Understanding Causal Descriptions of Physical Systems*

Gary C. Borchardt
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

This paper introduces the *causal reconstruction* task—the task of reading a causal description of a physical system, forming an internal model of the specified behavior, and answering questions demonstrating comprehension and reasoning on the basis of the input description. A representation called *transition space* is introduced, in which events are depicted as path fragments in a space of "transitions," or complexes of changes in the attributes of participating objects. By identifying partial matches between the transition space representations of events, a program called PATHFINDER is able to perform causal reconstruction on short causal descriptions presented in simplified English. Simple transformations applied to event representations prior to matching enable the program to bridge discontinuities arising from the writer's use of analogy or abstraction. The operation of PATHFINDER is illustrated in the context of a simple causal description extracted from the *Encyclopedia Americana*, involving exposure of film in a camera.

Introduction

Causal descriptions of the sort appearing in encyclopedias, reports and user manuals comprise an important source of knowledge about the behavior of physical systems. In circumstances where complex interactions, intuitive concepts or metaphorical understanding are involved, such descriptions often constitute the only way in which humans can express what they know about a causal situation. In this paper, I address the problem of getting programs to understand and reason on the basis of such descriptions.

Consider the following excerpt from the *Encyclopedia Americana* [1989]:

> **CAMERA.** The basic function of a camera is to record a permanent image on a piece of film. When light enters a camera, it passes through a lens and converges on the film. It forms a latent image on the film by chemically altering the silver halides contained in the film emulsion.

Given this description, a human previously unfamiliar with the operation of a camera should be able answer non-trivial questions such as the following:

What happens to the distance between the light and the film? (This distance decreases, then disappears as the light converges on the film.)

How does the light "converging on the film" relate to the light "forming the image on the film?" (The former leads to chemical alteration of the silver halides, which change appearance, constituting the latter.)

How could a building reflecting light into the camera cause the light to converge on the film? (This event ends with light entering the camera, from which it passes through the lens and converges on the film.)

Does the light come into contact with the film emulsion? (Yes. The light contacts the silver halides as it chemically alters them, and as these are a part of the emulsion, it therefore contacts the emulsion.)

Answering such questions involves: (1) recognizing unstated associations between events in the description and questions, and (2) using these to model the activity at the level of time-varying attributes of objects.

Causal Reconstruction

I call this task *causal reconstruction.* Causal reconstruction involves reading a causal description, forming an internal model of the activity, and answering questions testing for comprehension of the description. This task differs in two ways from the related task of *causal modeling* [Pearl and Verma, 1991; Doyle, 1989]. First, an initial causal model is already known to the writer of the description. Second, the writer's communication of this model is governed by conversational constraints; e.g., from Grice [1975]: (1) (The Maxim of Quantity) provide enough information but not too much, (2) (The Maxim of Quality) be truthful, (3) (The Maxim of Relation) provide relevant information, and (4) (The Maxim of Manner) be perspicuous.

From these considerations, we may state specific criteria for assessing success at causal reconstruction. Assuming that the *comprehender's* model must *also* be describable by the supplied description, we pose ques-

*This research was supported in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-85-K-0124.

tions to the comprehender, evaluating its responses to see if Grice's maxims would forbid use of the initial description as an account of this new model of activity. From the Maxim of Quantity: (1) does the new model introduce objects/events not motivated in the description (posing the description as underinformative)? From the Maxim of Quality: (2) does the new model disagree in any way with the description (posing it as untruthful), or (3) is the new model physically unrealizable (posing the description as vacuous)? From the Maxim of Relation: (4) does the new model fail to incorporate any information in the description (posing the description as containing irrelevant information), or (5) is the new model not fully connected (posing the description as containing one or more unrelated events)? From the Maxim of Manner: (6) does the new model make any piece of information in the description redundant (such that it could be condensed)?

This task characterization suffices for human or machine comprehension. Note, however, that Grice's Maxims depend on the intended audience of an utterance; e.g., in describing a situation to a child, one must include more information. For automated comprehension of causal descriptions, it is perhaps simplest to stipulate an absence of relevant background knowledge on the part of the program. Thus, we require definitions for events, static properties of objects, rules of inference and so forth to accompany an input description. This simplifies the task, but by no means trivializes it. Given the pieces of a puzzle, the program must still determine how these pieces fit together.

Transition Space

I now introduce a representation called *transition space,* supporting causal reconstruction in the program PATHFINDER. This representation describes events as paths in a space of "transitions," or complexes of changes in attributes for objects participating in a described scenario.[1] Such a representation finds motivation in the perceptual psychology literature [Michotte, 1946; Miller and Johnson-Laird, 1976], and is broadly consistent with research in qualitative reasoning [Forbus, 1984; de Kleer and Brown, 1984; Kuipers, 1986].

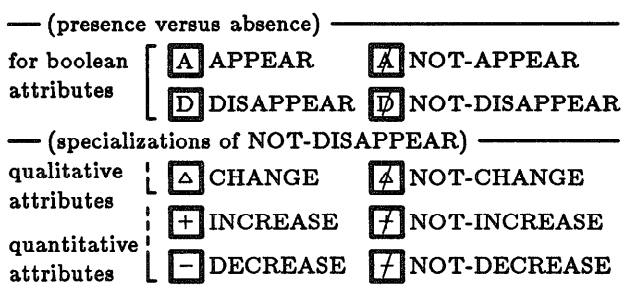
In contrast with the work in qualitative reasoning, transition space relies on *language* for attributes and their changes. Examples of assertions characterized by this representation appear below (attributes appear in boldface, indications of change in italics).

The **contact** between the steam and the metal plate *appears.*
The **concentration** of the solution *increases.*
The **appearance** of the film *changes.*
The pin *becomes* **a part of** the structure.
The water *remains* **inside** the tank.

[1] This representation is related to and in part based on previous representations described in Waltz [1982] and Borchardt [1985].

Miller and Johnson-Laird [1976] enumerate a large number of such attributes and characterize them as typically quantitative or qualitative (including boolean), and typically unary or binary.

Assuming an "absent" or "false" value in the range of each attribute, then if we know whether or not a specific attribute of one or more objects is present at each of two time points, one of which follows the other, and we know the qualitative relationship between the values at these two time points, then the following ten change characterizations are exhaustive, though overlapping. (The accompanying mnemonic symbols are used in a graphic representation described below.)



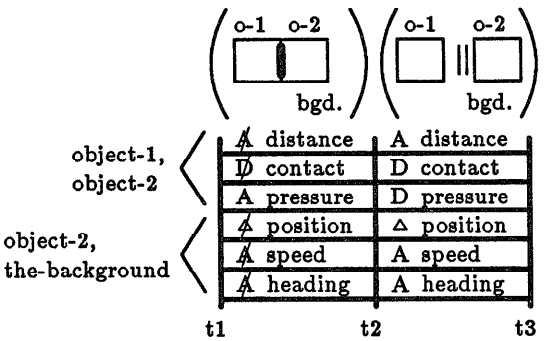
These characterizations are depicted as predicates taking four arguments: an attribute of concern, an object or tuple of objects, and two time points. The assertions below correspond to each of the English statements appearing above.

APPEAR(contact, ⟨the-steam, the-metal-plate⟩, t1, t2)
INCREASE(concentration, the-solution, t3, t4)
CHANGE(appearance, the-film, t5, t6)
APPEAR(a-part-of, ⟨the-pin, the-structure⟩, t7, t8)
NOT-DISAPPEAR(inside, ⟨the-water, the-tank⟩, t9, t10)

Two primitive predicates, EQUAL and GREATER, plus their negations, form a basis for defining these ten change characterizations. Additionally, six predicates are defined for assertions at a single time point. The definitions are omitted here, but may be found in [Borchardt, 1990] and [Borchardt, 1992].

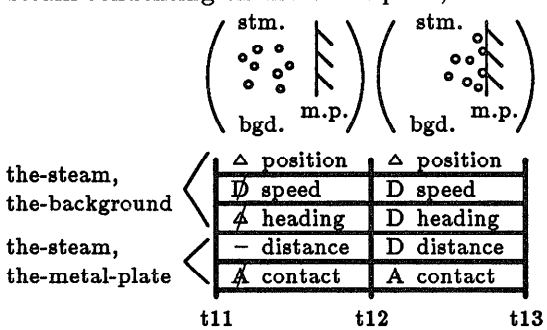
A *transition* is a set of assertions at and between two ordered time points. Events are sequences (or, more generally, directed acyclic graphs) of transitions. These representations are called *event traces,* as they correspond to simple paths in transition space.

While the underlying representation remains propositional, a simplified graphic format will be used here. The following diagram illustrates this format for an event trace depicting the event "push away." Only dynamic (across-time) information is portrayed, with the ten change characterizations coded using the mnemonic symbols specified above. Also, for ease of (our) visualization, a drawing is placed above each transition in the event. This event trace contains two transitions, the first corresponding to appearance of pressure between "object-1" and "object-2," the second, motion of "object-2" with respect to the background and parting of contact between the two objects.

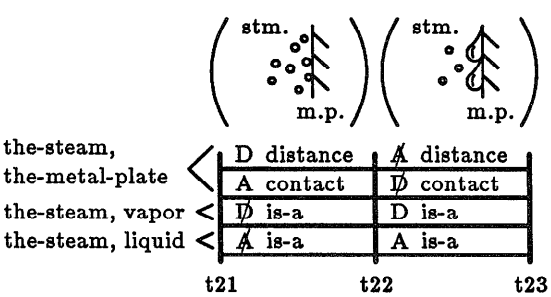


Matches Between Event Traces

Given a set of event traces for the events in a causal description, simple inter-event associations may be detected by identifying partial matches between the traces. As an example, suppose we are given two events, steam moving into contact with a metal plate, and steam condensing on the metal plate, as follows.



The steam moves into contact with the metal plate.

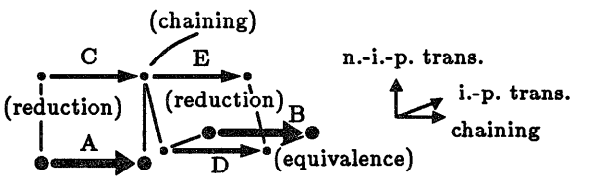


The steam condenses on the metal plate.

Partial matches are grouped into two classes: (1) *partial chaining* matches, where a non-initial transition in one trace partially matches the initial transition in another, and (2) *partial restatement* matches, where the traces match in some other way.[2] For the above traces, there are two possible partial chaining matches and three possible partial restatement matches (first transitions only, second only, both transitions). Of these possibilities, a single partial chaining match appears, leading from the first trace to the second and involving a disappearance of distance and appearance of contact between the steam and the metal plate.[3]

The match identified here is a *partial association*, as the transitions in question do not match completely. By performing simple transformations on the event traces, we may bring them into a complete match. Here, we distinguish between two classes of transformations: *information-preserving* and *non-information-preserving*. The former are members of inverse pairs of transformations in transition space; the latter are not.

A set of event traces linked by *complete associations*—transformations and complete matches—comprises an *association structure*. Association structures are diagrammed using a more abstract visual format. Here, event traces appear as arrows (or more generally, DAGs), with associations represented by alignment in three dimensions: horizontal for complete chaining associations, vertical for non-information-preserving transformations, and depthwise for information-preserving transformations. The following diagram illustrates an association structure elaborating the partial match discussed above.



A: The steam moves into contact with the metal plate.

B: The steam condenses on the metal plate.

The chain of associations may be summarized as follows. First, we perform an information-preserving transformation on the second trace, B, replacing time points "t21" and "t22" with their matched equivalents "t12" and "t13" (see previous illustration). This produces trace D. Next, we remove information concerning the attribute "is-a" from trace D (a non-information-preserving transformation), producing E, and likewise remove assertions involving "position," "speed" and "heading" from A, producing C. Finally, C and E are linked by a complete chaining association.

Inference and Background Statements

Symmetric, transitive and other properties of the primitives EQUAL and GREATER, plus properties of particular attributes can be mechanized into an inference step augmenting event traces with relevant new assertions. This provides additional material for matching. Background statements supplied with a description (e.g., "The water is inside the tank.") can assist in this process. Separately, inference can be used to test partial matches for logical consistency.

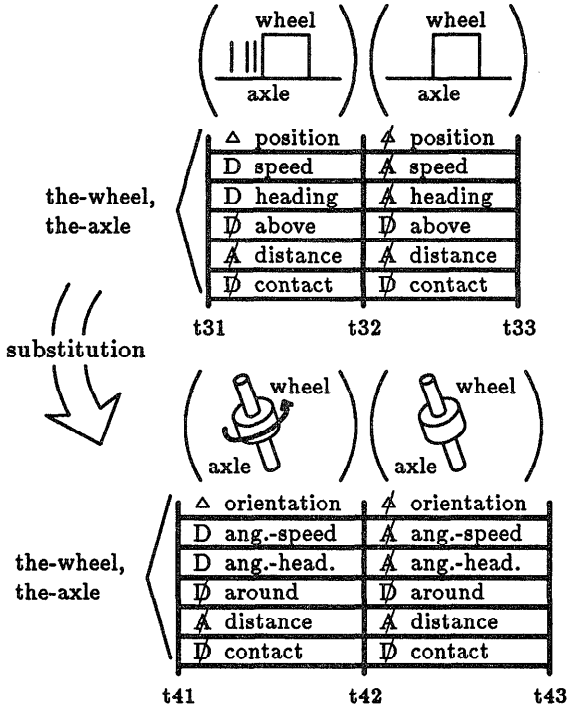
[2] This distinction is refined slightly in [Borchardt, 1992].

[3] Heuristics for ranking alternative partial matches are listed in the section entitled "PATHFINDER."

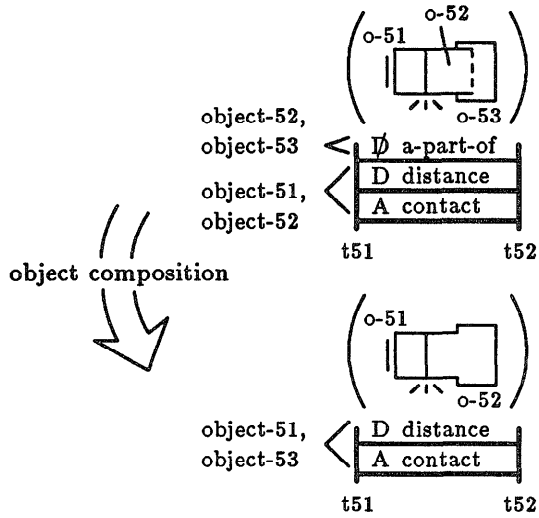
Exploratory Transformation of Event Traces

In addition, transformations of both varieties can be applied in an exploratory manner to form alternate characterizations of events at different levels of abstraction or in terms of different underlying metaphors. An event trace together with its transformed images forms a small "cluster" of traces, all of which participate in the association process. In this manner, we may bridge discontinuities arising from the writer's use of analogy or abstraction.

The following traces depict a simple exploratory information-preserving transformation (of type *substitution*). If we are told that an object "slides to a stop," it is natural to represent this by the first trace. For a rotating object like a wheel, however, a substitution of attributes taking us into the domain of spinning objects may be more appropriate. By including both traces in the association process, we can determine by matching which interpretation is correct.



Below is an exploratory non-information-preserving transformation (of type *object composition*). Suppose one object is specified as coming into contact with a second object, and the second object is a part of a third object. An alternate, more abstract characterization of the event portrays the first object as simply coming into contact with the third object. Such a situation arises in processing the camera description (as discussed below): light comes into contact with the silver halides as part of chemically altering them, yet this must be matched with light "converging on the film" which contains the silver halides.



Additional exploratory non-information-preserving transformations appear in [Borchardt, 1990] and [Borchardt, 1992]. These include: *generalization*, in which a reference term is replaced by a subsuming term; two additional varieties of composition (*interval composition, attribute composition*), and three varieties of reification (*attribute-object reification, event-attribute reification*, and *event-object reification*).

PATHFINDER

PATHFINDER is a 20,000 line program coded in Common Lisp and running on a Symbolics 3640. It consists of a parser operating on a simple, context free skeleton of English grammar, a simple language generation capability, a toolbox for representing, matching and conducting inference and transformations on events in transition space, and additional facilities for causal reconstruction. PATHFINDER has been applied to over 60 causal descriptions, most involving 2–4 events, in a wide range of physical domains including interaction between solid objects and liquids, condensation and melting, combustion, radio signals, light, chemical reactions and electric currents.

All input to PATHFINDER consists of statements in simplified English. (A sample appears in Figure 1, below.) First, PATHFINDER is given a causal description, consisting of (1) *event references* ("The light enters the camera."), (2) *background statements* ("The head is a part of the nail.") and, in some cases, (3) *explicit meta-level statements* ("The device starting to move causes the lever to start to move."). Next, a set of supplementary information is provided, possibly including (1) *additional background statements*, (2) *event definitions*, (3) *precedent events*, which may be of use in reconstructing the activity, (4) *rules of inference*, and (5) *rules of restatement*, including specifications of analogical mappings and rules of abstraction.

Given input in this form, PATHFINDER performs causal reconstruction in four phases. First, it uses the event definitions to form event traces for all events

(a) *(the causal description in simplified English)*

```
The camera records the image on the film. The recording of the image is a function of the camera.
The light enters the camera. The light passes through the lens. The light converges on the film.
The light forms the image on the film. The light chemically alters the silver halides. The silver
halides are contained in the emulsion. The emulsion is a part of the film.
```

(b) *(an event definition for "entering," involving physical objects)*

```
Object 11 entering object 12 translates to the following event. Concurrently, object 11 remains a
physical object, object 12 remains a physical object, object 12 remains hollow, the position of
object 11 changes, the speed of object 11 does not disappear, the heading of object 11 does not
change, and object 11 becomes inside object 12.
```

(c) *(a precedent event: change of appearance during chemical transformation)*

```
Object 61 changes appearance from chemical transformation.

Object 61 changing appearance from chemical transformation translates to the following event.
Concurrently, object 61 remains a physical object, object 61 becomes not made of substance 62,
object 61 becomes made of substance 63, and the appearance of object 61 changes.
```

(d) *(a restatement rule: light viewed as a physical object with respect to "contact")*

```
Concurrently, quantity 141 is a beam of light, object 142 is a physical object, and the contact
between quantity 141 and object 142 is present. The following statement parallels the preceding
statement. Concurrently, object 151 is a physical object, object 152 is a physical object, and the
contact between object 151 and object 152 is present.
```

(e) *(a restatement rule: contact with a part summarized as contact with whole)*

```
Concurrently, object 201 remains a part of object 202, the distance between object 203 and object
201 disappears, and the contact between object 203 and object 201 appears. The preceding statement
is summarized by the following statement. Concurrently, the distance between object 203 and object
202 disappears, and the contact between object 203 and object 202 appears.
```

Figure 1: Input text for the camera description (partial).

referenced in the description. Second, it extends the traces through inference and applies exploratory transformations—these motivated by rules of restatement in the input—producing for each event a cluster of traces describing that activity in different ways. Third, it constructs an agenda of partial matches identified between traces in different clusters. Iteratively choosing the top-ranked partial match and elaborating it in the manner illustrated above, it associates the clusters together. Inference is used to check each selected partial match for consistency. Fourth, when all of the events have been associated, it answers questions (described in the next section).

Several heuristics are used in choosing partial matches for elaboration. These heuristics are listed below. The first heuristic is given strongest influence over the selection process.

Matching between transitions. Definite changes (APPEAR, DISAPPEAR, CHANGE, INCREASE, DECREASE) are weighted most, other dynamic assertions next and static assertions least.

Proximity to description events. Penalties are introduced for matches involving precedent events or exploratory transformations of events.

Narrative ordering. Preference is given to chaining matches between events referenced consecutively in the description.

Current status of the association structure. Penalties are introduced for matches providing a second antecedent or consequent for an event, matches between events already connected via associations, and matches involving hypothesized objects (e.g., a conjectured part of a physical object).

Meta-level considerations. Restatement matches are penalized slightly, and matches violating explicit meta-level statements are penalized heavily.

Processing the Camera Description

Figures 1 and 2 contain an abbreviated session transcript for PATHFINDER regarding the camera description introduced in the Introduction. Input to PATHFINDER appears in Figure 1; question-answering appears in Figure 2.

Details of processing for this example are given in [Borchardt, 1992]. The association phase executes six iterations, selecting from 120 candidate partial matches on the first iteration. The diagram below illustrates a part of the association structure generated by PATHFINDER, connecting "The light converges on the film." with "The light chemically alters the silver halides.". Traces M, W and Z (outlined) have been formed by exploratory transformations. For M and W, an information-preserving transformation derived from the rule of restatement shown in Figure 1 (d) has

(a) *(a question of type 1: describing the time-varying status of a particular attribute)*

```
What happens to the distance between the light and the film?

  First, as the light passes through the lens, the distance between the light and the film decreases.
  Next, as the light converges on the film, the distance between the light and the film disappears.
  Next, as the camera records the image on the film, the distance between the light and the film does
  not appear.
```

(b) *(a question of type 2: describing the interrelationship between two events)*

```
How does the light converging on the film relate to the light forming the image on the film?
  The light converging on the film causes the light to chemically alter the silver halides, which ends
  with the silver halides changing appearance from chemical transformation, which occurs at the end of
  the light forming the image on the film.
```

(c) *(a question of type 3: identifying a plausible causal connection)*

```
How could the building reflecting the light into the camera cause the light to converge on the film?

  The building reflecting the light into the camera could end with the light entering the camera, which
  could cause the light to pass through the lens, which could cause the light to converge on the film.
```

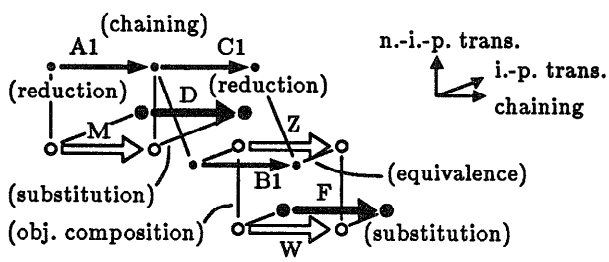
(d) *(a question of type 4: restating a portion of the activity)*

```
Does the light come into contact with the emulsion?

  Yes. The light coming into contact with the emulsion is a part of the light converging on the film.
```

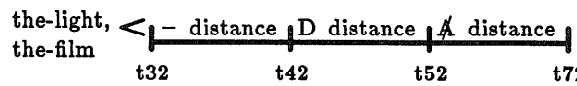
Figure 2: Question answering for the camera description.

recast activity involving a physical object with activity involving a beam of light. For Z, a non-information-preserving transformation derived from the rule of restatement shown in Figure 1 (e) has recast light contacting the silver halides as light contacting the film. A partial chaining match has been identified between traces M and Z, both of which specify the light contacting the film. Finally, traces A1, B1 and C1 have been produced by elaboration of this partial chaining match (involving an equivalence mapping from Z to B1, removal of assertions from M to A1 and from B1 to C1, and chaining from A1 to C1).



D: The light converges on the film.
F: The light chemically alters the silver halides.

PATHFINDER uses the association structure to answer four types of questions. The first type concerns the time-varying status of particular attributes of objects; e.g., "What happens to the distance between the light and the film?" To answer this, PATHFINDER merges the traces in the association structure, overlapping where indicated, extracts the portion relevant to the question, and expresses it in simple English. For the above question, the relevant portion is as below. PATHFINDER's response appears in Figure 2 (a).



The second type of question concerns inter-event relationships; e.g., "How does the light converging on the film relate to the light forming the image on the film?" To answer this type of question, PATHFINDER extracts the relevant path in the association structure and describes this path in simple English, highlighting important associations (see Figure 2 (b)).

The third and fourth types of questions also ask about inter-event relationships, but require PATHFINDER to do further association first. Supplementary information (e.g., event definitions) may be provided with these questions. The third type of question involves plausible causal associations; e.g., "How could the building reflecting the light into the camera cause the light to converge on the film?" (Figure 2 (c)). The fourth type asks if a new event may paraphrase part of the activity; e.g., "Does the light come into contact with the emulsion?" (Figure 2 (d)).

Discussion

This work is motivated by a range of research, as summarized in [Borchardt, 1992]. However, very little work has addressed the task of understanding written causal descriptions of physical systems. Rieger [1976] proposed a mechanism for such understanding, but his approach lacked an explicit notion of time and was never fully implemented and tested. More recently, Sembugamoorthy and Chandrasekaran [1986] and Bylander and Chandrasekaran [1985] provide interesting accounts of causal physical behavior that are consistent with human intuition. Their research targets the task of reasoning using knowledge entered directly in a

representational format, however, rather than the task of extracting causal knowledge from written material.

Several differences exist between the transition space representation and that used in qualitative physics, these arising primarily from differences in the problems being addressed. As noted above, representations for individual events are grounded in language, rather than a scientific model of a physical system. As a result, the transition space representation tends to be more macroscopic, with events often spanning several qualitative states of a device. Additionally, transition space explicitly represents differences in the *description* of events at alternate levels of abstraction or in terms of different underlying metaphors. Finally, the mechanism for reasoning is different, consisting of heuristic search rather than constraint propagation.

Work in spreading activation [Quillian, 1969; Alterman, 1985; Norvig, 1989] has addressed the problem of recognizing inter-event associations in natural language text. However, little of this work deals with reconstructing sequences of physical causation. I suspect that this is because inferring plausible causal links between physical events is very context dependent (e.g., a dropped object will fall, but only if it is not otherwise supported). Event traces in transition space can capture this context, whereas doing so in a semantic network would require considerable bifurcation of event nodes into sub-nodes depicting special cases.

The transition space representation works best when the activities are not all of the same type (e.g., all translational motion). In such cases, it is expected that incorporation of a more finely-grained spatial representation may be required. Other useful extensions include abstraction shifts for elaborating or summarizing repetetive events and feedback cycles, a means of estimating *likelihood* for causal sequences, and a means of classifying objects and events.

On the basis of descriptions processed by PATHFINDER, the heuristic of matching of transitions appears to be quite useful in causal reconstruction. The transition space representation is also easy to generate from simple, stylized verbal accounts of what happens during events. Since the representation is grounded in the variety of changes expressible in simple language, it is quite possible that this representation may find utility in other domains as well, beyond the current focus on physical systems.

Acknowledgements

I thank Patrick Winston, Randall Davis, David Waltz, Susan Carey and the members of the Learning Group at the MIT AI Laboratory—in particular, Rick Lathrop, Jintae Lee and Lukas Ruecker—for guidance and helpful criticism in the course of this research.

References

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.

Borchardt, G. C., "Transition Space," A.I. Memo 1238, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 1990.

Borchardt, G. C., *Causal Reconstruction: Understanding Causal Descriptions of Physical Systems,* Ph.D. Dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 1992, forthcoming.

Bylander, T. and Chandrasekaran, B., "Understanding Behavior Using Consolidation," *Proc. Ninth International Joint Conference on Artificial Intelligence,* 1985, 450-454.

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

Doyle, R. J., "Reasoning About Hidden Mechanisms," *Proc. Eleventh International Joint Conference on Artificial Intelligence,* 1989, 1343-1349.

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

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

Grice, H. P., "Logic and Conversation," in Cole, P. and Morgan, J. L. (eds.) *Syntax and Semantics, Volume 3: Speech Acts,* Academic Press, 1975, 41-58.

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

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.

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

Pearl, J. and Verma, T. S., "A Theory of Inferred Causation," *Proc. Second International Conference on Principles of Knowledge Representation and Reasoning,* 1991, 441-452.

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

Rieger, C., "An Organization of Knowledge for Problem Solving and Language Comprehension," *Artificial Intelligence* 7:2, 1976, 89-127.

Sembugamoorthy, V. and Chandrasekaran, B., "Functional Representation of Devices and Compilation of Diagnostic Problem-Solving Systems," in Kolodner, J. and Riesbeck, C. (eds.) *Experience, Memory, and Reasoning,* Lawrence Erlbaum Associates, 1986, 47-73.

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