

For the sake of readability, we will summarize event definitions by only stating the P_t-conditions and noting whether they are only necessary or are necessary and sufficient. For example, the above example will be written as:

Necessary and Sufficient Conditions for
$\text{OCCUR}(\text{CHANGE-POS}(\text{object}, \text{source}, \text{goal}), t)$:

$$\begin{aligned}&\exists t1, t2 .\\ &\text{MEETS}(t1, t) \,\&\, \text{MEETS}(t, t2) \,\&\\ &\quad \text{HOLDS}(\text{at}(\text{object}, \text{source}), t1) \,\&\\ &\quad \text{HOLDS}(\text{at}(\text{object}, \text{goal}), t2) .\end{aligned}$$

Axiom (O.1) allows us to count events and hence to construct events that are composites of other events. For example, the class of events of repeating an event twice can be defined by:

Necessary and Sufficient Conditions for
$\text{OCCUR}(\text{TWICE}(\text{event}), t)$:

$$\begin{aligned}&\exists t1, t2 .\\ &\quad \text{IN}(t1, t) \,\&\, \text{IN}(t2, t) \,\&\, t1 \neq t2 \,\&\\ &\quad \text{OCCUR}(\text{event}, t1) \,\&\, \text{OCCUR}(\text{event}, t2) .\end{aligned}$$

If we expand this definition out to its full form shown above, we can easily see that it captures repeating exactly twice rather than at least twice.

An event that is a sequence of other events can be defined so that the times of each successive event MEET or are strictly AFTER each other. For example, a two-event sequence with MEET can be defined as follows:

Necessary and Sufficient Conditions for
$\text{OCCUR}(\text{TWO-MEET-SEQUENCE}(\text{event1}, \text{event2}), t)$:

$$\begin{aligned}&\exists t1, t2 .\\ &\quad \text{STARTS}(t1, t) \,\&\, \text{FINISHES}(t2, t) \,\&\, \text{MEETS}(t1, t2) \,\&\\ &\quad \text{OCCURS}(\text{event1}, t1) \,\&\, \text{OCCURS}(\text{event2}, t2) .\end{aligned}$$

Finally, a composite event that will be useful later is the simple composite of two events occurring simultaneously.

Necessary and Sufficient Conditions for
OCCUR(COMPOSITE(event1, event2), t):

OCCUR(event1, t) & OCCUR(event2, t) .

All of the event classes we have considered so far have been fully specified. For example, the CHANGE-POS event class specified both the starting position and ending position of the object. It is possible to deal with less specific events in the same framework as well. For instance, consider the class of events of moving to somewhere (without specifying where from). Thus we can define this new class as follows:

Necessary and Sufficient Conditions for
OCCUR(MOVE-TO(obj, dest), T):

$\exists s$.OCCUR(CHANGE-POS(obj, s, dest), T) .

It is here that we see that axiom (O.1) is crucial to allow us to count MOVE-TO events. For example, consider a simple world with three distinguished positions, A, B, and C, and a ball that moves from A to C via B during time T. Now there are three CHANGE-POS events that OCCUR over or within T. CHANGE-POS(Ball1, A, B), CHANGE-POS(Ball1, B, C), and CHANGE-POS(Ball1, A, C). So it would appear that we could have two MOVE-TO (Ball1, C) events that OCCUR over or within T corresponding to the latter two CHANGE-POS events above. If this were the case, we should report that the ball moved to C at least twice during T, an obviously ridiculous conclusion. Axiom (O.1), which is embedded in the definition of MOVE-TO above, guarantees that only the CHANGE-POS from A to C produces a MOVE-TO C event. Thus we have the desired conclusion that the ball moved to C only once during T.

The other major subclass of occurrences is the processes. Processes differ from events in that axiom (O.1) does not hold. If a process occurs over a time interval T, it appears to occur over at least a substantial number of subintervals. For example, if I am walking during interval T, I must be walking during the first half of interval T. I could have stopped walking for a brief period within T, however, and still have been walking during T. Thus we appear to need some notion of grain of interval size for a given process, where in any subinterval larger than the grain size, the process also occurred. This is too difficult to formalize adequately at present. Thus we will make a weaker claim for processes than we might. In particular, if a process is occurring over an interval T, it must also be occurring over at least one subinterval of T. To formalize this, we introduce a new predicate, OCCURRING, for processes. We then have for all processes p, and time t:

$$\text{OCCURRING}(p, t) \Rightarrow \exists t'.\text{IN}(t', t) \,\&\, \text{OCCURRING}(p, t') . \quad \text{(O.2)}$$

Related classes of processes will be described using functions in a similar manner as with event classes. For certain classes of processes we can of course define stronger axioms than (O.2). For example, if we let FALLING(object) denote the class of processes involving the object falling, we could have an axiom that it is falling over all subintervals:

$$\text{OCCURRING}(\text{FALLING}(\text{object}), T) \Leftrightarrow (\forall t.\text{IN}(t, T) \Rightarrow \text{OCCURRING}(\text{FALLING}(\text{object}), t)) . \quad \text{(O.3)}$$

Many event classes have closely associated process classes. For example, the event of an object falling from x to y necessarily involves the process of falling. In fact, we could define the falling event as a composite of the CHANGE-POS event introduced above and a falling process. Using a definition of COMPOSITE extended to include processes, we define:

FALL(object, source, goal) =
COMPOSITE(CHANGE-POS(object, source, goal),
FALLING(object)) .

Using this definition, we can prove that

OCCUR(FALL(object, source, goal), t) $\Leftrightarrow$
OCCUR(CHANGE-POS(object, source, goal), t) &
OCCURRING(FALLING(object), t) .

Many events can be decomposed into a fairly 'neutral' event like CHANGE-POS and a process. This appears to formalize the intuition underlying many representations based on verb primitives (e.g., [12, 7]).

The relation between processes and events becomes more complicated once one considers sentences that describe processes in terms of events. For example, the sentence "John is walking from home to the store" appears to describe a process because it is in the progressive tense, yet does it in terms closer to an event. This sentence may be true even if the event does not occur: John might change his mind on the way and return home. A suggested solution to this problem is that the above sentence really means that John is walking with the intention of going to the store. This does not solve the entire problem, however, for similar sentences can be constructed for inanimate objects, as in "The ball was falling to the ground".

The above solution might be adapted by using a notion of expected outcome to subsume the agent's intention. A solution along these lines, however, is beyond the capabilities of the present formalism. Without further comment, we shall allow such sentences to be expressed by allowing events to be arguments to the OCCURRING predicate. For example, let us assume the sentence "The ball fell onto the table over T" is represented as:

OCCUR(FALL(ball, s, table1), T).

Then the sentence "The ball was falling to the table over T" would be represented by:

OCCURRING(FALL(ball, s, table1), T).

Generalizing from this example, we can see that if an event occurred, then it was occurring. In other words, for any event e and time t,

$$\text{OCCUR}(e, t) \Rightarrow \text{OCCURRING}(e, t). \tag{O.3}$$

The converse of this does not hold.

Defining necessary and sufficient conditions for many processes, especially those describing human activity, appears not to be possible. While there may be technical definitions of the differences between walking, strolling, and running, it is unlikely that they would be useful in language comprehension. Such terms appear to be primitive-like processes that may be recognized from the perceptual system. Of course, necessary conditions are likely to be found for these processes, and consequences, such as that they all involve moving, but at different rates, can be described if necessary.

Processes involving physical change, or motion (e.g., falling) may afford precise descriptions. What is necessary to describe these is a workable theory of naive physics (see [13]). Investigating these issues here will take us too far afield.

An important relationship that we shall need asserts that one event causes another. The nature of causality has been studied extensively in philosophy (e.g., [14]). Many of the issues considered there, however, will not affect this work. Let us introduce a predicate ECAUSE (event causation), where

ECAUSE($e1$, $t1$, $e2$, $t2$)

is true only if event $e1$'s occurrence at time $t1$ caused event $e2$ to occur at time $t2$. The following facts about causality are important.

If an event occurred that caused another event, then the caused event also occurred.

$$\text{OCCUR}(e, t) \,\&\, \text{ECAUSE}(e, t, e', t') \Rightarrow \text{OCCUR}(e', t'). \tag{O.4}$$

An event cannot cause events prior to its occurrence (though they may be simultaneous).

$$\text{ECAUSE}(e, t, e', t') \Rightarrow \text{IN}(t', t) \vee \text{BEFORE}(t, t') \vee \text{MEETS}(t, t') \vee \text{OVERLAPS}(t, t') \vee \text{EQUALS}(t, t'). \tag{O.5}$$

Furthermore, the ECAUSE relation is transitive, anti-symmetric, and anti-reflexive.

None of the axioms above can be used to infer a new causal relation from a set of facts involving no causal relations. Thus all inferences about causality come from other already known causal relations, or must be induced from outside the logic. This seems consistent with common treatments in philosophy and artificial intelligence in which causality is irreducible (e.g., [15, 16, 12]).

4. Defining Actions

An important subclass of occurrences are those that involve animate agents performing actions. There are actions that are processes (e.g., "John is running"), and actions that are events (e.g., "John lost his hat"). An action is an occurrence caused in a 'certain' way by the agent. This relation is not simple causality between events, for an agent may be involved in an event without it being an action of that agent. For example, consider two distinct interpretations of the sentence, "Cleo broke the window". The first describes an action of Cleo. The second arises, say, if Cleo is thrown through the window at a wild party. In this case, Cleo is the instrument of the breaking of the window. Thus not all events that are caused by animate agents are actions by that agent.

To avoid this difficulty, we introduce a new form of causality termed agentive causality or ACAUSE. An agent ACAUSES an event only in those cases where the agent caused the event in an appropriate manner for the situation to be called an action of the agent's.

Classes of actions can be characterized by the function

ACAUSE(agent, occurrence)

which for any agent and occurrence produces the action of the agent causing the occurrence. As with all other occurrences, such actions may OCCUR or be OCCURRING over a set of time intervals. Particular instances of actions can only be specified by specifying the time of the action.

We can also classify actions in the same manner as all other occurrences by

introducing a function for each related class of actions. For example, the class of actions consisting of an agent moving an object from one location to another can be generated by the function

MOVE-ACTION(agent, object, source-location, goal-location).

which can be defined as being equivalent to

ACAUSE(agent, CHANGE-POS(object, source-location, goal-location)).

It is hypothesized that every action can be characterized as an agent ACAUSEing an occurrence. For some actions, such as singing, an occurrence must be introduced that consists of the actual motions involved to produce the activity. Although such an occurrence might never occur independently of the action, introducing it preserves the simplicity of the model.

Again borrowing terminology from Mourelatos [6], we call the class of actions that consists of ACAUSEing an event to be *performances*, and those that consist of ACAUSEing a process to be *activities*.

We can capture much of the above discussion with the following axioms. If an agent ACAUSEs an occurrence over time t, then the occurrence was OCCURRING over t.

$$\text{OCCURRING}(\text{ACAUSE}(\text{agent}, \text{occurrence}), t) \Rightarrow \text{OCCURRING}(\text{occurrence}, t). \quad \text{(A.1)}$$

For every action there is a unique agent and a unique occurrence that the agent ACAUSEs which constitutes the action.

$$\forall \text{action}\ \exists!\ \text{agent, occurrence}\ \ \text{action} = \text{ACAUSE}(\text{agent}, \text{occurrence}). \quad \text{(A.2)}$$

For the subclass of performances, we have a stronger version of (A.1) with the OCCUR predicate.

$$\text{OCCUR}(\text{ACAUSE}(\text{agent}, \text{event}), t) \Rightarrow \text{OCCUR}(\text{event}, t). \quad \text{(A.3)}$$

The other important aspects of the ACAUSE relation remain to be considered in the section on intentional action. But first, let us reconsider the individuation of actions.

We have seen simple performances and activities as examples of actions. Using the constructors for composite events we can describe actions that consist of a sequence of actions, or consist of actions being performed simultaneously. Note that a composite action composed of a performance and an activity (e.g., "walking to the store while juggling three balls") is itself a performance. This is easily seen from the observation that we can count the number of occurrences of such composite actions. The only composites that are activities are those that consist entirely of activities.

There are situations which might appear to be simple composite actions, yet, on closer examination, have considerably richer structure. The composite actions we have seen so far consist of actions that can be considered independently of each other; neither is necessary for the success of the other. Thus walking to the store while juggling three balls consists of walking to the store, which could have been done independently, and juggling three balls, which also could have been done independently. Many other composites have subactions that are related in a considerably stronger fashion. Consider the actions performed in the situation described as "Sam hid his coat by standing in front of it".

Taking the position outlined in the introduction, there are at least two distinct actions performed here; namely, "hiding the coat" and "standing in front of the coat". These actions, however, are not independent of each other. They are intimately related, as one was performed by means of performing the other, i.e., the coat was hidden by means of standing in front of it. Note that this is not simply a causal relationship: standing in front of the coat didn't cause John's hiding the coat, it actually constituted the hiding.

A wide range of similar examples exists in the speech act literature (e.g., [17]). For example, I may perform a promise by telling you that I will come to your party, provided I have the appropriate intentions. Again, the act of speaking did not simply cause the promise act, but, in conjunction with the appropriate intentions, it constituted the promise act.

Goldman [8] terms this relationship between actions as generation. An act A *generates* an act B iff:

(i) A and B are cotemporal (they occur at the same time);

(ii) A is not part of doing B (such as "playing a C note" is part of "playing a C triad" on a piano);

(iii) A occurs in a context C, where the occurrence of A and C jointly imply the occurrence of B.

Goldman distinguishes different types of generation, depending on the nature of the context C. I have not found this a useful division as most examples seem to be combinations of the different types. He identifies three major components of C:

- causal laws: A generates B because the occurrence of A causes the occurrence of B;
- conventional rules: for example, "signaling a left turn on your bike" is generated by "putting your arm out";

- simple (or definitional): A generates B simply by the fact that B is defined as doing A in a certain context; an example of this, namely hiding an object from someone, will be discussed in detail below.

To continue the formal development, let us introduce a predicate GENERATES that takes two actions and a time:

GENERATES($a1$, $a2$, t).

This predicate is true only if action $a1$ generates action $a2$ during time t. GENERATES is transitive, anti-symmetric, and anti-reflexive with respect to its two action arguments.

For example, consider an agent JOHN playing a C triad on the piano. Other actions that JOHN performs simultaneously with this include playing a C note (a part of the first action) and waking up Sue (generated by the first action). We can express this in the formal notation using the following functions. Let:

- PLAY-C-TRIAD(agent, piano) be the action of the agent playing a C triad on the piano;
- PLAY-C(agent, piano) be the action of the agent playing a C note on the piano; and
- WAKE(agent, other) be the action of the agent waking up the other person.

Then the situation above is captured by

OCCUR(PLAY-C-TRIAD(JOHN, P), $T1$) &
OCCUR(PLAY-C(JOHN, P), $T1$) &
OCCUR(WAKE(JOHN, SUE), $T1$)

where

GENERATES(PLAY-C-TRIAD(JOHN, P),
WAKE(JOHN, SUE), $T1$)

and ∃-event

PLAY-C-TRIAD(JOHN, P) =
COMPOSITE(PLAY-C(JOHN, P), event).

The notion of generation is crucial for considering how an action was performed, or how to perform an action (i.e., planning). Investigating these issues will take us too far afield, but it is worth briefly considering how planning knowledge and definitional knowledge interact. We have seen two major classes of knowledge about actions. One, the definitional, outlines necessary (and sometimes sufficient) conditions for an action's occurrence. This is crucial for achieving a minimal understanding of what is implied by a sentence describing an action. The second, generational knowledge, outlines how actions can be performed, and is crucial for problem solving. But a sharp distinction in the uses of this knowledge is artificial. Generational knowledge can be used in understanding to infer plausible ways in which an action might have been accomplished, whereas definitional knowledge can be used by a problem solver to identify what actions might be appropriate to solve a certain task that has not been encountered previously.

5. Intentional Action

Before developing the notion of intentional action, let us consider an example that motivates the remainder of the paper, namely, defining the action of hiding an object from someone. We shall not consider the sense of hiding that is equivalent to simply concealing from sight accidentally. The sense here is the sense that arises from an accusation such as "You hid that book from me!" As we shall see, this is an intentional action.

The definition of hiding an object should be independent of any method by which the action was performed, for, depending on the context, the actor could hide the object in many different ways. In other words, the action can be generated in different ways. For instance, the actor could:

- put the object behind a desk;
- stand between the object and the other agent while they are in the same room; or
- call a friend Y and get her or him to do one of the above.

Furthermore, the actor might hide the object by simply not doing something s/he intended to do. For example, assume Sam is planning to go to lunch with Carole after picking Carole up at her office. If, on the way out of his office, Sam decides not to take his coat because he doesn't want Carole to see it, then Sam has hidden the coat from Carole. Of course, it is crucial here that Sam believed that he normally would have taken the coat. Sam couldn't have hidden his coat by forgetting to bring it.

This example brings up a few key points that may not be noticed from the first three examples. First, Sam must have intended that Carole not see the coat. Without this intention (i.e., in the forgetting case), no such action occurs. Second, Sam must have believed that it was likely that Carole would see the coat in the future course of events. Third, Sam must have decided to act in such a way that he then believed that Carole would not see the coat in the future course of events. Finally, for the act to be successful, Sam must have acted in that way. In this case, the action Sam performed was "not bringing the coat", which would normally not be considered an action unless it was intentional.

I claim that these four conditions provide a reasonably accurate definition of what it means to hide something. They certainly cover the four examples presented above. It is also important to note that one does not have to be successful in order to have been hiding something. The definition depends on what the hider believes and intends at the time, not what actually occurs.

However, the present definition is rather unsatisfactory, as many extremely difficult concepts, such as belief and intention, were thrown about casually. We shall investigate these issues in the next sections.

5.1. Belief and plans

There is much recent work on models of belief (e.g., [18, 19]). I will use a sentential model of belief based on the model of Haas [20], which is similar to that of Konolidge [21] and Kaplan [22]. In their work, belief is a predicate taking an agent and a description of a proposition as its arguments and is intended to mean that the agent believes that proposition. In computer models, it means that the agent has a data structure in its memory corresponding to the proposition. To develop this model requires a consistent method of introducing quotation into the predicate calculus so that the usual paradoxes are avoided. I will not develop this fully here as it is not necessary for this paper, but the interested reader should see [23, 20].

An important thing to notice, though, is that there must be two relevant time indices to each belief; namely, the time over which the belief is held, and the time over which the proposition is believed to hold. For example, I might believe *today* that it rained *last weekend*. This point will be crucial in modeling the action of hiding. To introduce some notation, let

"A believes (during T_b) that p holds (during T_p)"

be expressed as

HOLDS(believes(A, "HOLDS(p, T_p)"), T_b)

and which we shall abbreviate using

BELIEVES(A, p, T_p, T_b).

The quotation of formulas in this development must be viewed only as a notational convenience. A more elaborate system of quotation is required so that variables can be introduced into quoted expressions. In such cases, the variables range over names of terms. I will avoid these issues and simply allow variables within quotation marks. Once we have the capability of describing propositions in the logic, we can specify a wide range of beliefs using quantification over parts of formulas. Proof methods can be developed that allow the simulation of other agents' reasoning. For a rigorous treatment of these issues, see [20].

Plans are typically characterized in AI as a sequence or partial order of actions (e.g., [24, 25]). This characterization is adequate when modeling static worlds where the only change is by the agent that constructs the plan. In more general settings, however, plans may involve actions by other agents as well as naturally occurring events. In addition, plans are not made starting from a blank world. There are many external events occurring, and the agent believes that certain other events may occur in the future. This even includes actions that the agent has already decided to do in the future. Thus a more general notion of planning is called for.

Let each *agent* maintain three partial descriptions of the world. Each description is a set of propositions asserting properties and occurrences over time intervals. The first description is called the *expected* world, the second the *planned* world, and the third the *desired* world. The expected world is what the agent believes will happen given that certain known future events will occur and assuming that the agent does nothing. It is a view of the world assuming all agents are as lazy as possible but remain in accordance with the known future events. The expected world obeys a generalized law of momentum. Things in the process of changing continue to change unless prevented, and everything else remains the same unless disturbed.

The desired world contains a description of the properties and occurrences that the agent desires. The planned world is the same as the expected world except that the agent may add or remove actions (by that agent). Thus it is a simulation of what the world would be like if the agent acted differently than what is expected. The goal, of course, is to make the planned world subsume the desired world and then act according to it. A possible algorithm for this would be a generalized GPS model of finding differences between the worlds and introducing actions to reduce the differences. There is not space here to consider the planning algorithm further, but for an initial attempt at building a planner, see [26].

A plan is a set of decisions about performing or not performing actions by the agent. Thus, at any time, there is a plan that specifies the mapping from the expected world to the planned world. A plan can be constructed without the agent intending to actually act in accordance with the plan. Such plans arise from a wide range of activities, from abstract problem solving to the recognition of plans of other agents. It will be important for us here to have a notion of an agent *committing* to a plan. An agent is committed to a plan over a certain time interval if that agent believes he or she will act in accordance with the plan. That is, the planned world becomes the expected world. This then becomes part of the agent's predictions about the future state of the world, and must be considered in any further planning activity done by the agent. For instance, in the hiding example, when Sam decided not to take the coat, he changed an action he had committed to do in order to hide his coat.

Most models of plans in the literature consider a subclass of the set of plans allowed here. In particular, they only allow planning into the immediate future, and with only a few exceptions (e.g., [27]), do not allow occurrences other than

actions by the planner. We can express plans which contain occurrences independent of the planner, including actions by other agents, as well as plans not to do some action.

Let us introduce a little notation for use later on. Let the predication

TO-DO(action, time, plan)

be true if the plan specified includes performing the action at the given time, and let

NOT-TO-DO(action, time, plan)

be true if the plan specified includes not performing the action at the given time. This is of course much stronger than asserting ~TO-DO(action, time, plan), which simply asserts that the plan does not contain the indicated action.

The notion of a goal is not part of a plan directly. Rather, goals are part of the desired world and are the usual reason for committing to a plan. Let

IS-GOAL-OF(agent, goal, gtime, t)

be true if the agent's desired world at time t contains the specified goal holding during gtime. Finally, the predicate

COMMITTED(agent, plan, ctime)

is true if the agent is committed over ctime to act in accordance with the specified plan. Being committed to a plan means that the agent believes he or she will perform all actions (by the agent) in the plan.

One can think of a plan as a complex action, and it makes sense to talk of a plan occurring if all its decisions are carried out. This can be expressed by extending the OCCUR predicate to plans according to the following definition:

OCCUR(plan, t)⇔
  ∀action, t_a.TO-DO(action, t_a, plan)⇒
    OCCUR(action, t_a)&
  ∀action, t_a.NOT-TO-DO(action, t_a, plan)⇒
    (∀t.IN(t, t_a)⇒
      ~OCCUR(action, t_a)).

5.2. Intending

There are two senses of intention that are traditionally distinguished in the literature. The first has been termed *prior intention* (e.g. [28]) and arises in sentences such as

"Jim intends to run a mile today".

This intention is prior to the action, and can hold even if the action is never performed (i.e., Jim forgets or changes his mind). We shall model this form of intention simply by asserting that Jim has committed to a plan that has the action as a step. Thus

INTEND(agent, ACAUSE(agent, occurrence), atime, itime)⇔ (I.1)
  (∃plan COMMITTED(agent, plan, time)&
    TO-DO(ACAUSE(agent, occurrence), atime, plan)).

Note that Jim having this intention implies that he believes he will run the mile today. It does not imply that Jim wants to do the action, although in most cases this is a plausible inference. Also plausible is that Jim has some goal that results from this action being performed. The actual nature of this goal is uncertain. For instance, he may want to stay fit (and so he might change his mind and swim instead), or he may want to win a bet that he couldn't run a mile (and so swimming would not be a reasonable alternative). The example does not specify this information.

Many times an intention of an agent is so general that it is expressed in terms of a goal rather than an action. Examples are the actions of "achieving state x", "preventing event E", etc. To express these, we will simply use the IS-GOAL-OF predication introduced earlier.

The last sense of intention is that of intentional action, which arises in sentences such as

"Jack intentionally coughed".

This example is closely related to the notion of prior intention: it appears that Jack coughed as a result of a plan that he was committed to that involved him coughing. This is essentially the treatment suggested in Miller et al. [29].

But one must be careful with this definition. For instance, if one intentionally does an action A, does one intentionally do all the subparts of A? What about the actions one did to generate A, or the actions that A generates? For instance, if A was the action of Jack intentionally playing a C chord on the piano, did he also

(i) intentionally play a C note (a subpart of A);
(ii) intentionally move his finger (generating (i)); or
(iii) intentionally annoy his neighbors (generated by A).

One might say no to (i) and (ii), or allow subparts and not generation, or allow all as intentional. Each side has been proposed; which side we should

take depends on what is counted to be in a plan. For instance, if a composite action is in a plan, are its subparts in the plan? If so, then (i) is always intentional. If an action is in a plan, the actions that it generates may or may not be in the plan, depending on the knowledge and goals of the actor. Thus, with (iii), it might go either way, depending on why Jack was playing the piano. This last case shows we cannot simply define how intentionality relates to generation. Some actions which are generated could not be denied, however. If Sam intentionally aims and fires a gun at Sue, knowing it is loaded, he intentionally shot her. This appears to be so because there is no other plausible plan that Sam could have had. Thus, if we assume Sam is a rational being, and does not act at random, we can assume his plan must have been to shoot Sue. Thus, we will only get into difficulty with the plan-based model of intentionality if we make hard and fast rules, such as that all subparts of an action must be in a plan, or all generated actions of an action must be in a plan. If the contents of a plan is left up to plausible reasoning about the motivation of an agent's behavior, the plan model appears to provide a reasonable definition of intentionality.

There are remaining problems with the plan model, however. Davis [30] gives an example of a person driving a car when a small child runs in front of it. He claims the person intentionally slams on the brakes yet has no time to form a plan to do so. This difficulty may arise from an inadequate model of plan-directed behavior. For the present, however, these examples will not cause us any problems. Searle [28] presents other difficulties that arise with this simple model, but again, the problems can be ignored at present, as the examples that present them are fairly bizarre.

On the basis of this discussion, and following Goldman [8], we can say that an agent S intentionally performed an action A at time t iff

(i) S performed A at t;
(ii) A was part of a plan that S was committed to at time t; and
(iii) S performed A because of S's knowledge of (ii).

Introducing a new predicate, we can easily capture the first two conditions above, but it is not clear how to formalize condition (iii). For computational models in which an agent only acts because of an existing plan, however, this should not present any difficulties. Thus we can capture the first two conditions with:

$$\begin{aligned}&\text{INTENTIONAL}(\text{ACAUSE}(\text{agent}, \text{occurrence}), \text{time}) \Rightarrow \\ &\quad \text{OCCURS}(\text{ACAUSE}(\text{agent}, \text{occurrence}), \text{time}) \,\&\, \\ &\quad \text{INTEND}(\text{agent}, \text{ACAUSE}(\text{agent}, \text{occurrence}), \text{time}, \text{time}) .\end{aligned} \tag{1.2}$$

Finally, let us return to questions about the nature of the ACAUSE relation. The examples in which we have used it have all been intentional actions, so the question arises as to whether it is possible to have an ACAUSE relation to an unintentional action? For instance, if John broke the window unintentionally, did John perform an action? That, he certainly did, but the action he performed might not be breaking the window, it may have been hitting the baseball (which broke the window). If we claim that, even in this case, John performed the action of breaking the window, then we can make the example more complicated. What if John hit the baseball, which landed on the roof of a house, and a few minutes later rolled off and broke the window. Obviously, in this example, the delay and causal chain of events soon gets complicated enough that we would say John did not break the window. So where do actions stop and mere events caused by actions begin?

There seems no easy answer to this question, although a fruitful approach could be to consider the issue of responsibility. If an agent acts in a way that causes some effect which, while unintentional, should have been foreseen, we tend to term that as an action. We do not have the time to pursue this here, so make the simplifying assumption that all actions are intentional, i.e.,

$$\begin{aligned}&\text{OCCUR}(\text{ACAUSE}(\text{agent}, \text{occurrence}), t) \Rightarrow \\ &\quad \text{INTENTIONAL}(\text{agent}, \text{ACAUSE}(\text{agent}, \text{occurrence}), t) .\end{aligned} \tag{1.3}$$

In the unintentional case of "John broke the window", we analyze that as John did something intentionally that caused the window to break. This may seem to complicate the analysis of such sentences, but it can be handled in a relatively clean manner. The meaning of a sentence such as "John broke the window" could be

$$\begin{aligned}&\exists e, t1, t2.\text{OCCUR}(\text{ACAUSE}(\text{John}, e), t1) \,\&\, \\ &\quad \text{OCCUR}(\text{BREAK-EVENT}(\text{Window}), t2) \,\&\, \\ &\quad ((e = \text{BREAK-EVENT}(\text{Window})) \vee \\ &\quad\ \text{ECAUSE}(e, t1, \text{BREAK-EVEN}(\text{Window}), t2)) .\end{aligned}$$

The disjunction captures the ambiguity as to whether the breaking of the window was intentional or not. If e = BREAK-EVENT(Window), then the event that John ACAUSEd was the breaking of the window. If e caused BREAK-EVENT(Window), then the event John ACAUSEd was something else which ECAUSEd the breaking of the window. Finally, if John intentionally broke the window by hitting the baseball, then he performed two actions (intentionally) which are related by the GENERATEs relation.

6. How to Hide Revisited

With these tools, we can attempt a more precise definition of hiding. We first define the function

HIDE(agent, observer, object)

to generate the class of hiding actions. Let us also introduce an event function

SEE(agent, object),

which generates events of an agent seeing an object, and a property SEEN (agent, object) defined by

HOLDS(SEEN(agent, object), t) ⇔
  ∃t1.BEFORE(t1, t) &
  OCCUR(SEE(agent, object), t1).

So the necessary and sufficient conditions for Sam to hide the coat from Carole over interval T_h are as follows. He must have initially believed (during T_h) that Carole would have seen the coat during T_h:

∃t.IN(t, T_h) & STARTS(T_b1, T_h) &
  BELIEVES(Sam, SEEN(CAROLE, COAT), t, T_b1). (1)

He must have had an intention (during T_h) that Carole not see the coat:

IS-GOAL-OF(Sam, not(SEEN(CAROLE, COAT)), T_h, T_h) (2)

Restating conditions (1) and (2) in terms of Sam's plan during T_b1, we see that in his expected world Carole will see the coat, while in the desired world, she will not.

The next conditions describe Sam as he formulates a new plan which achieves his goal of Carole not seeing the coat. To describe this we introduce two new event classes, first the event of an agent committing to a plan, and second the event of an agent changing his or her mind about something. Let

COMMIT(agent, plan)

denote the class of events defined by

Necessary and Sufficient Conditions for
OCCUR(COMMIT(agent, plan), time):

∃t1, t2.MEETS(t1, time) &
  MEETS(time, t2) &
  ~ COMMITTED(agent, plan, t1) &
  COMMITTED(agent, plan, t2).

Furthermore, let us define the event class

CHANGE-MIND(agent, property, ptime)

by

Necessary and Sufficient Conditions for
OCCUR(CHANGE-MIND(agent, property, ptime), time):

∃t1, t2.MEETS(t1, time) &
  MEETS(time, t2) &
  BELIEVES(agent, property, ptime, t1) &
  BELIEVES(agent, not(property), ptime, t2).

Using these events, we can state that Sam is adopting a plan with the goal that Carole not see the coat and that he believes it will work:

∃plan, T_b2.MEETS(T_b1, T_b2) & (3)
  OCCUR(COMMIT(Sam, plan), T_b2) &
  ECAUSE(COMMIT(Sam, plan), T_b2,
    CHANGE-MIND(Sam, SEEN(CAROLE, COAT), T_h), T_b2).

These three conditions capture Sam's intention to hide the coat, and if Sam acts in accordance with the plan hypothesized in condition (3), the hide action is performed.

We can put these conditions all together into one definition as follows:

Necessary and Sufficient Conditons for
OCCUR(HIDE(agent, observer, object), T_h):

∃t, T_b1, T_b2, plan.
  STARTS(T_b1, T_h) &
  MEETS(T_b1, T_b2) &
  IN(t, T_h) &
  BELIEVE(agent, SEEN(observer, object), t, T_b1) &
  IS-GOAL-OF(agent, not(SEEN(observer, object)), T_h, T_h) &
  OCCUR(COMMIT(Sam, plan), T_b2) &
  ECAUSE(COMMITS(Sam, plan), T_b2,
    CHANGE-MIND(Sam, SEEN(CAROLE, COAT), T_h), T_b2) &
  OCCUR(plan, T_h).

The conditions that the agent changed his mind can be derived from the above conditions and the definition of ECAUSE.

One can see that much of what it means to hide is captured by the above. In particular, the following can be extracted directly from the definition:

- if you hide something, you intended it not to be seen (and thus can be held responsible for the consequences of this);
- you cannot hide something if you believed it was not possible that it could be seen, or if it were certain that it would be seen anyway;

– one cannot hide something simply by changing one's mind about whether it will be seen.

In addition, there are many other possibilities related to the temporal order of events. For instance, you can't hide something by performing an action after the hiding is supposed to be done.

7. Conclusion

In the introduction, three problems in representing actions were discussed. These problems have been addressed throughout the paper, but sometimes only implicitly. Let us reconsider each problem. The first problem concerned actions that involve non-activity, such as standing still. These can be modeled with no difficulty. An action class can be defined so that the agent remains in one position over the time of the action's occurrence. Note that such a non-activity must be intentional if it is to qualify as an action in this framework. Otherwise, such non-activity can only be modeled as an event. A more complicated form of non-activity involves not doing an action that was previously expected. These actions can be defined in terms of the beliefs and intentions of the agent. In particular, the agent must have been intending to do the action and later changed his or her mind.

The second problem concerned actions that cannot be defined by decomposition into subactions. The example of "hiding a book from Sue" is a prime example from this class. Any particular instance of hiding a book can be decomposed into a particular set of subactions, but there is no decomposition, or set of decompositions, that defines the class of hiding actions. Rather, hiding can only be defined in terms of the agent's beliefs and intentions. The speech acts also fall into this class. Each occurrence of a speech act is partially decomposable into the act of uttering something, but otherwise depends crucially on the speaker's intentions.

The third problem concerned actions that occur simultaneously and possibly interact. Simultaneous actions can be described directly since the temporal aspects of a plan are separated from the causal aspects. This enables us to describe situations where actions may interact with each other. Building a system that can reason about such interactions while problem solving, however, remains a difficult problem.

This framework is currently being used to study general problem-solving behavior, as well as the problem-solving behavior that arises in task-oriented dialogues. A simple problem solver has been built using this framework and is described by Allen and Koomen [26]. The model is also being used both for plan recognition and plan generation in a system under development at Rochester that comprehends and participates in task-oriented dialogues. The action models are being used to describe a useful set of conversational actions which include the traditional notion of speech acts.

Appendix A. Proof that (H.2) Entails (H.1)

$$\text{HOLDS}(p, T) \Leftrightarrow (\forall t.\text{IN}(t, T) \Rightarrow \exists s.\text{IN}(s, t) \wedge \text{HOLDS}(p, s)). \quad \text{(H.2)}$$

$$\text{HOLDS}(p, T) \Leftrightarrow (\forall t.\text{IN}(t, T) \Rightarrow \text{HOLDS}(p, T)). \quad \text{(H.1)}$$

We assume (H.2) as the definition, and prove (H.1) one direction at a time. The only assumptions we need about the IN relation is that it is transitive, and for every interval I, there exists an interval J such that IN(J, I).

Proof of $\text{HOLDS}(p, T) \Rightarrow (\forall t.\text{IN}(t, T) \Rightarrow \text{HOLDS}(p, T))$:

(1)	HOLDS(p, T)	hypothesis;
(2)	$\forall t.\text{IN}(t, T) \Rightarrow (\exists s.\text{IN}(s, t) \,\&\, \text{HOLDS}(p, s))$	by defn.(H.2), (1);
(3)	IN$(T1, T)$	assumption;
(4)	IN$(T2, T1)$	assumption;
(5)	IN$(T2, T)$	transitivity of IN using (3), (4);
(6)	$\exists s.\text{IN}(s, T2) \,\&\, \text{HOLDS}(p, s)$	MP (2), (5);
(7)	$\forall t'.\text{IN}(t', T1) \Rightarrow (\exists s.\text{IN}(s, t') \,\&\, \text{HOLDS}(p, s))$	discharging assumption (4);
(8)	HOLDS$(p, T1)$	by defn.(H.2), (7);
(9)	$\forall t.\text{IN}(t, T) \Rightarrow \text{HOLDS}(p, T)$	discharging assumption (3).

Proof of $(\forall t.\text{IN}(t, T) \Rightarrow \text{HOLDS}(p, t)) \Rightarrow \text{HOLDS}(p, T)$:

(1)	$\forall t.\text{IN}(t, T) \Rightarrow \text{HOLDS}(p, t)$	hypothesis;
(2)	IN$(T1, T)$	assumption;
(3)	HOLDS$(p, T1)$	MP (1), (2);
(4)	$\forall t.\text{IN}(t', T1) \supset \exists s'.\text{IN}(s', t') \,\&\, \text{HOLDS}(p, s')$	by defn.(H.2), (3);
(5)	$\exists s''.\text{IN}(s'', T1)$	axiom
(6)	IN$(T2, T1)$	existential elim., (5);
(7)	$\exists s'$, IN$(s', T2) \,\&\, \text{HOLDS}(p, s')$	MP (4), (6);
(8)	IN$(T3, T2) \,\&\, \text{HOLDS}(p, T3)$	existential elim., (7);
(9)	HOLDS$(p, T3)$	conj. elim., (8);
(10)	IN$(T3, T1)$	transitivity of IN, (6), (8);
(11)	IN$(T3, T1) \,\&\, \text{HOLDS}(p, T1)$	conj. intro.(9), (10);
(12)	$\exists s'.\text{IN}(s', T1) \,\&\, \text{HOLDS}(p, s')$	existential intro., (11);
(13)	$\exists s'.\text{IN}(s', T1) \,\&\, \text{HOLDS}(p, s')$	existential intro., [12];
(14)	$\forall t.\text{IN}(t, T) \supset \exists s'.\text{IN}(s', t) \,\&\, \text{HOLDS}(p, s')$	discharging assumption (2);
(15)	HOLDS(p, T)	by defn.(H.2), (14).

ACKNOWLEDGMENT

The author wishes to thank Henry Kautz for his detailed criticism of the penultimate version of this paper that forced the clarification of several murky areas. I would also like to thank Jerry Feldman, Alan Frisch, Andy Haas, Margery Lucas, Dan Russell, and Stuart Goldkind for many enlightening comments and improvements on previous versions of this paper, and Drew McDermott and Pat Hayes for discussions on general issues in representing action and time.

REFERENCES

1. Charniak, E., A common representation for problem-solving and language-comprehension information, *Artificial Intelligence* **16** (1981) 225–255.
2. Allen, J.F. and Perrault, C.R., Analyzing intention in utterances, *Artificial Intelligence* **15** (1980) 143–178.
3. Fillmore, C.J., The case for case, in: Bach and Harms (Eds), *Universals in Linguistic Theory* (Holt, Rinehart and Winston, New York, 1968).
4. McCarthy, J. and Hayes, P.J., Some philosophical problems from the standpoint of artificial intelligence, in: B. Meltzer and D. Michie (Eds), *Machine Intelligence* 4 (Edinburgh University Press, Edinburgh, 1969).
5. McDermott, D., A temporal logic for reasoning about processes and plans, RR 196, Computer Science Dept., Yale Univ., New Haven, CT, 1981; also in *Cognitive Sci.* 6(2) (1982).
6. Mourelatos, A.P.D., Events, processes, and states, *Linguistics and Philosophy* 2 (1978) 415–434.
7. Jackendoff, R., Toward an explanatory semantic representation, *Linguistic Inquiry* 7(1) (1976) 89–150.
8. Goldman, A., *A Theory of Human Action* (Princeton University Press, Princeton, NJ, 1970).
9. Davidson, D., The logical form of action sentences, in: N. Rescher (Ed.), *The Logic of Decision and Action* (University Pittsburgh Press, Pittsburgh, PA, 1967).
10. Allen, J.F., Maintaining knowledge about temporal intervals, TR 86, Computer Science Dept., Univ. Rochester, January 1981; also in *Comm. ACM* **26** (1983) 832–843.
11. Mays, E., A modal temporal logic for reasoning about change, *Proc. 21st Meeting, Association for Computational Linguistics* (MIT, Cambridge, CA, 1983).
12. Schank, R.C., *Conceptual Information Processing* (North-Holland, New York, 1975).
13. Hayes, P.J., Naive physics I: Ontology for liquids, Working Paper 63, Institut pour les Etudes Semantiques et Cognitives, Geneva, 1978.
14. Sosa, E. (Ed.), *Causation and Conditionals* (Oxford University Press, Oxford, 1975).
15. Taylor, R., *Action and Purpose* (Prentice-Hall, Englewood Cliffs, NJ, 1966).
16. Norman, D.A. and Rumelhart, D.E., *Explorations in Cognition* (Freeman, San Francisco, CA, 1975).
17. Searle, J.R., *Speech Acts: An Essay in the Philosophy of Language* (Cambridge University Press, London, 1969).
18. Cohen, P.R., On knowing what to say: Planning speech acts, TR 118, Computer Science Dept., Univ. of Toronto, 1978.
19. Moore, R.C., Reasoning about knowledge and action, Ph.D. Thesis, MIT, Cambridge, MA, 1979.
20. Haas, A., Planning mental actions, TR 106 and Ph.D. Thesis, Computer Science Dept., Univ. of Rochester, Rochester, NY, 1982.
21. Konolidge, K., A first-order formalization of knowledge and action for a multiagent planning system, TN 232, AI Center, SRI International, Menlo Park, CA, 1980.
22. Kaplan, D., Quantifying in, *Synthese* **19** (1968) 178–214.
23. Perlis, D., Language, computation, and reality, TR 95 and Ph.D. Thesis, Computer Science Dept., Univ. of Rochester, Rochester, NY, 1981.
24. Fikes, R.E. and Nilsson, N.J., STRIPS: A new approach to the application of theorem proving to problem solving, *Artificial Intelligence* 2 (1971) 189–205.
25. Sacerdoti, E.D., The nonlinear nature of plans, *Proc. 4th IJCAI*, Tbilisi, USSR, 1975.
26. Allen, J.F. and Koomen, J.A., Planning using a temporal world model, *Proc. 8th IJCAI*, Karlsruhe, West Germany, 1983.
27. Vere, S., Planning in time: Windows and durations for activities and goals, Jet Propulsion Laboratory, California Institute of Technology, 1981.
28. Searle, J.R., The intentionality of intention and action, *Cognitive Sci.* 4(1) (1980).
29. Miller, G.A., Galanter, E. and Pribram, K.H., *Plans and the Structure of Behavior* (Holt, Rinehart and Winston, New York, 1960).
30. Davis, L.K., *Theory of Action* (Prentice-Hall, Englewood Cliffs, NJ, 1979).

Received March 1982: revised version received October 1983