MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 1238 November, 1990

Transition Space

Gary C. Borchardt

Abstract

Informal causal descriptions of physical systems abound in sources such as encyclopedias, reports and user's manuals. Yet these descriptions remain largely opaque to computer processing. This paper proposes a representational framework in which such descriptions are viewed as providing partial specifications of paths in a space of possible transitions, or *transition space*. In this framework, the task of comprehending informal causal descriptions emerges as one of completing the specifications of paths in transition space—filling causal gaps and relating accounts of activity varied by analogy and abstraction. Drawing on results from perceptual psychology, a specific representation for events is outlined, along with a set of operations for use in comprehending and answering questions regarding informal causal descriptions. The use of the representation and its operations is illustrated in the context of a simple description concerning rocket propulsion.

Copyright © Massachusetts Institute of Technology, 1990

This report describes research done 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 Projects Agency of the Department of Defense under Office of Naval Research contract N00014-85-K-0124.

1 Introduction

Informal causal descriptions of the sort appearing in encyclopedias, reports and user's manuals comprise a rich yet elusive source of knowledge about the behavior of physical systems. Particular aspects of these descriptions which frustrate automated comprehension are: their incomplete nature, such that the comprehender must fill in missing components; incorporation of metaphorical references; varying levels of abstraction; and inclusion of conceptualizations which have no set physical extension in the world (e.g., spaces, paths, collections, events and so forth). In spite of these difficulties, there are a number of important applications which rely on an ability to accept, apply and recount informal causal descriptions: intelligent on-line user's manuals and encyclopedias, tutoring programs, literature search software, knowledge acquisition and explanation facilities for expert systems, intelligent user interfaces, text summarization software, intelligent service directories, and general causal reasoning systems that must operate in domains of extreme complexity (e.g., transportation or manufacturing) or domains which critically involve human conceptualizations (e.g., human-factors engineering or remote-sensing data interpretation).

There are two general categories of research relevant to the task of understanding informal causal descriptions of physical systems. The first category concerns the utilization of somewhat more formal causal descriptions of physical systems: for example, the research in naive physics [Hayes, 1985a; Hayes, 1985b], qualitative physics [Forbus, 1984; De Kleer and Brown, 1984; Kuipers, 1986], model-based reasoning [Davis, 1984; Davis and Hamscher, 1988] and temporal reasoning [McDermott, 1982; Allen, 1984; Shoham and Goyal, 1988]. Three factors make it difficult to transfer the techniques in this body of work to application on informal causal descriptions. First, these approaches typically utilize a representational ontology deriving from a technical perspective, including such quantities as mass, density, momentum and so forth, whereas we need an ontology deriving from everyday language, excluding many technical quantities but including a number of human conceptualizations such as spaces, paths, collections, events, systems, and attributes such as "part-of," origin, ownership, function and value.[1] Second, these approaches focus primarily on causal reasoning, whereas we require the additional use of extended modes of reasoning such as analogy and abstraction.[2] Finally, these approaches generally adhere to a philosophical or scientific characterization of causation as a necessitating force in nature—if A occurs, then B must occur—

[1]This difference does not apply to all of the approaches listed above. In particular, Hayes' naive physics [Hayes, 1985a; Hayes, 1985b] does target a number of human conceptualizations.

[2]The use of abstraction does appear in the work on model-based reasoning and in some of the work on qualitative physics.

leading to a model in which individual assertions and groups of assertions in a causal explanation support one another via logical implication. For the purposes of processing informal causal descriptions, we need more of a psychological account of causation as a non-necessitating, heuristic device used in the construction of explanations—if A occurs, a number of tentative expectations may be raised, to be evaluated according to their participation in an overall explanation of the concerned phenomena—this leads to a model in which strict, logical implication plays a diminished role, and each assertion is dependent on the entire explanatory structure for support.

Turning to the second category of related research—work tending generally toward a cognitive orientation—we find an abundance of related work in the areas of mental models [Gentner and Stevens, 1983; Johnson-Laird, 1983], analogy [Winston, 1982; Gentner, 1983; Holland et al., 1986], case-based reasoning [Kolodner et al., 1985; Hammond, 1986], explanation-based learning [Mitchell et al., 1986; DeJong and Mooney, 1986], natural language understanding [Grosz, 1977; Schank and Riesbeck, 1981; Lehnert, 1981; Dyer, 1983; Kolodner, 1983; Alterman, 1985; Norvig, 1989] and semantics [Jackendoff, 1983; Talmy, 1988; Lakoff and Johnson, 1980; Levin, 1985]. While these approaches generally advance the sort of ontology, modes of reasoning and view of explanation required for the interpretation of informal causal descriptions, there is also a tradeoff: these approaches typically lack the degree of detail in the representation of time and change that appears in the more technically-motivated literature. In the cognitively-oriented literature, events are typically represented by the equivalent of atomic formulae (e.g., "COLLIDE(block-1, block-2)"), possibly accompanied by specifications of roles for the participants and explicit associations with other events, but excluding an account of the actual *changes* taking place (e.g., appearance of contact, changes in speed, heading, etc. for "collide"). In the absence of change-level information, it is difficult to reason about components of events (e.g., "striking," "approaching," "applying pressure," "departing" "changing direction" and others all qualify as components of the event "bouncing") or to recognize similarities or analogies between events (e.g., "swinging" and "turning" are quite similar, while "flying" and "slowing down" are not). Additionally, it is difficult to reason about side-effects or failures arising from extra or missing changes in the unfolding of events. Yet, these capabilities are central to understanding the similarities and differences between informal causal descriptions and the applicability of these descriptions to particular dynamic situations.

What is needed is a synthesis of these two general approaches: a cognitively-oriented framework for representation and reasoning about events and causality that is sufficiently detailed at the level of time and change. A good source of guidance in this endeavor is the literature in perceptual psychology. From this literature, several useful representational guidelines emerge, as listed below.

Representation in terms of objects and attributes. Specific classes of objects appear in everyday accounts of causal situations, and these objects are described by specific attributes. A good summary appears in [Miller and Johnson-Laird, 1976]. Objects and attributes may be either perceptual or conceptual in nature, attributes are typically unary or binary and may also be distinguished as qualitative (e.g., color or contact) or quantitative (e.g., length or pressure). Additionally, special object "prototypes" serve as standards for comparison.

Time as a sequence of moments. Investigation by Newtson et al. [1976; 1977] indicates that time is perceived as a sequence of moments, and that events are delimited by specific "breakpoints"—time points at which significant changes are perceived by the observer.

Qualitative comparisons and changes. Following from the superiority of humans at relative as opposed to absolute estimation of attributes, it is natural to represent both static comparisons and dynamic changes in a qualitative manner (see, for example, [Miller and Johnson-Laird, 1976]).

Causation as an association between changes. Experimental evidence characterizes perceived causality as an association between consecutive changes in a scene [Michotte, 1946]. Concurring accounts appear in the cognitive development literature [Leslie and Keeble, 1987; White, 1988].[3]

The final point above is significant. While a number of AI representations—generally derived from the state space model used in problem solving and planning situations—have permitted the antecedents or consequents of causality to be *states,* it would seem to be changes which are really involved.[4] An intuitive argument offers further support for this point. Regarding causal antecedents, if states are to do the causing, then one may ask why the causal effect occurs precisely when it does and not earlier; thus, one is led to suspect that some additional ingredient has fallen into place just prior to the causal effect. This final change may then be ascribed as the antecedent of causality. Regarding causal effects, if there is no change in a scene, then causation contributes nothing to the reasoning process; we can reason just as well with what we knew to be true beforehand.

[3] There is also arguably a conceptual level of causality, where rules of causal attribution are articulated and metaphorical models of the causal process appear (e.g., events connected into chains and cycles, "transfer" of energy from event to event) [Bullock et al., 1982; Lakoff and Johnson, 1980; Shultz, 1982].

[4] The state space representation is of course well suited to tasks such as problem solving and planning (given a minimal amount of causal interaction between plan steps). The contention here is that informal causal explanation is a sufficiently different enterprise as to require a different underlying representation.

If changes are the true antecedents and consequents of causality, then how is it possible to construct representations in which states serve this role? One way is to factor out ongoing processes: in the human body, a state of having low blood sugar may lead to various effects, given the ongoing circulatory and metabolic processes of the body. Another, more common technique is simply to include dynamic information in the specification of states, by associating predicates like "MOVING" with objects or by specifying instantaneous directions of change for quantities in terms of values "increasing," "steady" and "decreasing," as in the work on envisioning [De Kleer and Brown, 1984; Kuipers, 1984].

In the context of informal causal descriptions, however, these techniques are not always applicable. In particular, the technique of specifying instantaneous directions of change for quantities only works with quantitative attributes, like "temperature" or "elevation," which may be differentiated with respect to time, but not with qualitative attributes, like "contact," "support," and "inside." For example, if two objects are specified as coming into contact, with one object breaking as a result, there is no "state of coming into contact" to serve as the causal antecedent. At one instant the objects are not in contact, and this certainly does not cause the breakage, and at a subsequent instant they are in contact—but this cannot be attributed as the antecedent either, because they could have been simply resting, in contact, for a long time. It is precisely the transition from non-contact to contact which serves as the causal antecedent of the breakage. Capturing this sort of transition requires that we expand our view to the consideration of differences in attributes measured between *pairs* of time points.

Given this motivation and incorporating the representational guidelines outlined above, we may characterize changes as collections of qualitative differences in one or more attributes, as measured across pairs of time points. Let us call these characterizations of changes "transitions." Since causality is taken to be an association between two changes, or transitions, we may think of informal causal explanations as carving out paths in a space of all possible transitions. This is the motivation for the transition space representation. In this conceptualization, individual points correspond to particular transitions, informal causal descriptions are viewed as providing partial specifications of paths connecting known or given transitions with transitions to be explained, and the comprehension of these causal descriptions is seen as a process of completing the specifications of these paths.

2 Representing Transitions and Events

Given the above, general characterization of transition space, we may proceed to specify a particular representation of this space. In doing so, we are forced to make a number of commitments where the guidelines listed above underspecify an implementation. In particular, we choose a symbolic representation (rendering object types, attributes and so forth discrete entities), with the assertions comprising causal explanations represented propositionally (not weighted by a measure of probability, for example). In addition, we choose a resolution of time points which seems "natural" in the specification of events, and we describe these events in terms of objects, attributes and prototypes which are motivated in natural language descriptions of events.[5] Given these considerations, we set forth a set of five syntactic categories from which to construct the representation for transitions and events, as follows.

Objects, both perceptual and conceptual. For example: solids; quantities of liquid, gas, fire, etc.; spaces, surfaces, paths and edges; events, sequences, cycles and systems; collections and lists—ultimately, anything that may participate in an event.

Attributes, both perceptual and conceptual. For example: length, width, depth, size, weight and color; position, elevation, orientation, speed, heading, direction, distance; "insideness," pressure, contact, restraint; "is-a," "a-kind-of," "made-of," "part-of," "before" and "after," age, origin, function and value. As in everyday language, a degree of overlap appears in the set of attributes.

Time points, as needed to distinguish points of comparison within events.

Prototypes, used as standards of reference for comparison. For example: object types (solid, event, collection...), colors, substances, numbers, qualitative directions such as "up" and "forward," and a quantity "null" serving as a point of reference for the "false" or "absent" state for all attributes.

Predicates, used in assertions comparing attribute values. For qualitative attributes, the predicates EQUAL and NOT-EQUAL suffice. For quantitative attributes, these plus GREATER and NOT-GREATER suffice.

The predicates EQUAL, NOT-EQUAL, GREATER and NOT-GREATER take five arguments: an attribute for comparison, a first object and associated time point, and a second object and associated time point. For binary attributes, tuples of objects are used in the second and fourth positions.[6] These predicates may be used in assertions comparing attributes of objects at a single time point or between two time points and serve as a foundation for the entire representation of transitions and events. Three simple assertions involving these predicates are given below.

[5]The representation of events here is similar to, and in part based on, precursors described in [Waltz, 1982] and [Borchardt, 1985].

[6]Additionally, for unusual attributes such as "between," a nesting of tuples is employed.

GREATER(length, object-1, t1, object-2, t1)
NOT-EQUAL(position, <object-1, background>, t1,
    <object-1, background>, t2)
EQUAL(distance, <object-1, object-2>, t1, null, t1)

The first assertion states that "object-1" has greater length than "object-2" at time "t1." The second, that the position of "object-1" relative to "background" is different at time "t1" and time "t2" (i.e., "object-1" has moved). The third, utilizing a comparison to the "null" prototype, states that there is a (known) absence of distance between "object-1" and "object-2" at time "t1."

While the above predicates are sufficient for the representation of arbitrary, compound transitions, the following definitions provide a useful shorthand.[7] For comparisons at a single time point:

PRESENT(?attribute, ?object, ?t1) $\Longleftrightarrow$
    NOT-EQUAL(?attribute, ?object, ?t1, null, ?t1)

NOT-PRESENT(?attribute, ?object, ?t1) $\Longleftrightarrow$
    EQUAL(?attribute, ?object, ?t1, null, ?t1)

MATCH(?attribute, ?object-1, ?object-2, ?t1) $\Longleftrightarrow$
    PRESENT(?attribute, ?object-1, ?t1) AND
    PRESENT(?attribute, ?object-2, ?t1) AND
    EQUAL(?attribute, ?object-1, ?t1, ?object-2, ?t1)

NOT-MATCH(?attribute, ?object-1, ?object-2, ?t1) $\Longleftrightarrow$
    PRESENT(?attribute, ?object-1, ?t1) AND
    PRESENT(?attribute, ?object-2, ?t1) AND
    NOT-EQUAL(?attribute, ?object-1, ?t1, ?object-2, ?t1)

EXCEED(?attribute, ?object-1, ?object-2, ?t1) $\Longleftrightarrow$
    PRESENT(?attribute, ?object-1, ?t1) AND
    PRESENT(?attribute, ?object-2, ?t1) AND
    GREATER(?attribute, ?object-1, ?t1, ?object-2, ?t1)

NOT-EXCEED(?attribute, ?object-1, ?object-2, ?t1) $\Longleftrightarrow$
    PRESENT(?attribute, ?object-1, ?t1) AND
    PRESENT(?attribute, ?object-2, ?t1) AND
    NOT-GREATER(?attribute, ?object-1, ?t1, ?object-2, ?t1)

[7] Variables are denoted by labels beginning with "?". Also, note that several of the "NOT-" forms are not strict logical negations: they assume some of the same information as the positive forms, in line with common usage of these terms.

For comparisons across time points:[8]

```
APPEAR(?attribute, ?object, ?t1, ?t2) ⟺
        PRESENT(after, <?t2, ?t1>, null) AND
        NOT-PRESENT(?attribute, ?object, ?t1) AND
        PRESENT(?attribute, ?object, ?t2)

NOT-APPEAR(?attribute, ?object, ?t1, ?t2) ⟺
        PRESENT(after, <?t2, ?t1>, null) AND
        NOT-PRESENT(?attribute, ?object, ?t1) AND
        NOT-PRESENT(?attribute, ?object, ?t2)

DISAPPEAR(?attribute, ?object, ?t1, ?t2) ⟺
        PRESENT(after, <?t2, ?t1>, null) AND
        PRESENT(?attribute, ?object, ?t1) AND
        NOT-PRESENT(?attribute, ?object, ?t2)

NOT-DISAPPEAR(?attribute, ?object, ?t1, ?t2) ⟺
        PRESENT(after, <?t2, ?t1>, null) AND
        PRESENT(?attribute, ?object, ?t1) AND
        PRESENT(?attribute, ?object, ?t2)

CHANGE(?attribute, ?object, ?t1, ?t2) ⟺
        NOT-DISAPPEAR(?attribute, ?object, ?t1, ?t2) AND
        NOT-EQUAL(?attribute, ?object, ?t1, ?object, ?t2)

NOT-CHANGE(?attribute, ?object, ?t1, ?t2) ⟺
        NOT-DISAPPEAR(?attribute, ?object, ?t1, ?t2) AND
        EQUAL(?attribute, ?object, ?t1, ?object, ?t2)

INCREASE(?attribute, ?object, ?t1, ?t2) ⟺
        NOT-DISAPPEAR(?attribute, ?object, ?t1, ?t2) AND
        GREATER(?attribute, ?object, ?t2, ?object, ?t1)

NOT-INCREASE(?attribute, ?object, ?t1, ?t2) ⟺
        NOT-DISAPPEAR(?attribute, ?object, ?t1, ?t2) AND
        NOT-GREATER(?attribute, ?object, ?t2, ?object, ?t1)

DECREASE(?attribute, ?object, ?t1, ?t2) ⟺
        NOT-DISAPPEAR(?attribute, ?object, ?t1, ?t2) AND
        GREATER(?attribute, ?object, ?t1, ?object, ?t2)

NOT-DECREASE(?attribute, ?object, ?t1, ?t2) ⟺
        NOT-DISAPPEAR(?attribute, ?object, ?t1, ?t2) AND
        NOT-GREATER(?attribute, ?object, ?t1, ?object, ?t2)
```

[8] "Null" is also used here to designate an irrelevant time point argument.

The above ten predicates, describing comparisons across two time points, form an exhaustive, though overlapping set of assertions concerning qualitative changes in attributes, provided that it is known that one time point follows another, and it is known whether or not the specified attribute is non-null at each time point. Situations beyond the range of these special circumstances must be described using (NOT-)PRESENT, (NOT-)MATCH, (NOT-)EXCEED, and, where necessary, the primitive predicates (NOT-)EQUAL and (NOT-)GREATER. For qualitative attributes (e.g., shape, color, and qualitative "position," "orientation," and "heading"), changes are typically specified in terms of the predicates (NOT-)APPEAR, (NOT-)DISAPPEAR and (NOT-)CHANGE. Boolean attributes (e.g., contact, "inside," support and "part-of") are viewed as a subset of the qualitative attributes—namely, those having only a single non-null value. Changes in boolean attributes are specified solely in terms of (NOT-)APPEAR and (NOT-)DISAPPEAR. Finally, changes in quantitative attributes (e.g., speed, elevation, distance and pressure) are specified in terms of all ten of the above predicates for comparisons across time points.

A point in transition space corresponds to a *set* of assertions within and across two time points. For example, the following information (grouped into static versus dynamic assertions) would appear in the specification of a transition involving "object-11" sliding across "surface-11" between the time points "t11" and "t12." (In summary, the solid object "object-11" changes position, decreases in speed, maintains an identical heading and orientation with respect to the solid object "surface-11," maintains its relationship above "surface-11," an absence of distance to "surface-11," contact with "surface-11" and support by "surface-11.")[9]

PRESENT(is-a, <object-11, solid-object>, t11)
PRESENT(is-a, <surface-11, solid-object>, t11)

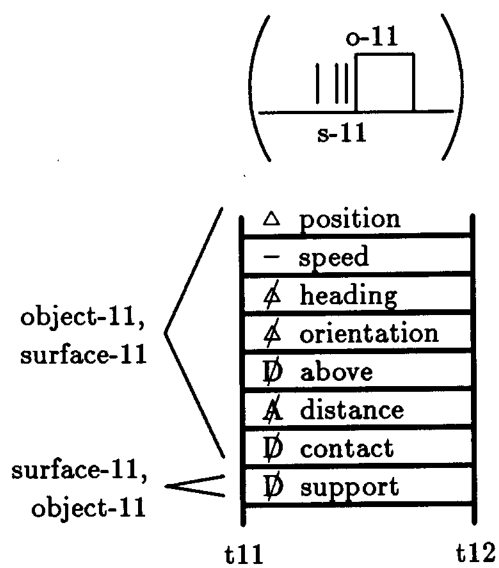
CHANGE(position, <object-11, surface-11>, t11, t12)
DECREASE(speed, <object-11, surface-11>, t11, t12)
NOT-CHANGE(heading, <object-11, surface-11>, t11, t12)
NOT-CHANGE(orientation, <object-11 surface-11>, t11, t12)
NOT-DISAPPEAR(above, <object-11, surface-11>, t11, t12)
NOT-APPEAR(distance, <object-11, surface-11>, t11, t12)
NOT-DISAPPEAR(contact, <object-11, surface-11>, t11, t12)
NOT-DISAPPEAR(support, <surface-11, object-11>, t11, t12)

[9]There is some latitude in the choice of attributes for inclusion in the specification of particular transitions (e.g., "weight" and "color," among other attributes, are omitted here, and "position" is a 3-dimensional specification which could be replaced by a 2-dimensional "location" plus a specification of "elevation"). Varying accounts of the same physical behavior may be related by the abstraction operations discussed in the following section.

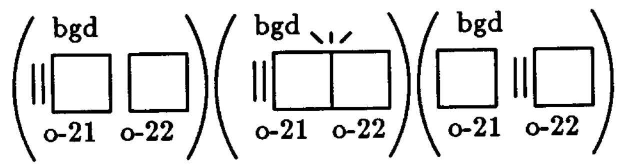
As the above form is rather unwieldy from a human perspective, we employ a special graphic representation of transitions in the remainder of this paper. Specifically, we indicate only the dynamic information in a transition, listing changes in attributes occurring between time points depicted as vertical lines. In these diagrams, the ten qualitative change assertions are coded as follows:

A	APPEAR,	A̸	NOT-APPEAR,
D	DISAPPEAR,	D̸	NOT-DISAPPEAR,
△	CHANGE,	△̸	NOT-CHANGE,
+	INCREASE,	+̸	NOT-INCREASE,
−	DECREASE,	−̸	NOT-DECREASE.

The graphic representation for the above transition involving sliding is then as follows. The specific objects to which each attribute or group of attributes applies are supplied at the left. Also, for ease of (our) visualization, a drawing of the activity appears above.



Finally, events are represented as sequences (or, more generally, as directed, acyclic graphs) of transitions. These representations are called "event traces," as they correspond to simple paths in transition space. As such, events containing two or more transitions constitute the simplest causal explanations. Below is an illustration of an event "propel-by-collision" ("launching" in the psychology literature [Michotte, 1946]). In this event, "object-21" approaches a stationary "object-22" during the first transition, the two objects come into contact during the second transition, and "object-22" departs from a now-stationary "object-21" during the third transition.



		t21–t22	t22–t23	t23–t24
object-21, background		Δ position	Δ position	$\not\Delta$ position
		$\not\Delta$ speed	D speed	$\not A$ speed
		$\not\Delta$ heading	D heading	$\not A$ heading
object-21, object-22		$-$ distance	D distance	A distance
		$\not A$ contact	A contact	D contact
object-22, background		$\not\Delta$ position	$\not\Delta$ position	Δ position
		$\not A$ speed	$\not A$ speed	A speed
		$\not A$ heading	$\not A$ heading	A heading

t21 t22 t23 t24

Event traces meet our representational objectives outlined above: they are detailed at the level of time and change, yet depict events from a cognitive/perceptual standpoint. Given suitable object types, attributes and granularity of time points, traces matching verbal accounts of perceived events may be constructed.

It should be noted that event traces are not limited to the representation of concrete events. The representation may also be used to depict events involving *conceptual* objects and attributes: for example, traversal of a path by an object [Jackendoff, 1983], the metaphorical "movement" of upcoming events toward us as if they were a passing parade [Lakoff and Johnson, 1980], and enabling and prevention of events by other events [Talmy, 1988]. Additionally, different traces may represent a single event at alternate levels of abstraction or in terms of different underlying metaphors, as illustrated in the following section.[10]

Given a knowledge base of event traces depicting simple causal scenarios, larger causal explanations may be constructed by combining traces which overlap in their specified transitions. Overlapping is computed by a matching process at the level of EQUAL and GREATER assertions. These and other associations between event traces are outlined in the following section.

[10]As specified, event traces are similar to histories [Hayes, 1985a; Hayes, 1985b; Kuipers, 1986; Forbus, 1987; Forbus, 1988], modulo the sorts of distinctions raised in Section 1 (regarding representational ontology, latitude with respect to analogy and abstraction, view of causation and basis in transitions rather than states).

3 Reconstructing Informal Causal Explanations

Informal causal descriptions portray sets or sequences of events, sometimes explicitly stating the causal relationships ("Closing the switch results in a flow of current.") and sometimes omitting or only hinting at these relationships ("The coil heats up the water. Eventually, the water begins to boil.")[11] The comprehender's task is one of assembling these events into a coherent whole. One source of discontinuity in these descriptions is the occurrence of causal gaps. For instance, a description of the operation of a pistol might begin with a statement "The trigger is pulled, releasing a hammer that strikes a firing pin." Comprehension of this statement involves the construction of a scenario in which the pulled trigger moves (rather than stretching, for example), subsequently unlatching the hammer, and in which the hammer, once released, is propelled (as by a spring) against the firing pin. Similarly, discontinuities in a description may arise through the use of analogy (the metaphorical use of a verb, for example) or abstraction (for example, specification of activity in terms of the whole rather than the parts).

Interestingly, the transition space formalism suggests a view of causal explanation as involving three major types of inter-event associations, these corresponding rather closely to the general notions of causality, analogy and abstraction. The intuitive argument supporting this characterization is as follows. If we model causal explanation as a process of constructing chains of event traces drawn from a central knowledge base, then if we are in a position of requiring a causal antecedent or consequent for a particular point in transition space—say, concerning propagation of a radio signal from a transmitting antenna—yet the extent of our knowledge is such that no available event traces involve that point, then we have no choice but to enlist the use of traces associated with other positions in the space—for example, metaphorically-relevant traces concerning the throwing or carrying of physical objects. To accomplish this "borrowing" of information requires that we apply a transformation to our position in transition space, and, following the utilization of traces in the new location, a second transformation returning us to the vicinity of our point of departure. Here, we distinguish between two classes of such transformations: those that belong to an inverse pair of transformations, and those that do not. If a transformation belongs to an inverse pair, then no information is lost in the course of either transformation. This corresponds roughly to the notion of analogy: we may reason about motion of a radio signal or a quantity of liquid as if it were a solid, or temperature as if it were elevation, and we may map sequences of transitions involving the motion or elevation back to equivalent transitions for the original items. On the other hand, if the transformations do not form an inverse

[11] A good summary of grammatical devices used in the indication of causality appears in [Leech and Svartvik, 1975].

pair (i.e., one or the other does not have a unique inverse), then information is not preserved at some point. This corresponds roughly to the notion of abstraction: for example, we may reason about the whole instead of the parts, but upon completion, we must make assumptions about the current status of the parts.

Carrying the intuitive argument a bit further, we might note that in a more general sense, it is somewhat more complex *combinations* of chaining associations, information-preserving transformations and non-information-preserving transformations that, in transition space, correspond to the more general notions of causality, analogy and abstraction. For example, we might have to edit out extraneous information in an event trace (a non-information-preserving transformation) or substitute names of objects and time points in a trace serving as a precedent (an information-preserving transformation) before we may establish the exact match between two transitions required for a chaining association. Likewise, analogies between events may require an initial editing of one or both traces before a substitution transformation may be applied to one, producing the other.

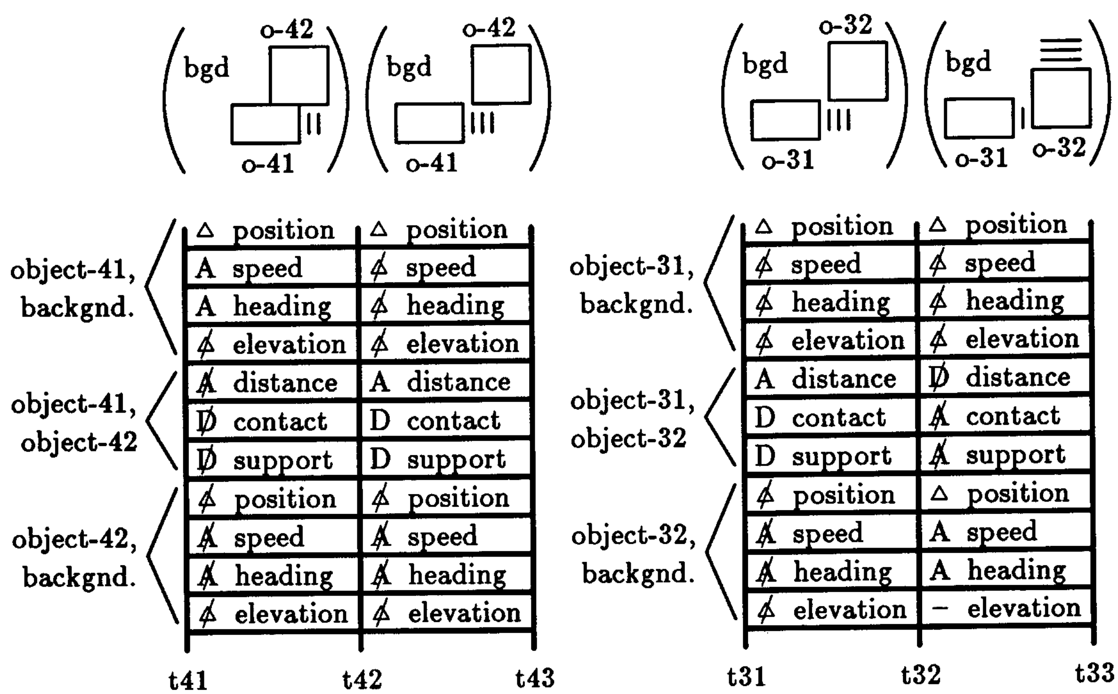
Given the above general characterization, we may model the reconstruction of informal causal explanations as a process of constructing two related structures in transition space. The first is an *association structure*, a network of chaining associations and transformations of both varieties relating the events in a description to one another and to events in the knowledge base. As part of this structure, a number of additional traces may be generated, serving as intermediate, transformed images of the original traces. (An illustration of a simple association structure appears in Figure 1, Section 4, in connection with the example in that section.) The second structure is a *composite trace* summarizing the overall activity, formed by a merging of relevant event traces appearing in the association structure. (A corresponding illustration appears in Figure 7, Section 4.) The difficult part in the reconstruction process is elaboration of the association structure, involving a search activity similar to the spreading activation process described in the text comprehension and natural language understanding literature [Quillian, 1969; Collins and Loftus, 1975; Kintsch, 1988; Alterman, 1985; Norvig, 1989]. A critical difference here, however, is that we do not use a pre-elaborated semantic network of inter-event associations as a medium for conducting the search. Here, the associations arise *implicitly* through comparisons and transformations involving the internal, transition-level representations of events.

The remainder of this section outlines a basic repertoire of inter-event associations suited to the reconstruction of informal causal explanations in transition space. These associations are grouped according to the three association types discussed above. In the next section, a number of these associations are used in illustrating the reconstruction of a simple causal explanation involving rocket propulsion.

Chaining of Event Traces

Consequence

The consequence association concerns chaining of event traces by exact match between transitions. For this relation to hold, one of the transitions in a first trace (the antecedent trace) must contain the same set of assertions as the initial transition in a second trace (the consequent trace).[12] The two traces below illustrate a slightly more complex causal relation, with the final transition of the first trace matching the initial transition of the second trace, modulo a replacement of objects and time points—a substitution operation as described below. In the first trace, "object-41" moves sideways, leading to a disappearance of contact and support for "object-42," while in the second trace, "object-31" begins by moving out from under "object-32," resulting in "object-32" beginning to fall. Suppose the second trace is already present in the knowledge base, resulting from the processing of a previous causal scenario, and the first trace arises in the processing of a new causal scenario. Following a substitution transformation applied to the second trace, a consequence association may be established between the two traces.

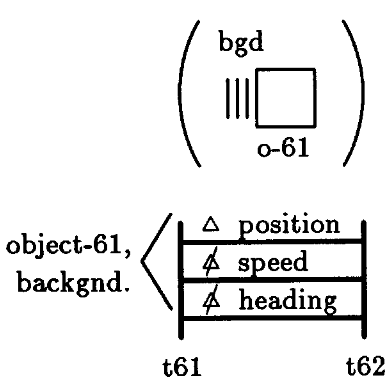


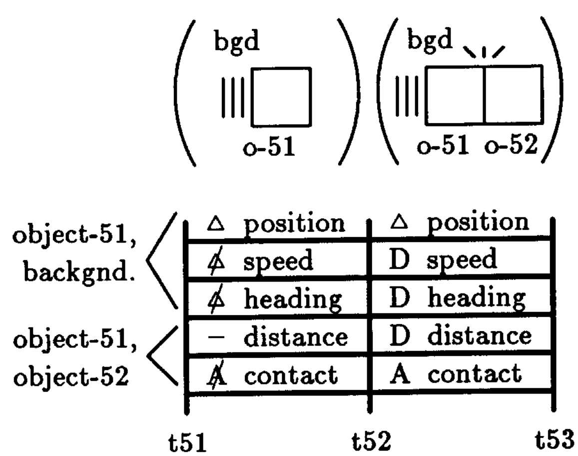
[12]In a more general sense, we would require one of the transitions in the antecedent trace to contain a *superset* of the assertions appearing in the initial transition of the consequent trace. As described above, this corresponds to an initial editing operation, followed by a consequence association.

Separately, using the consequence association we may implement a transition-space version of the STRIPS assumption [Fikes and Nilsson, 1971]. Here, in the absence of conflicting information, we assume that asserted *changes* continue into succeeding intervals (except that APPEAR and DISAPPEAR progress to NOT-DISAPPEAR and NOT-APPEAR, respectively). These defaults are established on demand by drawing consequence associations between changes asserted over preceding intervals, on the one hand, and on the other hand, substituted images of special traces maintained in the knowledge base, specifying, for example, a NOT-CHANGE followed by a NOT-CHANGE or an INCREASE followed by an INCREASE.

Contingency

Similarly, partial matches may be used in chaining. If a transition of one trace contains a *subset* of the assertions in the initial transition of a second trace (the opposite circumstance to that requiring editing, as noted above), a tentative consequence/antecedence relation may be drawn, pending explanation of the missing components. The following example illustrates this "contingency" association. In the first trace, "object-61" is simply moving with respect to the background. Suppose it is desired to form an explanation for "object-61" colliding with another object, say "object-62," and the second trace is present in the knowledge base. In working memory, we may then set up a contingency relation between the first trace and the product of a substitution operation (again, involving objects and time points) on the second trace, supplying the needed explanation while noting the absence of direct information that "object-61" is indeed approaching "object-62."





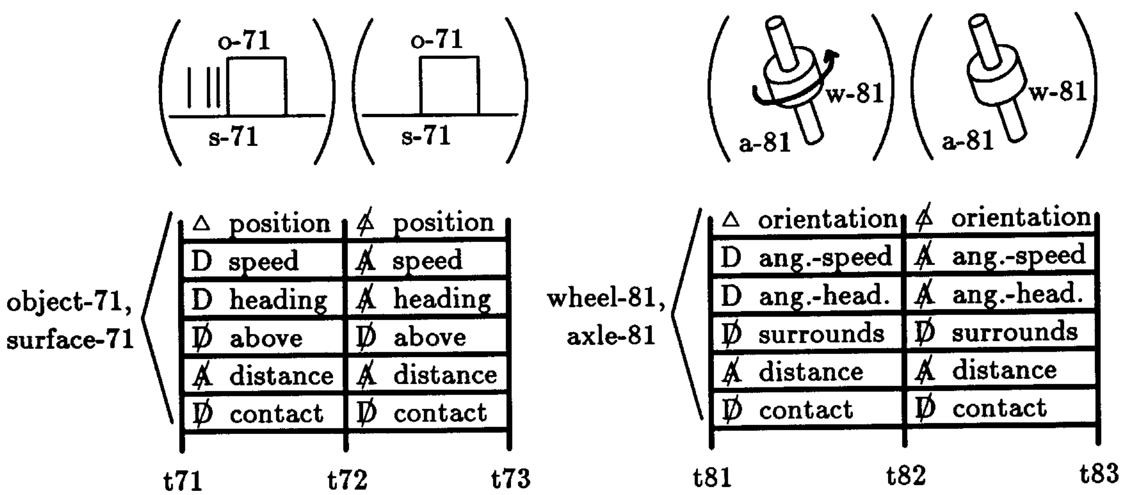
Information-Preserving Transformations

Substitution

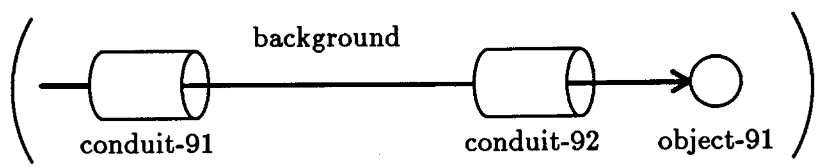
Substitution involves the replacement of time points, objects, prototypes or attributes in event traces with alternate quantities, leaving the remaining structure of the traces unaltered. In general, the use of substitution must be restricted to particular contexts (constrained, for example, by a library of allowable analogical associations), and further governed by current reasoning goals (limited to analogical associations concerning attributes of immediate interest, for example).

At the simplest extreme, an event and its reoccurrence may be related by a substitution association involving only the replacement of time points. Progressing to more complex varieties, a reoccurrence of an event involving different objects of the same type may be related through a substitution association involving time points and objects, a reoccurrence involving different objects of different types through an association involving time points, objects, and prototypes specifying object classes, and finally, an event involving parallel changes in different attributes for objects of different types through an association involving all of these quantities plus one or more attributes.

The example below illustrates a substitution association involving time points, objects, prototypes and attributes. In the first trace, an object slides to a halt, while in the second trace, a wheel mounted on an axle spins to a halt.



Similarly, substitution associations may exist between traces depicting concrete, physical situations and traces depicting situations in abstract domains. For example, event occurrences may be viewed metaphorically in terms of several physical scenarios: events as objects which pass by us in sequence [Lakoff and Johnson, 1980]; events as objects which set each other in motion [Talmy, 1988]; events as positions or containers, with an "energy" object moving between them (one interpretation of the research by Shultz [1982]); and events as conduits leading between positions corresponding to states or transitions. The traces below illustrate this last scenario, with positions corresponding to states as in the state space analogy employed in AI problem solving and planning. Here, an event trace depicting an object passing through two conduits is contrasted with a trace depicting a device changing state while engaging in a sequence of two events.

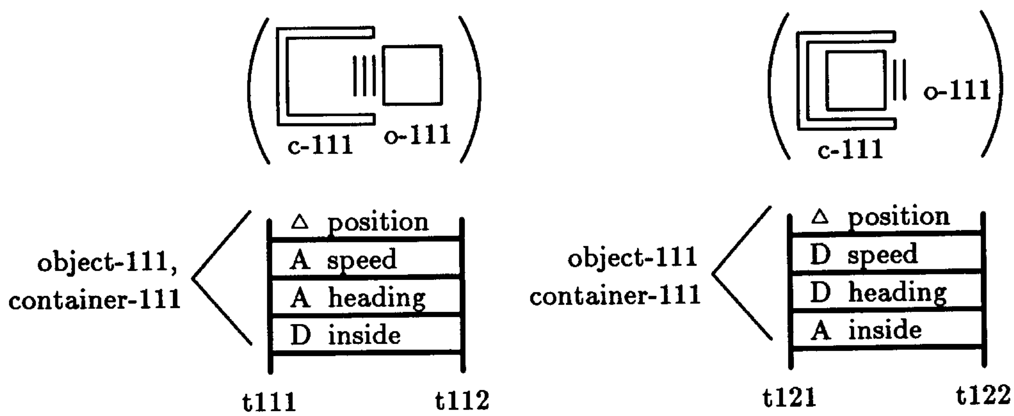


	t91–t92	t92–t93	t93–t94	t94–t95
object-91, backgnd. <	△ position	△ position	△ position	△ position
object-91, conduit-91 <	A inside	D inside	A̸ inside	A̸ inside
object-91, conduit-92 <	A̸ inside	A̸ inside	A inside	D inside

	t101–t102	t102–t103	t103–t104	t104–t105
device-101 <	△ state	△ state	△ state	△ state
device-101, event-101 <	A engaged-in	D engaged-in	A̸ engaged-in	A̸ engaged-in
device-101, event-102 <	A̸ engaged-in	A̸ engaged-in	A engaged-in	D engaged-in

Interchange

Substitutions may also be performed within the sets of time points, objects, prototypes and attributes of particular traces. This transformation is primarily of use in searching for paths of recovery from failures or deviations, and may involve exchange of quantities only in some respects, leaving other assertions of an event trace in their original form. For example, in the first trace below, "object-111" moves outside "container-111." If the goal is to discover a means of counteracting this transition, we may first form a new trace by interchanging the time points in the first trace with respect to temporal order, subsequently renaming them (a substitution operation), resulting in a retrograde of the original transition. Following this, we may search for a possible causal antecedent of this second trace (e.g., another object strikes "object-111," altering its heading so that it approaches and reenters "container-111").



Non-Information-Preserving Transformations

Generalization

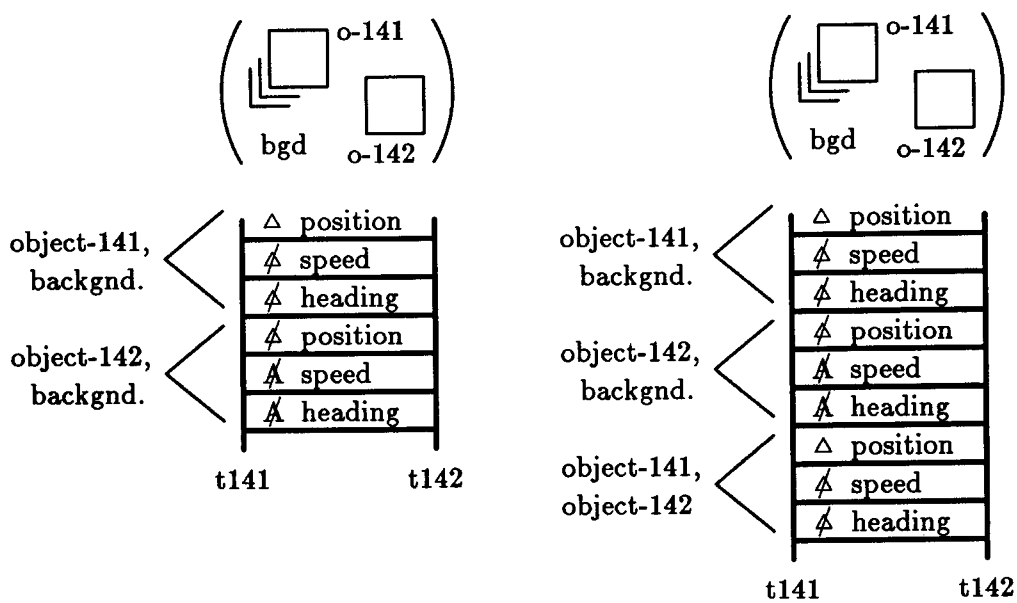
Generalization involves the replacement of prototypes with superordinate prototypes. The following two pairs of static assertions illustrate this association in the context of the attributes "is-a" and "made-of." In each pair, the first statement contains a more specific prototype and the second a general, encompassing quantity. Traces may also be specialized by replacing prototypes with subordinate prototypes; however, as with all abstraction operations, this requires independently-supplied information or the making of assumptions.

PRESENT(is-a, <object-131, wheel>, t131)
PRESENT(is-a, <object-131, solid-object>, t131)

PRESENT(made-of, <object-131, steel>, t131)
PRESENT(made-of, <object-131, metal>, t131)

Editing

One way to abstract an event trace is simply to delete information pertaining to particular time points, objects, prototypes or attributes. In the other direction, we may specialize a trace by including logical consequences arising from reflexivity, symmetry or transitivity of particular predicates or attributes, or by including consequences arising from particular physical constraints, as illustrated below. Here, a trace specifying "object-141" moving with respect to the background while "object-142" remains at rest is transformed into a second trace which augments this information with the assertion that "object-141" is moving with respect to "object-142."[13]

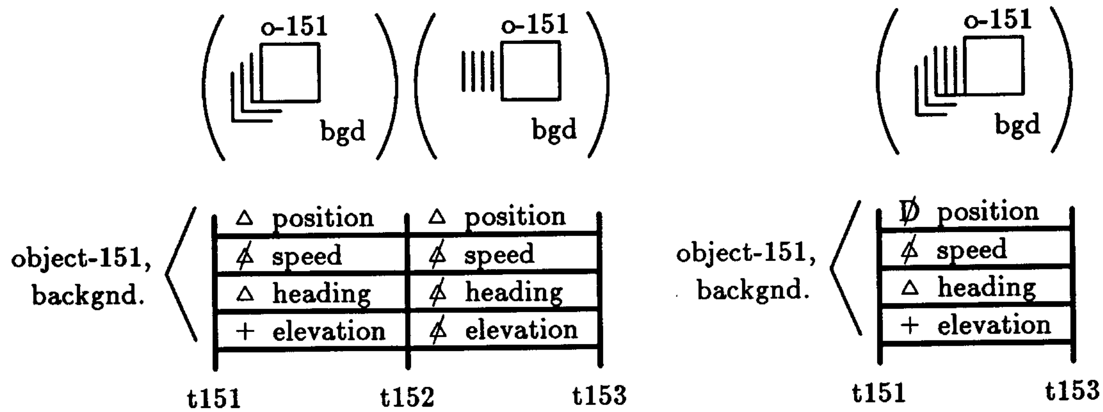


Composition

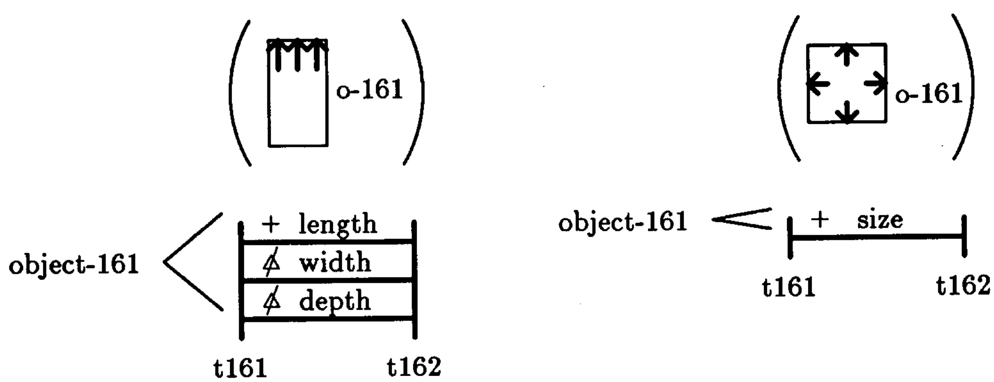
Composition concerns movement from descriptions concerning parts to descriptions concerning wholes. Three varieties of composition are illustrated below. The first example concerns temporal composition—the merging of temporal intervals to form composite intervals. In this example, "object-151" rises from time "t151" to "t152," changing its heading so that from "t152" to "t153" it moves horizontally.

[13]Most inferences of this sort are best made on demand only, using backward chaining.

Merging the two intervals into a single interval (and applying rules of composition for the indicated changes) produces the second trace. Here, the object is seen as simply rising and changing heading from time "t151" to "t153."[14]

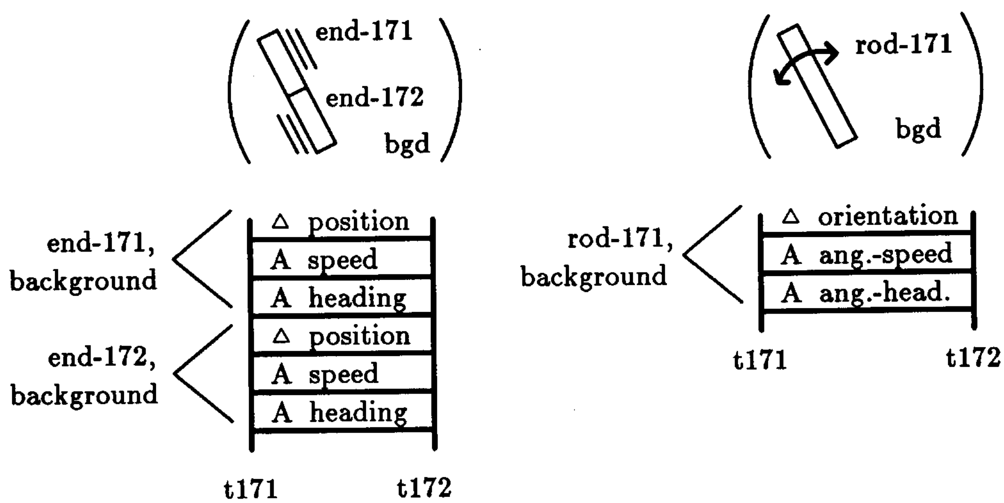


The second form, attribute composition, involves replacement of sets of attributes with encompassing (or roughly encompassing) attributes. In the example below, description of changes in terms of length, width and depth is transformed to a simpler description in terms of size.



[14]Regarding the attribute "position," the composition of "CHANGE" with "CHANGE" results in a weaker assertion "NOT-DISAPPEAR," as the concatenation of two changes does not necessarily imply an overall change. In this particular case, however, the object must have changed position, as it has increased in elevation. This could be established by a subsequent inference (editing) operation.

The third form, object composition, involves movement from description in terms of objects which are parts of a whole to description in terms of the whole. In the example below, differing movements on the part of two ends of a rod are translated into an overall rotation of the rod.[15]



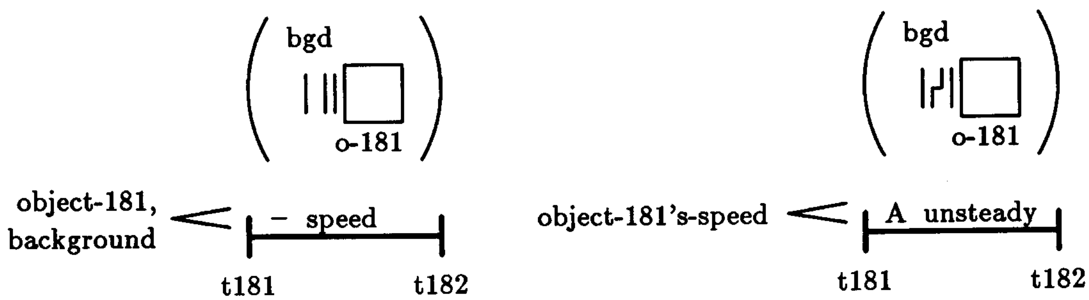
NOT-MATCH(heading, <end-171, background>,
<end-172, background>, t172)

Reification

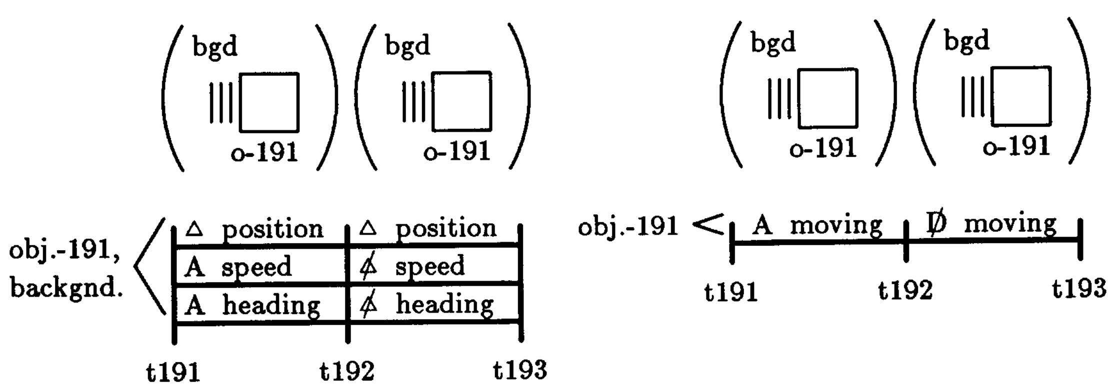
Finally, reification refers to the introduction of new symbols into the representation, representing quantities previously unrepresented or represented in a different form (see, for example, [Genesereth and Nilsson, 1987]). There are three varieties of reification applicable to our representation of transition space. In attribute-object reification, an attribute (as applied to a specific set of objects) is reified as a new object. This allows us to make assertions regarding the properties or behavior of a particular attribute as applied to a specified set of objects. In the example below, a trace involving "object-181" decreasing in speed is transformed into a trace in

[15]Static assertions relevant to particular examples are simply listed below the diagrammed event traces.

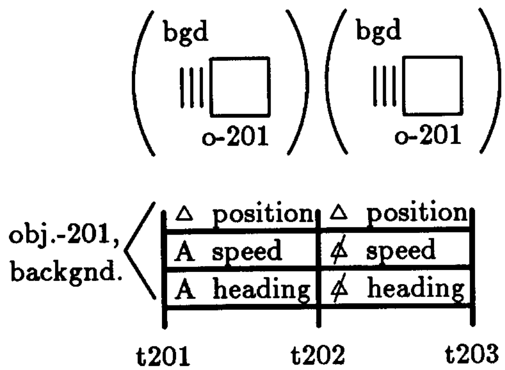
which the activity is described in terms of the appearance of instability in a new object, "object-181's-speed."

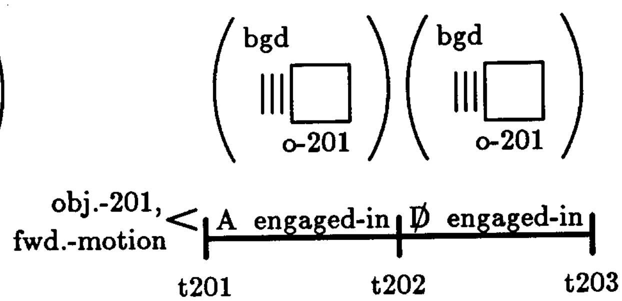


In the second form, event-attribute reification, an entire event trace is depicted in terms of an attribute applying to one or more of the participants in the event. In the following example, two ways of describing motion are illustrated. In the first trace, motion is represented as before, in terms of changes in the attributes position, speed and heading for "object-191." In the second trace, motion is described simply in terms of appearance and subsequent non-disappearance of an attribute "moving" applied to "object-191."



Finally, in event-object reification, an event trace is reified as an object. In the following example, forward movement of "object-201" in the first trace translates to appearance and subsequent non-disappearance of "object-201" being "engaged-in" the (event-)object "forward-motion."





MATCH(heading, <object-201, background>,
forward, t202)

As the last two examples illustrate, there are several levels of abstraction at which we may represent events. These levels correspond to alternate forms of description available in everyday language: we may say that an object has changed position, we may speak of "moving" objects (treating motion as a property), or we may speak of objects that are "engaged in movement." Reification associations in transition space relate these alternate forms of description.

Given the range of inter-event associations and levels of representation outlined in the preceding paragraphs, the appearance of multiple paths of association between events appearing in informal causal descriptions is to be expected. In particular, paths of association involving the reified event-object level, as illustrated above in the example concerning "device-101" successively engaging in "event-101" and "event-102," must not be accepted as adequate reconstructions of informal causal explanations where equivalent reconstructions exist at the lower "non-reified" level, although the reified event-object level may certainly be used in the manipulation of templates for the construction of paths of association at the lower non-reified level.

Similarly, metaphorical descriptions must not be accepted where literal description is possible. For example, we might attempt to explain the fact that a loudspeaker has become broken by saying that the signal fed to it was "too large for it to handle." This could be explained better in terms of forces too great for the diaphragm or its wiring or electrical potentials causing arcs in the voice coil, depending on the particular circumstances. Of course, in many cases, it may be claimed that various physical phenomena are ultimately *comprehended* metaphorically (this is the central thesis advanced in [Lakoff and Johnson, 1980] and [Johnson, 1987]). Or, it may be the case that a more "literal" description is possible, but the author of a particular informal causal description has nonetheless advanced a metaphorical explanation in view of the intended audience or due to a limited comprehension of the domain. In such cases, the metaphorical explanation must suffice.

4 An Example

The following example, drawn from the *Encyclopedia Americana* [1989], illustrates the reconstruction of a simple informal causal explanation using a number of the associations outlined in the previous section. The emphasis here is on illustrating the suitability of the representational framework in general, and not on the evaluation of specific search heuristics, which must necessarily be done in the context of a large knowledge base containing both relevant and irrelevant event traces. Consequently, we confine the exposition to the specification of an input to the comprehension process plus an association structure and composite event trace constituting a plausible reconstruction of the causal explanation. In context, the following statement appears as the opening sentence of the entry for "rocket."

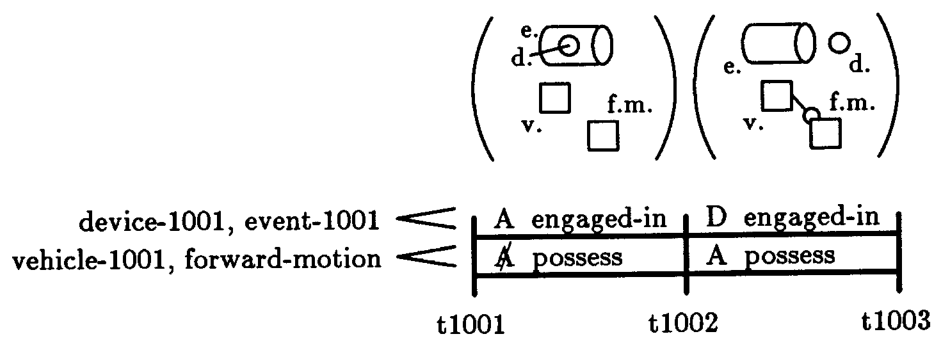
Rocket, a propulsion device that provides forward motion to a vehicle by expelling to the rear a jet of matter, usually heated gases.

Bypassing the problem of parsing this sentence, we start with the following pre-parsed, somewhat simplified rendition, accompanied by an indication that the former event is to be explained in terms of the latter.[16]

[16] As the central concern here is with reconstruction of causal chains, a de-emphasis has been placed on a number of other significant problems in understanding free natural language text—for instance, dealing with the range in syntactic forms, lexical disambiguation, anaphora, quantification, metonymy, and so forth.

[S [NP device] [VP [V provides] [NP forward motion] [PP to [NP vehicle]]]]
[S [NP device] [VP [V expels] [NP jet] [PP to [NP rear]]]]

At face value, the first statement describes an event involving a transfer of possession, treating "forward motion" as a thing to be given or possessed. In transition space, then, we take as a starting point an event trace depicting this literal interpretation of "provide..to." Since in general, it is not necessary for the provider of an item to be originally in possession of the item being provided (one might give a person money to purchase an item or lead the person to a spot where such an item is free for the taking, for example), the event trace for the first statement depicts "device-1001" engaging in an arbitrary activity "event-1001," leading to "vehicle-1001" coming into possession of "forward motion."[17] [18]

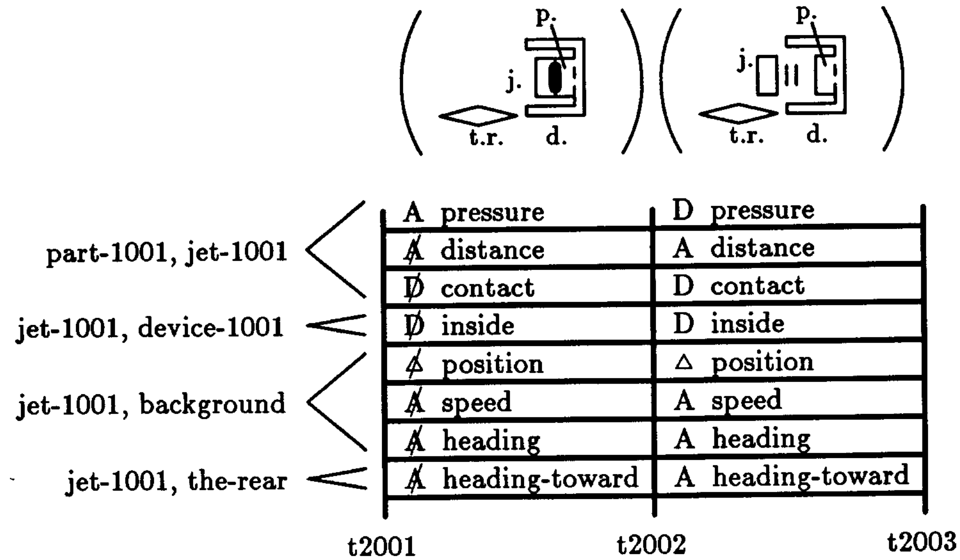


The second statement is depicted as follows. Again, a literal interpretation of the activity is taken. Here, this entails treating "the rear" as a physical location in whose direction the jet is expelled.[19] In this event trace, "jet-1001" is initially inside "device-1001," and the appearance of pressure between a part of the device, "part-1001," and "jet-1001" leads to "jet-1001" moving outside of the device, in the direction of the location "the rear."

[17] We treat "forward-motion" as a reified event-object here, represented as a prototype with an accompanying rule for translation into an equivalent event trace at the non-reified level.

[18] In the drawings accompanying event traces in this section, conceptual attributes such as "engaged-in" and "possess" are depicted by simple physical analogs. In particular, the conduit analog for engagement in an event is employed. As before, these drawings are provided solely for human visualization of the events described by the diagrammed traces.

[19] Similar to the treatment of "forward-motion" above, "the-rear" is represented as a prototype, with a corresponding rule for translation into a non-reified form.



The comprehension problem involves construction of an association structure relating the above two traces, using operations of the sort illustrated in the previous section. Such a structure is diagrammed in overview in Figure 1, with details of the operations provided in Figures 2 through 6. In Figure 1, a special 3-dimensional illustration of transition space is employed, with chaining associations proceeding along the horizontal axis, information-preserving transformations along the depth axis, and non-information-preserving transformations along the vertical axis.[20] In the diagram, the axes are labeled by their intuitive correlates "causality," "analogy" and "abstraction." The event trace for the first statement above, depicting "device-1001" providing forward motion to "vehicle-1001," is represented by the solid, heavy arrow labeled "A." The two points connected by this arrow represent the two transitions in the event trace, with the direction of the arrow specifying the direction of flow from the first transition to the second. Similarly, the event trace for the second statement, depicting "device-1001" expelling "jet-1001" to the rear, is represented by the solid, heavy arrow labeled "K." The dotted, heavy arrow labeled "H" corresponds to an event trace of special importance which must be supplied by the knowledge base. This trace portrays one object pushing backward on a second object, with the effect that the first object moves forward. Trace "H" is specified in terms of two objects "object-501" and "object-502."

The remaining, light arrows in Figure 1 correspond to intermediate event traces formed by various transformation operations on the three principle traces. Keeping Figure 1 handy as a source of overall context, we proceed by inspecting in detail five segments of the association structure.

[20]This 3-dimensional representation is, of course, imprecise, since many specific varieties of associations are compressed into each of the three dimensions. The diagram should be taken as simply a visualization of the space suited to human inspection.

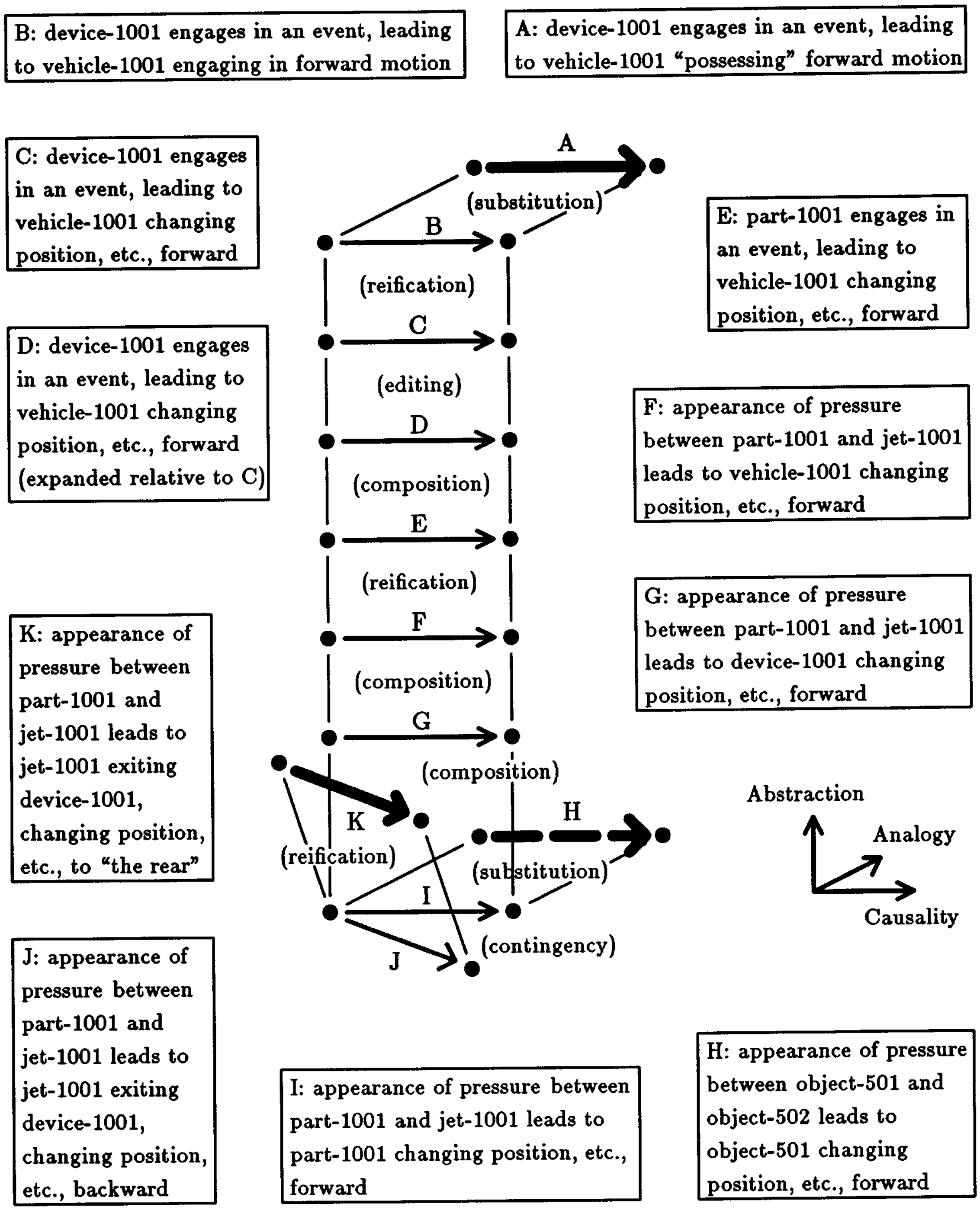


Figure 1: Association structure for the rocket example (overview).

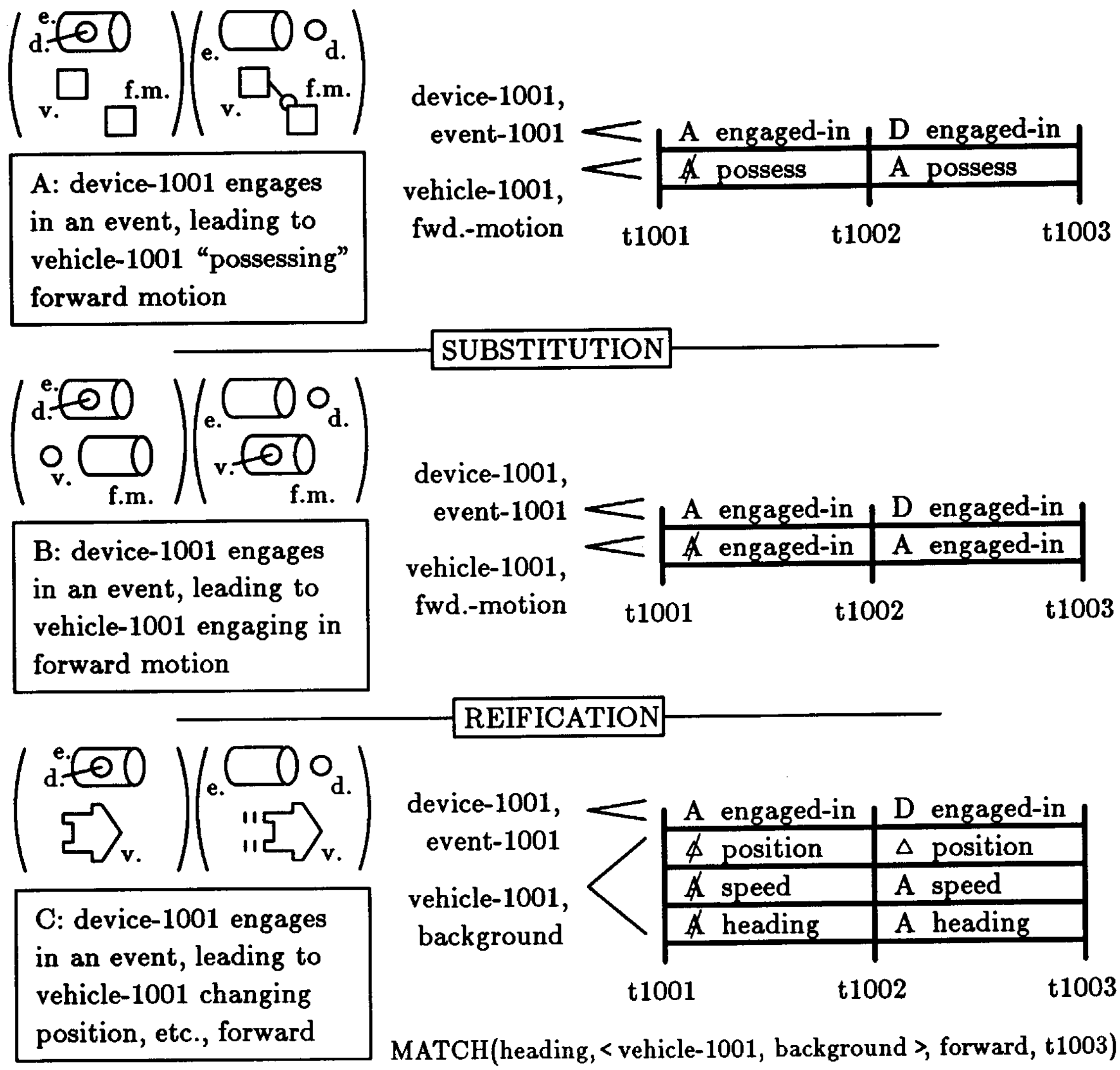


Figure 2: Association structure for the rocket example, traces "A" to "C."

The first segment, from "A" and "C," is illustrated in Figure 2. Here, in the course of two transformations, an explicit representation of forward motion for "vehicle-1001" is constructed. Specifically, from "A" to "B," a substitution operation replaces "possession" of forward motion with the property of being "engaged in" forward motion, and from "B" to "C," an inverse reification operation replaces engagement in forward motion with actual changes in position, speed and heading.[21]

[21]For simplicity, we align the background such that forward and backward motion for the vehicle are forward and backward in the global context, relative to the background.

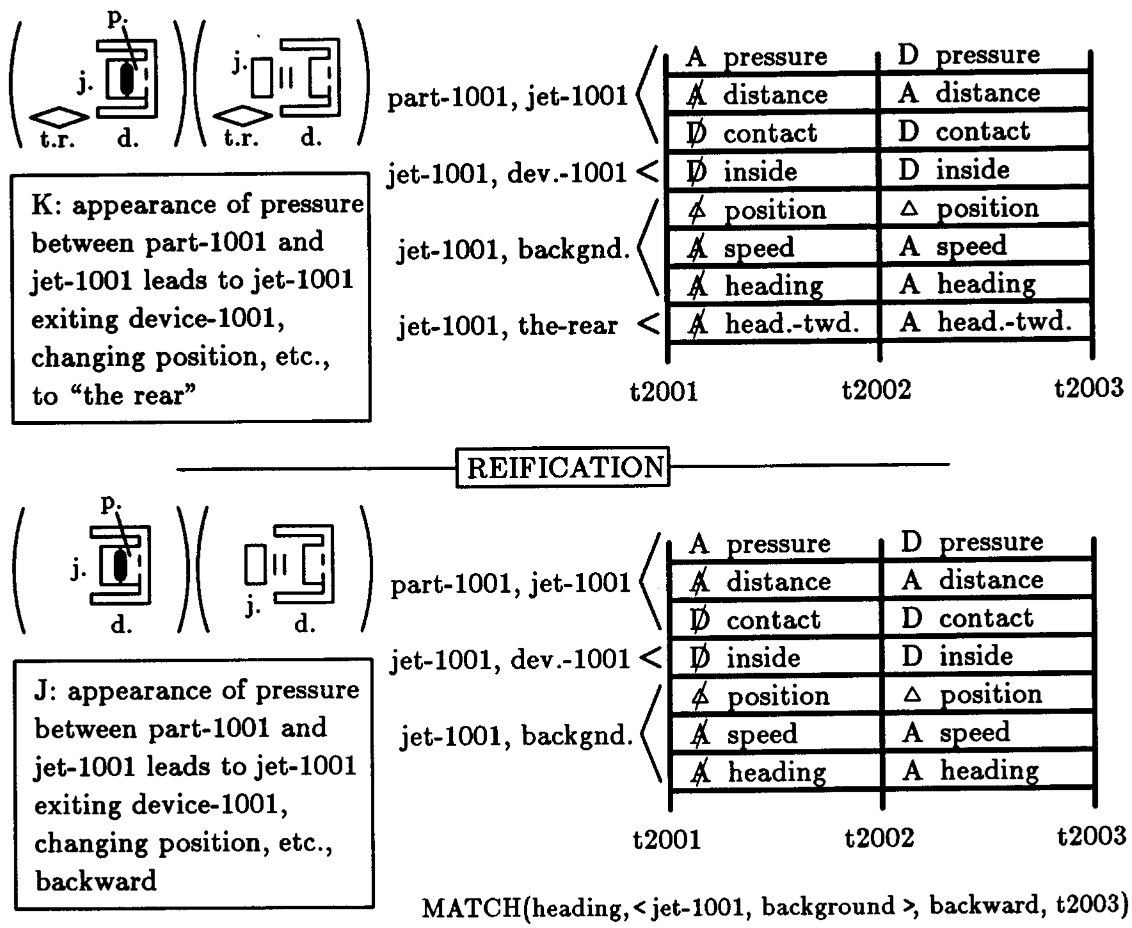


Figure 3: Association structure for the rocket example, traces "K" to "J."

Figure 3 illustrates the second segment, from trace "K" to trace "J." In this segment, we arrive at an explicit representation of backward motion for "jet-1001." Specifically, a single inverse reification operation eliminates the conceptual object "the-rear," substituting a representation in which the heading of "vehicle-1001" matches the direction prototype "backward."

In Figure 4, the segment between traces "J" and "H" is illustrated. Here, a contingency relation is drawn between the initial transition in trace "J"—involving the appearance of pressure between "part-1001" and "jet-1001"—and the causal scenario of such pressure leading to forward motion on the part of the object in front. As part of this relation, an intermediate trace "I" is constructed, in which the objects "part-1001" and "jet-1001" replace "object-501" and "object-502" in "H"[22]

[22]The initial transition in "I" is from "t2001" to "t2002" in anticipation of the match between this transition and the initial transition of "J." A new third time point is generated, however, since the second transition in "I" need not coincide with the second transition in "J."

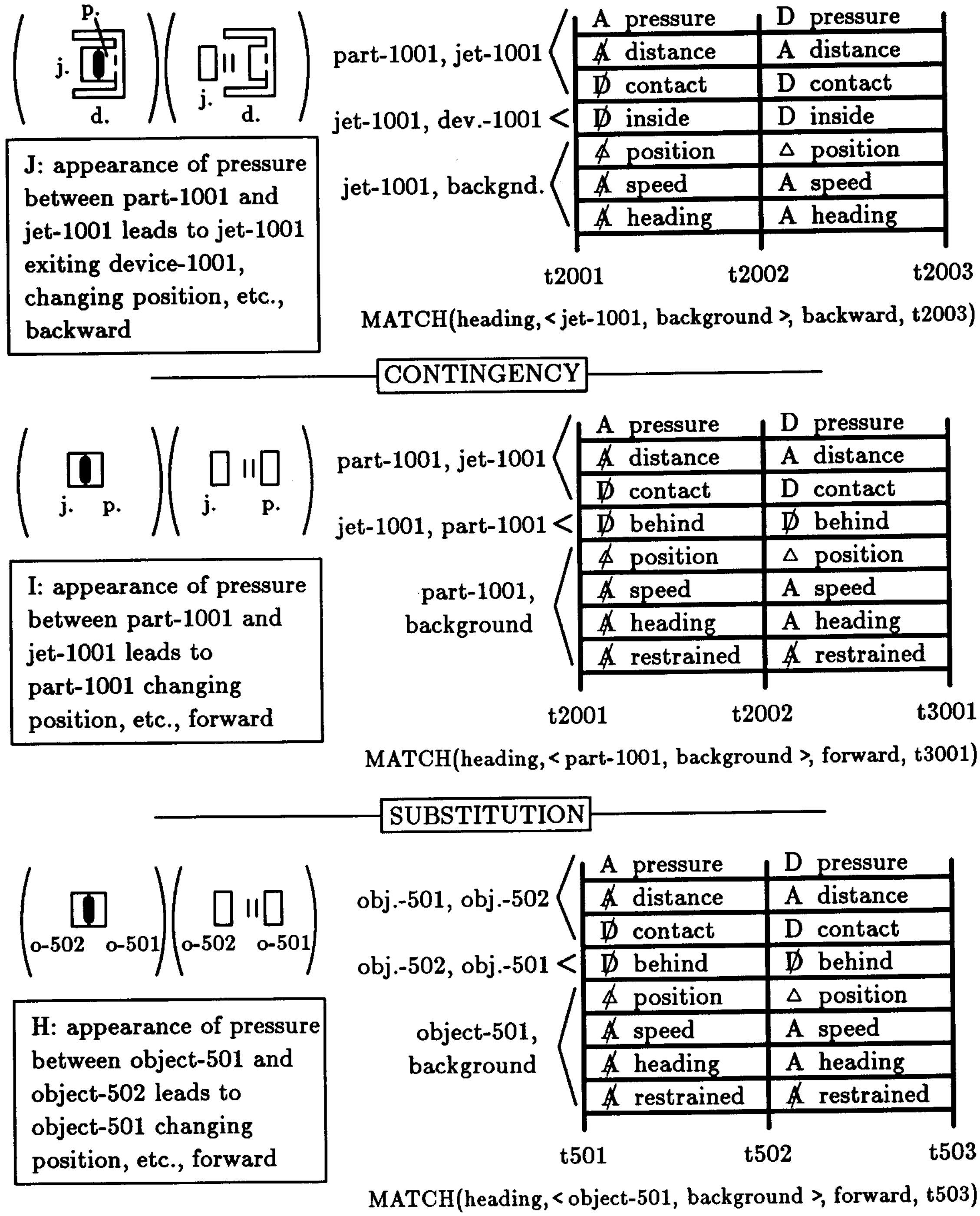


Figure 4: Association structure for the rocket example, traces "J" to "H."

The relation is one of contingency since the initial transition of trace "J" contains no information concerning motion or presence of restraint for "part-1001" with respect to the background—information which is present in the initial transitions of "H" and "I"[23] As discussed below, the incomplete match at this point may be used as a source of information regarding possible failure of the overall activity.

Figure 5 illustrates the next segment, from trace "I" up to trace "F." Here, a sequence of two object composition operations replaces forward motion for "part-1001" with forward motion for "vehicle-1001," assuming "part-1001" is rigidly attached to "device-1001" and "device-1001" is a rigidly attached part of "vehicle-1001." As before, the posing of assumptions provides a basis for subsequent reasoning about possible failures of the overall scenario.

Figure 6 illustrates the final segment of the association structure, from trace "F" up to trace "C." In this segment, the appearance of pressure between "part-1001" and "jet-1001" is equated with the activity "event-1001" engaged in by "device-1001" as a precursor to the forward movement of "vehicle-1001." Specifically, from trace "F" to trace "E," an event-object reification operation equates the interaction between "part-1001" and "jet-1001" in the initial transition of "F" with appearance of an activity "event-1002" in which "part-1001" engages. Next, from "E" to "D," an object composition operation relates "part-1001" engaging in "event-1002" with "device-1001" engaging in "event-1001."[24] Lastly, extraneous information regarding the attribute "restrained" is removed from "D," and the time point "t3001" generated for trace "I" is matched with point "t1003" (technically a substitution operation), producing a match with trace "C" and completing the overall association structure.

Regarding the search process responsible for constructing an association structure such as that described above, three factors are expected to provide assistance in overcoming the inherent combinatorics of the task. The first two are natural constraints, arising from the fact that informal causal descriptions are constructed by humans and intended for human consumption. First, intuitively it would seem that informal causal descriptions provide a sufficient number of "milestones" such that the depth of search required to associate pairs of events in a description remains reasonably bounded. Second, the order of exposition appears to be such that focusing techniques may be used to limit the breadth of search [Grosz, 1977].

[23]It is possible to infer that "jet-1001" is behind "part-1001" from the fact that the appearance of pressure results in "jet-1001" moving backwards. Also, an editing operation may remove extraneous information in "J" not needed for the contingency match. These operations are omitted here.

[24]In general, while it is safe to say that if a part of a device is participating in an event, the device as a whole is participating in an event, these events are not necessarily the same (e.g., the wheels of an automobile roll forward while the automobile as a whole simply *moves* forward).

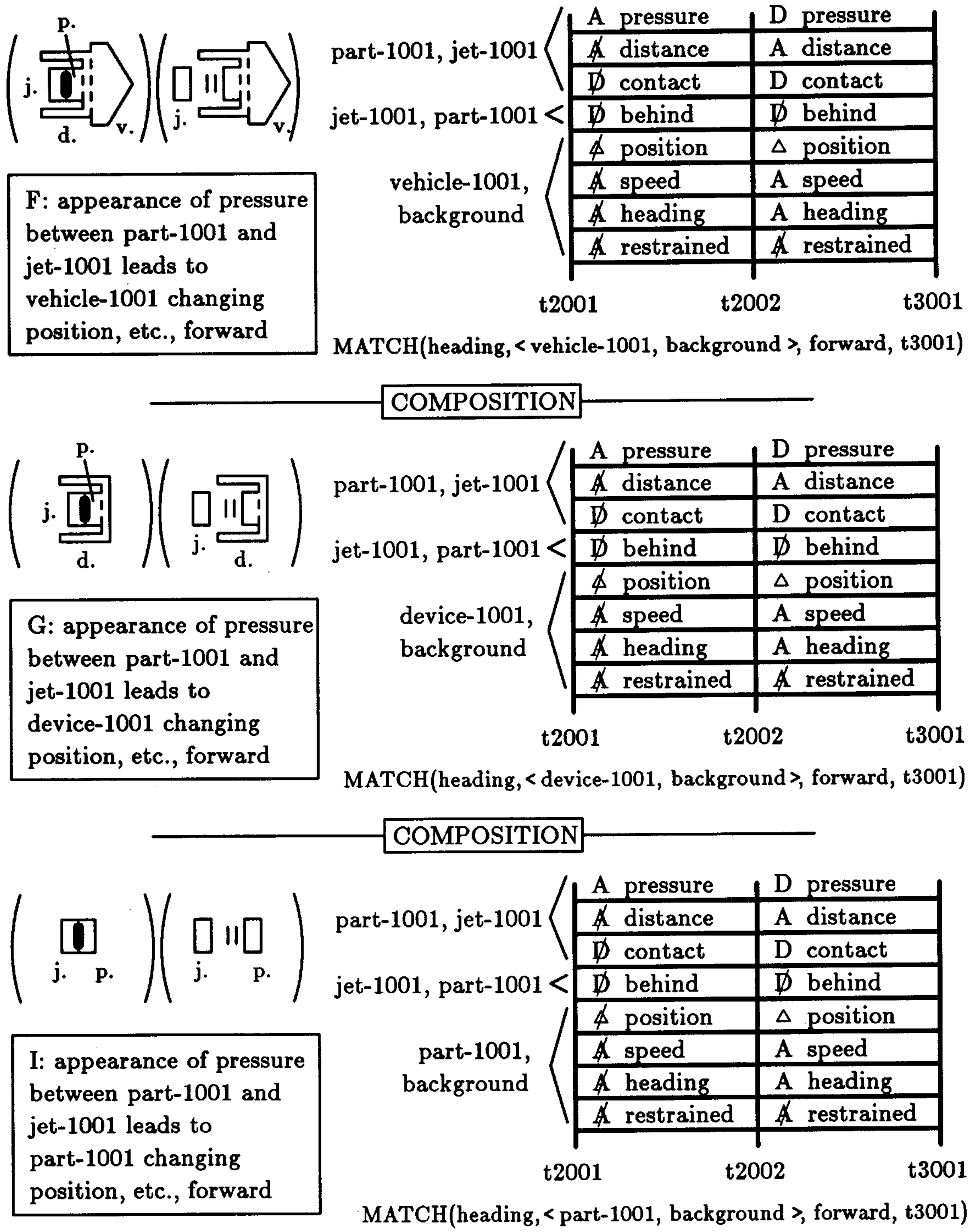


Figure 5: Association structure for the rocket example, traces "I" up to "F."

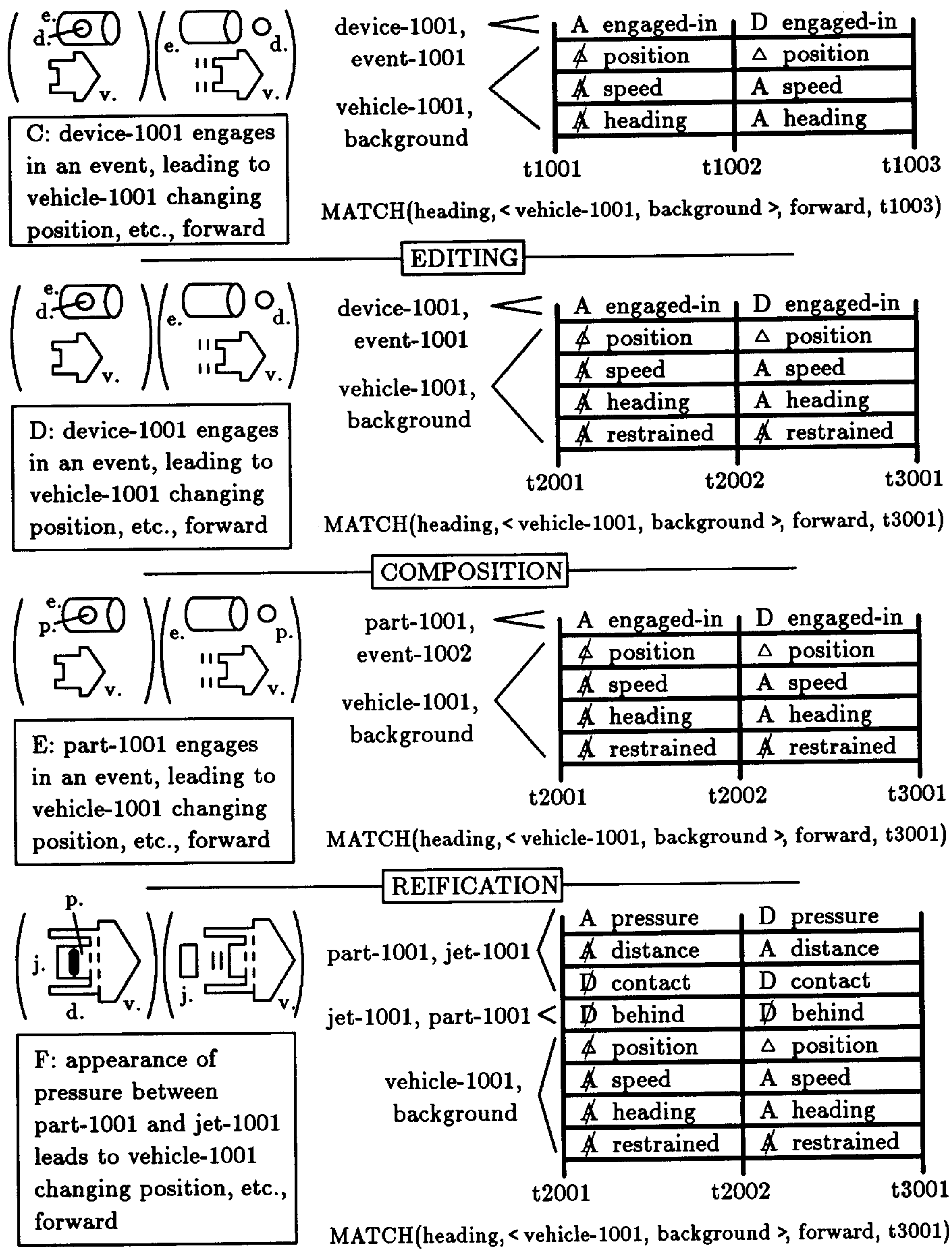


Figure 6: Association structure for the rocket example, traces "F" up to "C."

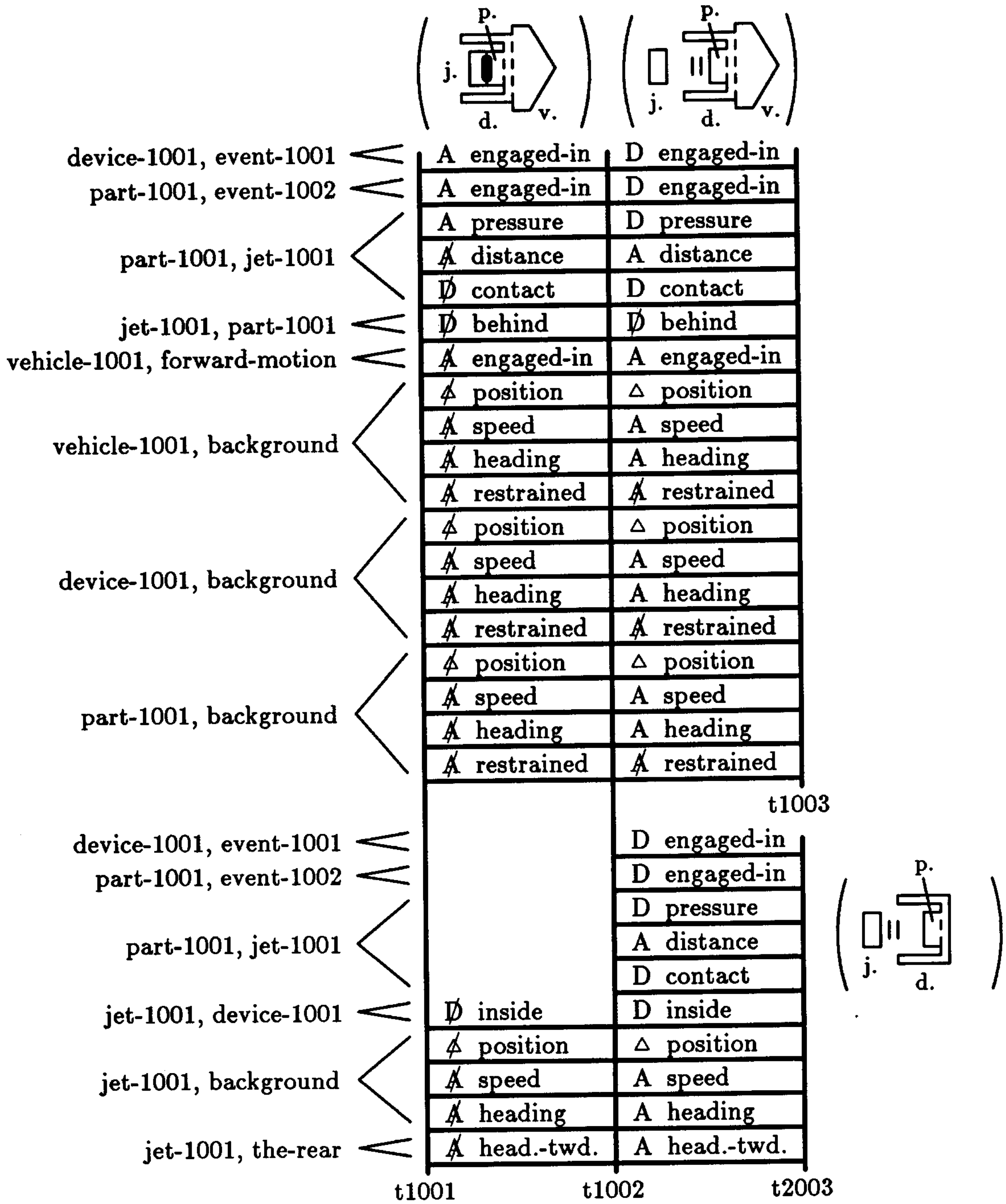
The third source of assistance in constructing association structures is the application of appropriate search heuristics. An informal analysis of about 50 short causal descriptions of physical systems from the *Encyclopedia Americana* suggests the following general strategy. First we move, to the extent possible, from a literal rendering of the specified events to a representation in terms of attributes like "position," "speed," "heading," "contact," "pressure" and so forth, resolving metaphorical descriptions (e.g., possession of forward motion) and moving from abstract representations (involving reified events, for example) to more specialized ones. Next, we search for causal chains of increasing complexity that connect the specified types of events, allowing for partial matches or differences in participants. (In the above example, trace "H" fills this role, providing a path from appearance of pressure to forward motion.) In some cases, there are several disjoint causal chains in a description, or causal cycles to be recognized. Finally, we apply short sequences of operations to tentatively associated event traces in an effort to reconcile their remaining differences.

Once an association structure has been worked out for a description, the causal "path" itself—that is, the overall chain of events—may be assembled in the form of a composite trace of the activity, as illustrated in Figure 7 for the rocket example. A composite trace is formed by merging information from the traces appearing in the association structure.[25] The composite trace in Figure 7 includes information from all of the traces appearing in the association structure for the rocket example, with the exception of trace "H," which was drawn from the knowledge base and refers to a prior scenario, and trace "A," whose metaphorical use of the attribute "possess" is supplanted by the description provided in trace "B."[26]

Given the association structure and composite trace as described above, plus a knowledge base of event traces depicting causal scenarios for a range of specific, physical transactions, we may envision a related process involved in answering various questions regarding causal relationships, analogies and levels of abstraction. Concerning causality, such questions may call for the enumeration of predictions, explanations or sources of failure for particular subevents within the described scenario. For instance, interaction between the force of propulsion and the force of gravity may alter the vehicle's heading. Asked why the vehicle moves forward, individual links in the association structure may be recounted; for example, the vehicle moves forward because the device moves forward and it is assumed that the device is a rigidly attached part of the vehicle. Sources of failure may be iden-

[25]The merging of two traces to form a third trace may be thought of as a further variant of the "composition" operations described in the previous section.

[26]In the composite trace, the time points "t2001," "t2002" and "t3001" have been equated with "t1001" through "t1003," in line with the connectivity of associations in the association structure. Point "t2003" remains distinct from "t1003" however, as the two are not associated in the association structure.



MATCH(heading, < vehicle-1001, background >, forward, t1003)
MATCH(heading, < device-1001, background >, forward, t1003)
MATCH(heading, < part-1001, background >, forward, t1003)
MATCH(heading, < jet-1001, background >, backward, t2003)

Figure 7: Composite trace for the rocket example.

tified from assumptions made in the course of reconstructing the explanation; the propulsion may fail if the vehicle is restrained or moving backward with respect to the background, or if the device is not connected to the vehicle. Concerning analogy, questions may call for the construction of paraphrases of the overall activity, reassembling the transitions into a different set of event traces. For instance, the above scenario might alternatively be expressed in terms of the device "pushing" on the jet, causing the vehicle to "fly" forward. Concerning abstraction, questions may call for an overall summarization of the causal scenario. For example, the above sequence might be summarized in terms of the device simply "carrying" the vehicle through space.

Generalizing from the above example, we can see three particular strengths of the transition space framework. First, this sort of representation makes explicit the differences between specifications of events at varying levels of abstraction and utilizing a variety of metaphors. This allows us to model the comprehension process starting at a literal, face-value rendering of the events specified in informal causal descriptions. Second, causal associations between events arise implicitly, through the interrelationships appearing between their contained transitions. Thus, we are not forced to enumerate explicitly all of the possible causal interactions between a new event being added to our knowledge base and other events existing previously in the knowledge base: we simply specify the particular changes comprising the new event. Third, varying degrees of partial matching may be employed in the interassociation of events by causality, analogy or abstraction. This allows us to form plausible reconstructions of explanations from causal descriptions that leave out details of the events they specify. As well, assumptions made in the course of partial matching may serve as sources of information regarding possible failures in the described causal scenarios.

5 Discussion

In combining a cognitively-motivated treatment of events and causality with a representation explicit at the level of time and change, the transition space formalism provides a powerful metaphor, allowing us to conceptualize the reconstruction of informal causal explanations in terms of a spatial search process, much the way the state space representation allows us to conceptualize problem solving and planning problems in terms of spatial search. This search process begins with a literal, face-value rendering of the events specified in a description and involves the elaboration of an association structure relating the specified events to one another and to events in the knowledge base via chains of causal, analogical and abstraction associations, these arising from comparisons and transformations over the underlying change-level representations for the events. Given an association structure relating the events specified in an informal causal description, a composite trace of the activity may be constructed, specifying the overall causal chain of events at several levels of abstraction.

Regarding the particular representation of transitions and events outlined in this paper, we can indeed see a number of limitations, generally arising from the discrete, propositional nature of the representation. These limitations include a restricted utility in spatial reasoning (i.e., concerning shapes, sizes, directions, etc.), in assessment of *likelihood* for causal sequences, and in dealing with the *classification* of objects and attributes, of use both in recognizing situations and in constraining the use of analogy. On the other hand, this sort of representation has a benefit of simplicity. In a similar manner, the work in qualitative physics compromises some of the same capabilities for the benefits of simplicity arising from a discrete, propositional representation [Forbus, 1988].

One aspect of the representation requiring further elaboration is the ontology of attributes and prototypes. At present the approach has been to employ, rather freely, terms that appear in everyday language. In a sense, this is as it should be. Given the range of expression possible in everyday language, a representation ought to be able to distinguish between the alternate forms of description (e.g., "moving forward" versus "moving in the forward direction," "possessing forward motion," "engaging in forward motion," etc.). Maintaining such a variety at the representational level, given appropriate rules for translation between the various forms, provides a broader basis from which to explore associations of all types between events. On the other hand, it is also useful to have a restricted, "core" set of attributes and prototypes available, such that causal interconnections, for example, may be worked out in a common representational arena. Ultimately, the nature of such a "core" set must be determined empirically, based on the usefulness of particular quantities in the reconstruction of informal causal explanations.

The construction of a general system for the comprehension of informal causal descriptions of physical systems, stated, say, in simplified English, will require the addition of two critical ingredients, the development of which must go hand in hand. These are: (1) a knowledge base and associated lexicon of sufficient breadth, describing common physical causal scenarios at the level of individual transitions, and (2) a set of search heuristics of general applicability in the reconstruction of informal causal explanations. Starting with an empty knowledge base and lexicon, a suitable approach is to construct an *interactive* comprehension system, which may ask questions in the course of processing informal causal descriptions. Such an effort is currently in progress. While the initial explanations processed by this system may require elaboration in terms of a rather large pyramid of underlying knowledge (presumably bottoming out in a set of simple physical interactions

[Lakoff and Johnson, 1980]), it is expected that subsequent explanations may eventually require less elaboration due to a sharing of knowledge base quantities, as has been proposed in connection with the CYC project [Lenat et al., 1986]. The projection of this process to a general enumeration of quantities and events for use in comprehension of informal causal descriptions is, of course, an extensive enterprise; however, the potential gains of such an endeavor are considerable.

Acknowledgements

I would like to thank Patrick Winston, Randall Davis, David Waltz, Susan Carey and the members of the Learning Group at the MIT AI Lab, in particular Rick Lathrop and Jintae Lee, for helpful comments and guidance in the course of this research.

References

Allen, J.F., "Towards a General Theory of Action and Time," *Artificial Intelligence* 23:2, 1984, 123-154.

Alterman, R., "A Dictionary Based on Concept Coherence," *Artificial Intelligence* 25:2, 1985, 153-186.

Borchardt, G.C., "Event Calculus," *Proc. Ninth International Joint Conference on Artificial Intelligence,* 1985, 524-527.

Bullock, M., Gelman, R. and Baillargeon, R., "The Development of Causal Reasoning," in W.J. Friedman (ed.), *The Developmental Psychology of Time,* Academic Press, 1982.

Collins, A.M. and Loftus, E.F., "A Spreading-Activation Theory of Semantic Processing," *Psychological Review* 82:6, 1975, 407-428.

Davis, R., "Diagnostic Reasoning Based on Structure and Behavior," *Artificial Intelligence* 24:1-3, 1984, 347-410.

Davis, R. and Hamscher, W., "Model-Based Reasoning: Troubleshooting," in Shrobe, H.E. (ed.), *Exploring Artificial Intelligence,* Morgan Kaufmann, 1988, 297-346.

DeJong, G. and Mooney, R., "Explanation-Based Learning: An Alternate View," *Machine Learning* 1:2, 1986, 145-176.

De Kleer, J. and Brown, J.S., "A Qualitative Physics Based on Confluences," *Artificial Intelligence* 24:1-3, 1984, 7-83.

Dyer, M.G., *In-Depth Understanding,* MIT Press, 1983.

The Encyclopedia Americana, International Edition, Grolier Inc., 1989.

Fikes, R.E. and Nilsson, N.J., "STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving," *Artificial Intelligence* 2:3-4, 1971, 189-208.

Forbus, K.D., "Qualitative Process Theory," *Artificial Intelligence* 24:1-3, 1984, 85-168.

Forbus, K.D., "The Logic of Occurrence," *Proc. Tenth International Joint Conference on Artificial Intelligence,* 1987, 409-415.

Forbus, K.D., "Qualitative Physics: Past, Present and Future," in Shrobe, H.E. (ed.), *Exploring Artificial Intelligence,* Morgan Kaufmann, 1988, 239-296.

Genesereth, M.R. and Nilsson, N.J., *Logical Foundations of Artificial Intelligence,* Morgan Kaufmann, 1987.

Gentner, D., "Structure-Mapping: A Theoretical Framework for Analogy," *Cognitive Science* 7:2, 1983, 155-170.

Gentner, D. and Stevens, A.L. (eds.), *Mental Models,* Lawrence Erlbaum Associates, 1983.

Grosz, B.J., "The Representation and Use of Focus in a System for Understanding Dialogs," *Proc. Fifth International Joint Conference on Artificial Intelligence,* 1977, 67-76.

Hammond, K.J., "CHEF: A Model of Case-Based Planning," *Proc. AAAI Fifth National Conference on Artificial Intelligence,* 1986, 267-271.

Hayes, P.J., "The Second Naive Physics Manifesto," in Hobbs, J.R. and Moore, R.C. (eds.), *Formal Theories of the Commonsense World,* Ablex Publ. Co., 1985a.

Hayes, P.J., "Naive Physics I: Ontology for Liquids," in Hobbs, J.R. and Moore, R.C. (eds.), *Formal Theories of the Commonsense World,* Ablex Publ. Co., 1985b.

Holland, J.H., Holyoak, K.J., Nisbett, R.E. and Thagard, P.R., *Induction: Processes of Inference, Learning, and Discovery,* MIT Press, 1986.

Jackendoff, R., *Semantics and Cognition,* MIT Press, 1983.

Johnson, M., *The Body in the Mind,* University of Chicago Press, 1987.

Johnson-Laird, P.N., *Mental Models,* Harvard University Press, 1983.

Kintsch, W., "The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model," *Psychological Review* 95:2, 1988, 163-182.

Kolodner, J.L., "Maintaining Organization in a Dynamic Long-Term Memory," *Cognitive Science* 7:4, 1983, 243-280.

Kolodner, J.L., Simpson, R.L. and Sycara-Cyranski, K., "A Process Model of Case-Based Reasoning in Problem Solving," *Proc. Ninth International Joint Conference on Artificial Intelligence,* 1985, 284-290.

Kuipers, B., "Commonsense Reasoning about Causality: Deriving Behavior from Structure," *Artificial Intelligence* 24:1-3, 1984, 169-203.

Kuipers, B., "Qualitative Simulation," *Artificial Intelligence* 29:3, 1986, 289-338.

Lakoff, G. and Johnson, M., *Metaphors We Live By,* University of Chicago Press, 1980.

Leech, G. and Svartvik, J., *A Communicative Grammar of English,* Longman, 1975.

Lehnert, W.G., "Plot Units and Narrative Summarization," *Cognitive Science* 5:4, 1981.

Lenat, D., Prakash, M. and Shepherd, M., "CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks," *AI Magazine* 6:4, 1986, 65-85.

Leslie, A.M. and Keeble, S., "Do Six-Month-Old Infants Perceive Causality?" *Cognition* 25, 1987, 265-288.

Levin, B. (ed.), *Lexical Semantics in Review,* Lexicon Project Working Paper 1, Center for Cognitive Science, MIT, 1985.

McDermott, D., "A Temporal Logic for Reasoning about Processes and Plans," *Cognitive Science* 6:2, 1982, 101-155.

Michotte, A., *The Perception of Causality,* Translated by T. and E. Miles from French edition, 1946, Methuen, London, 1963.

Miller, G.A. and Johnson-Laird, P.N., *Language and Perception,* Harvard University Press, 1976.

Mitchell, T.M., Keller, R.M. and Kedar-Cabelli, S.T., "Explanation-Based Generalization: A Unifying View," *Machine Learning* 1:1, 1986, 47-80.

Newtson, D. and Engquist, G., "The Perceptual Organization of Ongoing Behavior," *Journal of Experimental Social Psychology* 12, 1976, 436-450.

Newtson, D., Engquist, G. and Bois, J., "The Objective Basis of Behavior Units," *Journal of Personality and Social Psychology* 35:12, 1977, 847-862.

Norvig, P., "Marker Passing as a Weak Method for Text Inferencing," *Cognitive Science* 13:4, 1989, 569-620.

Quillian, M.R., "The Teachable Language Comprehender: A Simulation Program and Theory of Language," *Communications of the ACM* 12:8, 1969, 459-476.

Schank, R.C. and Riesbeck, C.K., *Inside Computer Understanding: Five Programs Plus Miniatures,* Lawrence Erlbaum Associates, 1981.

Shoham, Y. and Goyal, N., "Temporal Reasoning in Artificial Intelligence," in Shrobe, H.E. (ed.), *Exploring Artificial Intelligence,* Morgan Kaufmann, 1988, 419-438.

Shultz, T.R., "Rules of Causal Attribution," *Monographs of the Society for Research in Child Development* 47:1, No. 194, 1982.

Talmy, L., "Force Dynamics in Language and Cognition," *Cognitive Science* 12:1, 1988, 49-100.

Waltz, D.L., "Event Shape Diagrams," *Proc. AAAI Second National Conference on Artificial Intelligence,* 1982, 84-87.

White, P.A., "Causal Processing: Origins and Development," *Psychological Bulletin,* 104:1, 1988, 36-52.

Winston, P.H., "Learning New Principles from Precedents and Exercises," *Artificial Intelligence* 19:3, 1982, 321-350.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
November 1990

3. REPORT TYPE AND DATES COVERED
Memo

4. TITLE AND SUBTITLE
Transition Space

5. FUNDING NUMBERS
N00014-85-K-0124

6. AUTHOR(S)
Gary C. Borchardt

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Artificial Intelligence Laboratory
545 Technology Square
Cambridge, MA 02139

8. PERFORMING ORGANIZATION REPORT NUMBER
AIM 1238

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
Information Systems
Arlington, VA 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AD-A231404

11. SUPPLEMENTARY NOTES
None

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Informal causal descriptions of physical systems abound in sources such as encyclopedias, reports and user's manuals. Yet these descriptions remain largely opaque to computer processing. This paper proposes a representational framework in which such descriptions are viewed as providing partial specifications of paths in a space of possible transitions, or *transition space*. In this framework, the task of comprehending informal causal descriptions emerges as one of completing the specifications of paths in transition space—filling causal gaps and relating accounts of activity varied by analogy and abstraction. Drawing on results from perceptual psychology, a specific representation for events is outlined, along with a set of operations for use in comprehending and answering questions regarding informal causal descriptions. The use of the representation and its operations is illustrated in the context of a simple description concerning rocket propulsion.

14. SUBJECT TERMS
knowledge representation
informal causal descriptions
analogy
natural language
causal reasoning
abstraction

15. NUMBER OF PAGES
41

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
unclassified	unclassified	unclassified	unclassified

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

CS-TR Scanning Project
Document Control Form

Date : 10/21/94

Report # AIM-1238

Each of the following should be identified by a checkmark:
Originating Department:

- [x] Artificial Intellegence Laboratory (AI)
- [ ] Laboratory for Computer Science (LCS)

Document Type:

- [ ] Technical Report (TR)
- [x] Technical Memo (TM)
- [ ] Other:

Document Information

Number of pages: 41
Not to include DOD forms, printer intstructions, etc... original pages only.

Originals are:

- [ ] Single-sided or
- [x] Double-sided

Intended to be printed as :

- [ ] Single-sided or
- [x] Double-sided

Print type:

- [ ] Typewriter
- [ ] Offset Press
- [x] Laser Print
- [ ] InkJet Printer
- [ ] Unknown
- [ ] Other:

Check each if included with document:

- [x] DOD Form ORG
- [ ] Funding Agent Form
- [ ] Cover Page
- [ ] Spine
- [ ] Printers Notes
- [ ] Photo negatives
- [ ] Other:

Page Data:

Blank Pages(by page number):

Photographs/Tonal Material (by page number):

Other (note description/page number):

Description : Page Number:

Scanning Agent Signoff:

Date Received: 10/21/94 Date Scanned: 10/26/94 Date Returned: 10/27/94

Scanning Agent Signature: Michael W. Cook

Rev 9/94 DS/LCS Document Control Form cstrform.vsd