Advances in Cognitive Systems 2 (2013) 1–18 Submitted 9/2013; published 12/2013

Reasoning from Radically Incomplete Information: The Case of Containers

Ernest Davis DAVISE@CS.NYU.EDU
Computer Science Dept., New York University, 251 Mercer St. New York, NY 10012 USA

Gary Marcus GARY.MARCUS@NYU.EDU
Psychology Dept., New York University, New York, NY 10012 USA

Angelica Chen AC17@PRINCETON.EDU
College of Arts and Science, Princeton University, Princeton, NJ, 08540 USA

Abstract

In physical reasoning, humans are often able to carry out useful reasoning based on radically incomplete information. One physical domain that it is ubiquitous both in everyday interactions and in many kinds of scientific applications, where reasoning from incomplete information is very common, is the interaction of containers and their contents. We have developed a preliminary knowledge base for qualitative reasoning about containers, expressed in a sorted first-order language of time, geometry, objects, histories, and events. We have demonstrated that the knowledge suffices to justify a number of commonsense physical inferences, based on very incomplete knowledge.

1. Physical Reasoning Based on Radically Incomplete Information

In physical reasoning, humans, unlike programs for scientific computation, are often able to carry out useful reasoning based on radically incomplete information. If AI systems are to achieve human levels of reasoning, they must likewise have this ability. Extant automated reasoners based on simulation cannot fully address the challenges of radically incomplete information (Davis & Marcus, 2013); rather such challenges require alternative reasoning techniques specifically designed for incomplete information.

As a vivid example, consider the human capacity to reason about containers — boxes, bottles, cups, pails, bags, and so on — and the interactions of containers with their contents. For instance, you can reason that you can carry groceries in a grocery bag and that they will remain in the bag with only very weak specifications of the shape and material groceries being carried, the shape and material of the bag, and the trajectory of motion. Containers are ubiquitous in everyday life, and children learn to use containers at a very early age (figure 1).

Containers likewise are central in a wide range of applications and domains. For example, in a separate study we have recently begun of the reasoning needed to understand a biology textbook (Reece, et al., 2011), we find that containers of many different kinds and scales (literal, physical containers, not even including metaphorical containers) appear in domains relevant to biology. Some examples:

- The membrane of a cell is a container that holds the contents of the cell. Many of the primary processes in the cell are concerned with bringing material into the container and expelling material from the container.
- The skin or other outer layer of an animal is a container for the animal. Again, many of the central life processes — eating, breathing, excreting — deal with transporting material into and out of the container.
- In a discussion of speciation, it is mentioned that a subpopulations of a water creature can be isolated if the water level of a lake falls, dividing it into two lakes. Here the container is the lake bed, and the phenomenon depends on the somewhat non-obvious fact that a liquid container that bounds a single connected region at one level may bound two regions at a lower level (figure 2).

Figure 1: Infant learning about containers

Figure 2: A lake divides into two lakes when the water level falls

In this paper we describe the initial stages of development of a knowledge-based system for reasoning about manipulating containers, in which knowledge of geometry and physics and problem specifications are represented by propositions. Below, we outline the system, and show that in skeletal form this approach suffices to justify a number of commonsense physical inferences, based on very incomplete knowledge of the situation and of the dynamic laws that govern the objects involved. We have implemented one of these inferences in the first-order theorem proving system SPASS (Weidenbach, et al., 2009).

1.1 Incomplete information

The issues of complete and incomplete information can easily be misunderstood, so let us make clear what we have in mind. Of course, few representations are truly complete or entirely precise; in virtually any representation, some aspects are omitted, some are simplified, and some are approximated. However, techniques such as simulation, or STRIPS-like representations, require that the initial conditions of the scenario and that the dynamics of the microworld *be fully specified* relative to a given level of description. That is, the representational framework specifies some number of critical relations between entities and properties of entities. A complete representation of a situation relative to that framework enumerates all the entities that are relevant to the situation, and specifies all the relations in the framework that hold between those entities. The description must be detailed and precise enough that the situation at the next time step is likewise fully specified, in the same sense.

For instance, the standard blocks world representation omits the size, shape, and physical characteristics of the blocks involved, and the trajectory of the actions. Situations are describe purely in terms of the fluent On(x,y) and actions are described in terms of Puton(x,y,z). However, the dynamic theory is a complete account at this level of description; that is, a complete enumeration of the On relations that hold in one situation completely determines what actions are feasible, and determines all the On relations that will hold once the action is executed. Additionally, most projection and most planning problems provide a complete enumeration of the On relations that hold in the initial situation.

2. Containers

We begin with a general discussion of the properties of containers as encountered in everyday situations and of the characteristics of commonsense reasoning about containers.

A container can be made of a wide range of materials, such as rigid materials, paper, cloth, animal body parts, or combinations of these. The only requirement is that the material should maintain its shape to a sufficient degree that holes do not open up through which the contents can escape. Under some circumstances, there can even be a container whose bottom boundary is a *liquid*; for instance, an insect can be trapped in a region formed by the water in a basin and an upside-down cup. A container can also have a wide range of shapes (precise geometric conditions for different kinds of containers are given in section 4.1.)

The material of the contents of a container is even less constrained. In the case of a closed container, the only constraint is that the material of the contents cannot penetrate or be absorbed into the material of the container (e.g. you cannot carry water in a paper bag or carry light in a cardboard box); and that the contents cannot destroy the material of the container (you cannot keep a gorilla in a balsa wood cage). Using an open container requires additionally that the contents cannot fly out the top (Davis, 2011). Using a container with holes requires that the contents cannot fit or squeeze through the holes.

Those are all the constraints. In the case of a closed container, the material of the contents can be practically anything with practically any kind of dynamics. For instance, you can infer that an eel will remain inside a closed fish tank without knowing anything at all about how the mechanisms that eels use to swim or about the motions that are possible for eels.

A container can serve many different purposes, including: carrying contents that are difficult or impossible to carry directly (e.g. a shopping bag or a bottle); ensuring that the contents remain in a fixed place (e.g. a crib or a cage); protecting the contents against other objects or physical influences (e.g. a briefcase or a thermos bottle); hiding the contents from inspection (e.g. an envelope); or ensuring that objects can only enter or exit through specific portals (e.g. a tea-kettle). In some cases it is necessary that some kinds of material or physical effects can either fit through the portals or pass through the material of the container, while others cannot. For instance, a pet-carrying case has holes to allow air to go in and out; a display case allows light to go in and out but not dust.

There are four primary kinds of physical principles involved in all of these cases. First, matter must move continuously; if the contents could be teleported out of the container, as in Star Trek, these constraints would not apply. Second, the contents (or the externality being kept out, such as dust) cannot pass through the material of the container. Third, there are constraints on the deformations possible to the shapes of the container and of the content. Fourth, in the case of an upright open container, gravity prevents the contents from escaping.

Simple, natural examples of commonsense physical reasoning reveal a number of important characteristics.

First, human reasoners can use very partial spatial information. For example, consider the text, "There was a beetle crawling on the inside of the cup. Wendy trapped it by putting her hand over the top of the cup, then carried the cup outside, and dumped the beetle out onto the lawn." A reader understands that the cup and the hand formed a closed container for the beetle, and that Wendy removed her hand from the top of the cup before dumping the beetle. Thus, qualitative spatial knowledge about cups, hands, and beetles suffices for interpreting the text; the reader does not require the geometry of these to be specified precisely.

Second, human reasoners can often infer that a material is confined within a closed container even if they have only a vague idea of the physics of the material of the container and almost no idea at all of the material of the contents. For example, the text above can be understood by a reader who does not know whether a “beetle” is an insect, a worm, or a small jellyfish.

Third, human reasoners can predict qualitative behavior of a system and ignore the irrelevant complex details; unlike much software, they are often very good at seeing the forest and not being distracted by the trees. For example, if you pour water into a cup, you can predict that, within a few seconds it will be sitting quietly at the bottom of the cup; and you do not need to trace through the complex trajectory that the water goes through in getting to that equilibrium state.

Finally, knowledge about containers, like most high-level knowledge, can be used for a wide variety of tasks in a number of different modalities, including prediction, planning, manipulation, design, textual or visual interpretation, and explanation. The container relation is also often used metaphorically; e.g. for the relation between a memory location and a value in computer science.

3. Physical reasoning: Overall architecture.

We conjecture that, in humans, physical reasoning comprises several different modes of reasoning, and we argue that machine reasoning will be most effective if it follows suit. **Simulation** can sometimes be effective; for example, for prediction problems when a high-quality dynamic theory and precise problem specifications are known (Davis & Marcus, 2013) (Battaglia, Hamrick, & Tenenbaum, 2013). An agent can use **highly trained, specialized manipulations and control regimes,** such as an outfielder chasing a fly ball. **Analogy** is used to relate a new physical situation that has some structural similarities to a known situation, such as comparing an electric circuit to a hydraulic system. **Abstraction** reduces a physical situation to a small number of key relations, for instance reducing a physical electric device to a circuit diagram. **Approximation** permits the simplification of numerical or geometric specification; for instance, approximating an oblong object as a rectangular box. Moreover, all of these modes are to some degree **integrated**; if an outfielder chasing a fly ball and a fan throws a bottle onto the field, the outfielder may alter his path to avoid tripping on it.

Where knowledge of the dynamics of a domain or of the specifications of a situation are extremely weak, the most appropriate reasoning mode would seem to be **knowledge-based reasoning**; that is, a reasoning method in which problem specifications and some part of world knowledge are represented declaratively, and where reasoning consists largely in drawing making inferences, also represented declaratively, from this knowledge. Such forms of representation and reasoning are particularly flexible in their ability to express partial information and to use it in many directions.[1] Our objective in this paper is to present a part of a knowledge-based theory of containers and manipulation.

The knowledge-based theory itself has many components at different levels of specificity and abstraction. For example:

- We use a theory of *time* that only involves order relations between instants: time TA occurs before time TB. A richer theory might involve also order relations between durations (duration DA is shorter than DB); or order-of-magnitude relations between durations (DA is much shorter than DB); or a full metric theory of times and durations (DA is twice as long as DB). However, the examples we have considered do not require those.
- Our theory of *spatial and geometrical relations* has a number of different components. For the most part, we use topological and parthood relations between regions, such as "Region RA is part of region RB," "RA is in contact with RB", or "RA is an interior cavity of RB." However we also incorporate a theory of order-of-magnitude relations between the size of regions ("RA is much smaller than RB") and a very attenuated theory of the vertical direction, to enable us to distinguish open containers that are upright from open containers in general.
- Our theory of the *spatio-temporal characteristics of objects* includes the relations "Object O occupies region R at time T'', "Region R is a feasible shape for object O" (that is, O can be manipulated so as to occupy R), and "The trajectory of object O between times TA and TB is history H."

Key concepts that are used at an abstract level of description may not be fully specifiable except in terms of a more concrete level. For example, the full definition of a "rigid object" requires a metric spatial theory that is powerful enough to express the notion of congruence. However, one can assert some of the properties of rigid objects in our qualitative language; for example, if a rigid object is a closed container at one time, it is always a closed container; if it is an open container at one time, it is always an open container. Therefore we include the concept of a "rigid object" in the qualitative level even though the full concept implicitly involves a more powerful geometric theory.

Another, more complex, example: A key concept in the theory of manipulation that the feasibility of moving an object O from place A to place B. It is sometimes possible to show that this action is infeasible using purely topological information; for example, if place A is inside a closed container and B is outside it, then the action is not feasible.

Giving necessary and sufficient conditions, however, is much more difficult. In delicate cases, where one has to rely on bending the object O through a tight passage, reasoning whether it is feasible to move O from A to B or not requires a very detailed theory of the physical and geometric properties both of O and of the manipulator.[2] Moreover, though humans cannot, of course, always do this accurately, because of the frequency and importance of manipulation in everyday life, they are implicitly aware of many of the issues and complexities involved.

---

[1] How knowledge-based reasoning can be implemented in the neural hardware is a difficult problem which we do not attempt to address here. However, we subscribe to the theory (Newell, 1981) that the cognitive processes can be usefully described at the knowledge level in terms of symbolic representations and symbolic reasoning.

[2] Fully detailed physical and spatial *theories* can in principle support inference from radically incomplete *problem specifications* using theorem proving techniques; however, automating these tends to be computationally intractable.

However, at this stage of our theory development, we are not attempting to characterize a complete theory of moving an object, or even of the commonsense understanding of moving an object. Rather, we are just trying to characterize some of the knowledge used in cases where the information is radically incomplete and the reasoning is easy. Therefore, rather than presenting general conditions that are necessary and sufficient, our knowledge base incorporates a number of specialized rules, some stating necessary conditions, and some stating sufficient conditions.

The theory that we envision, and the fragment of it that we have worked out, is frankly neither an elegant system of equations nor a system of necessary and sufficient conditions expressed at a uniform level of description. It is much more piecemeal: there are constraints, there are necessary conditions, there are sufficient conditions; but these are not "tight". Some of these are very general (e.g. two objects do not overlap), others quite specialized. Some require only topological information, some require qualitative metric information, some require quite precise geometric information. Nonetheless, we believe that this is on the right track because it seems to address the problem and reflect the characteristics of radically incomplete reasoning much more closely than any alternative.

More importantly, perhaps, the knowledge base does not conform to any well-defined metalogical framework, along the lines of (Sandewall, 1995) or (Reiter, 1995). Such frameworks, when available, have many advantages: they guide theory construction, guide efficient implementation, and allow the possibility of proving metalogical properties such as consistency or computational complexity. However, neither of these particular systems cited above, nor others that we know of, seem to be particularly well fitted to the kind of domain and inferences that we are working with; and we have not worked out a new framework. At some future date, we would certainly hope to revisit this question and develop such a framework for these kinds of theories. However, in the early stages of theory development, we prefer to focus purely on questions of content, without restricting ourselves by a too-early commitment to a restrictive framework.

The design of this knowledge base must also face the issues of the redundancy of rules and of the level of generality at which rules should be stated. Contrary to common practice in axiomatizing mathematical theories, we have made no attempt to state a minimal collection of axioms, since for our purposes there is no advantage in that. There remains the question, however, of choosing the level of abstraction at which to state the rules, and our choice may strike some readers as leaning implausibly to the abstract side. The motivation for this is to bring out the commonality in different situations.

Consider, for example, the following three facts:

Fact 1: An object inside a solid closed container cannot come out of the container, even if the container is moved around.

Fact 2: In the situation shown in figure 3, the ball must go through the red region before it can reach the green region.

Fact 3: The water in a tea kettle with the lid on can only come out the spout.

It is certainly possible that a human reasoner is applying three entirely separate rules specific to these particular situations. (Undoubtedly, the way in which reasoning is done varies from one person to another, and also changes developmentally.) However, it certainly seem plausible that often people will use the same knowledge in solving all three problems, that they will think of the two problems in the same way, and, if they are presented with all these problems, they will realize that they are similar. An automated reasoner should do likewise. Note, though, that the specifics of the three situations are quite different: in fact 1, there is a single moving object that is a closed container; in fact 2, there is a closed container formed by the union of the solid walls with the purely spatial region marked in red; in fact 3, there is a closed container formed by the kettle plus lid plus an imaginary cork in the spout. To formulate a principle that subsumes both cases, therefore, requires the fairly abstract concept of a *history*, a function from time to regions of space, that can move around (needed for facts 1 and 3) but is not tied to a physical object (needed for facts 2 and 3).

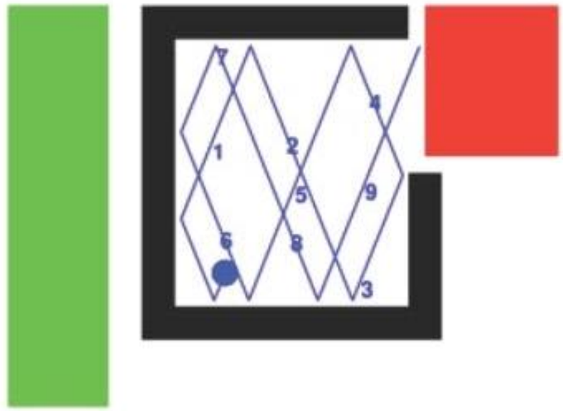


Figure 3: Reasoning about a bouncing ball (from (Smith, Dechter, Tenenbaum, & Vul, 2013))

We use first-order logic with equality as a convenient notation, without at all claiming, either that this is an ideal formalism for an automated system or that it is especially close to cognitive realities. First-order logic has the advantage that it is a standard *lingua franca* (Hayes, 1977), and that there exist standard software inference engines.

3.1 Methodology

Our approach is that of knowledge-based analysis of commonsense reasoning (Hayes, 1979) (Davis, 1998), though with some distinctive features as consequences of the features of the particular domain and of the goal.

A collection of examples of problems in the domain whose solution seem commonsensically obvious form the focus of the study; next, we formulate a *microworld*, a well-defined idealization of the domain, with some limited collection of relations and sorts of entities. This in turn grounds a knowledge base, ensuring that the symbols are being used consistently and that the knowledge base is consistent.

We next formulate an axiomatic system with the following properties:

- The meanings of the symbols can be defined in terms of the microworld.
- The axioms are true in the microworld.
- The axioms justify the solutions to the problems.
- The axioms are reasonably easily stated in first-order logic. In particular, in the knowledge-based system described here, we avoid the use of *axiom schemas*, infinite collections of axioms, such as the principle of induction or the comprehension axiom from set theory.[3] Axiom schemas are certainly problematic in terms of computational efficiency of inference, and perhaps also in terms of cognitive plausibility.

There are also two further desiderata that we try to achieve for the axioms (these two often conflict, so there is a trade-off to be managed). First, symbols should correspond to concepts that

[3] This is the major technical difference between the methodology we use here and that which we have previously used in (Davis, 2008) and (Davis, 2011).

seem reasonably natural in a cognitive model. For instance, ClosedContainer seems plausible; HausdorffDistance, used in (Davis, 2011), seems less so. Second, axioms should be stated at a fairly high-level of generality and abstraction, so that each axiom can be used for many different problems (For simplicity, we have above described our methodology as sequential: first problems, then microworld, then axioms. In practice it is cyclical and iterative. In particular the process of formulating the axioms suggests new problems, improved formulations for old problems, and improvements to the scope and characteristics of the microworld)

Our aim here is not to be fully comprehensive, but rather to establish a baseline. We have not yet incorporated any kind of probabilistic or plausible reasoning, though in the long run, of course, a complete theory of commonsense reasoning would have to include such mechanisms. Likewise, some features of reality go beyond what is considered in the microworld; othera difficult forms of inference, such as induction, are also not included; still, we believe that by systematizing the machinery in this domain, we provide insight into possible mechanisms for coping with incomplete information.

3.2 Evaluation

The difficulties of systematically evaluating such a theory are formidable. As we have argued elsewhere (Davis, 1998), it is in general difficult to evaluate theories of commonsense reasoning in a limited domain, because there is rarely any natural source of commonsense problems limited to a given domain. The class of commonsense physical reasoning problems studied in the AI literature or in the cognitive psychology literature (e.g. (Hegarty, 2004)) is more a reflection of the interests or imagination of the researcher (on the AI side, also a reflection of what is easily implemented) than of the kinds of problems that occur in an ecologically natural setting. The criteria in the methodology do not lend themselves to numerical measures of success, and the iterative nature of theory development means that the goal itself is a moving target.

What we have done is to demonstrate that the symbols and rules in the knowledge base are adequate to express and justify simple commonsensical qualitative inferences, discussed below in section 5.2. We have used the automated theorem prover SPASS (Weidenbach, et al., 2009) to validate the correctness of some of these inferences.

4. The Microworld

Our theory is grounded in a microworld of the following properties:

Time is dense and forward-branching. Branching corresponds to the choice by the agent among different feasible actions.

Space is Euclidean three-space $\mathbb{R}^3$.

Objects are distinct; that is one object cannot be part of another or overlap with another. They are eternal, neither created nor destroyed. They move continuously. Distinct objects cannot overlap spatially. They are not required to be connected. An object occupies a region of some three-dimensional extent (technically, a topologically regular region); it cannot be a one-dimensional curve or two-dimensional surface.

This object ontology works well with solid, indestructible objects. It works much less well for liquids, thought it does not entirely exclude them; better ontologies for liquids are developed in (Hayes, 1985) and (Davis, 2008)

For any object O, there is some range of regions that O can in principle occupy, consistent with its own internal structure; these are called the *feasible* regions for O. For instance, a rigid object can in principle occupy any region that is congruent (without reflection) to its standard shape. A string can occupy any tube-shaped region of a specific length. A particular quantity of liquid can occupy any region of a specific volume.

There is a single agent, which itself is an object. The agent is capable of moving by itself and of directly manipulating one other object. If the object is a container, then its contents move with it. If the agent is holding a box, and there is a lid placed correctly on the box, then the lid will move with the box. Otherwise, the agent cannot use one object to move another.

At any time, if the agent is holding an object, it can release the object. If the object is not in a stable position, it will fall. If the object is a container, the contents will fall with the object; if it is a box with a lid, the lid will fall with the object. Dropping an object does not cause any other objects to move.

4.1 Containers

We distinguish three kinds of geometric containment relations (figure 4):

Region R is a *closed container* for cavity C if C is an interior-connected, bounded component of the complement of R.

Region R is an *open container* for cavity C (a region) if there exists a region A, between two parallel planar surfaces S1 and S2 such that:

- A and R do not overlap. The intersection where they meet $R \cap A$ is equal to the ring around A separating S1 and S2: $R \cap A = Bd(A) \setminus (S1 \cup S2)$.
- C is a cavity of the union $R \cup A$, but is not a cavity of either R or A separately.

Region R is an *upright open container* for cavity C if the planar surfaces S1 and S2 associated with A are horizontal and A is above C.

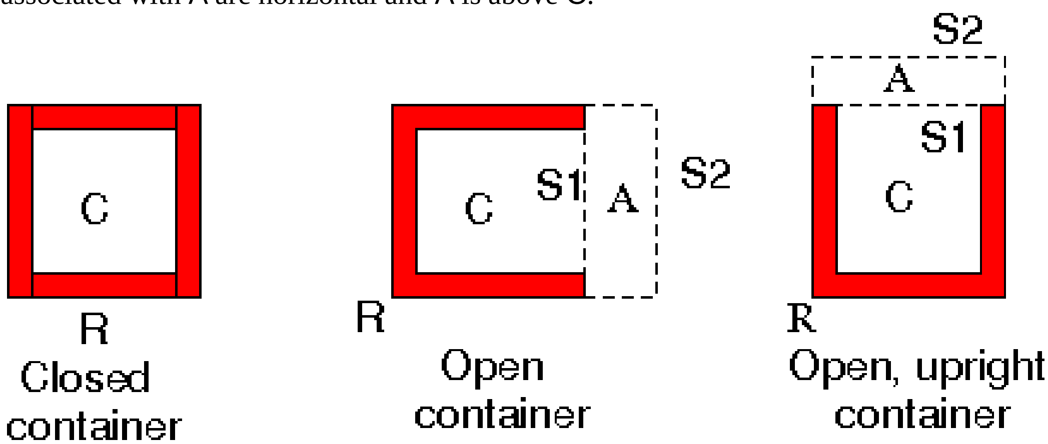


Figure 4: Containers

4.2 Physical laws

At the most general level of our theory, the class of general physical laws is very limited:

- Two distinct objects do not overlap spatially.
- The trajectory of an object is a continuous function of time.
- An object O occupies a region feasible for O. (The axioms that specify what regions *are* feasible for a given object are all at more specific levels of the theory.)
- The agent can hold an object O only if the agent contacts O along an extended face, and can manipulate the object only if he is holding it.
- If the agent is holding an object, he can release it at any time.

- If the agent is holding an object and releases it in an unstable position, then the object will fall for a short period of time. (We do not develop any theory of stability at all; and all that is asserted in our theory of falling is that the object does not move outside any upright container that it is currently in.)
- Object U "goes with" object V if either
    a. U = V.
    b. U is contained in container V.
    c. V is an open box and U is placed as the lid of V.

  In essence, we assume that, in other situations where moving V would cause U to move, the agent cautiously avoids moving V, since the effect on U lie outside the scope of the theory.
- An object O changes its position only under one of the following three conditions:
    a. O is the agent.
    b. The agent is manipulating object V and O goes with V.
    c. Object V is falling and O goes with V.

  Technically, this rule is a frame axiom for the place of object O using explanation closure (Schubert, 1990).
- As discussed below, there are some specialized feasibility axioms, giving sufficient conditions for certain kinds of actions to be feasible.

4.2 Theory of Action

At the most abstract level of our theory, we consider only two actions: Travel(r), the agent travels to occupy a specified region r without moving any other object, and MoveTo(o,r), the agent moves object o to occupy region r.[4] There is additionally an event Falling(o), which is triggered by the agent releasing object o in an unstable position. The basic causal rules for the domain are stated in terms of these actions and events, as are the frame axioms, such as the frame axiom for the positions of objects discussed above. However, if the geometric language is limited to purely topological relations, then only a very limited dynamic theory of these events can be constructed. In particular, as discussed above, it does not seem to be possible to formulate in a purely topological language any preconditions for Travel or MoveTo that are sufficient to guarantee the feasibility of the action.

Therefore, we additionally develop a collection of more specialized actions for which more adequate qualitative theories can be formulated. The actions we have defined include CarrySimple(o,r), CarryClosed(o,r), CarryUpright(o,r), CarryBoxWithLid(ob,ol,r), PutLidOnBox(ol,ob,r), TakeLidOffBox(ol,ob,r), CloseBag(ob,r), OpenBag(ob,r) — the meanings of these can be guessed from their names. We also define some specific sequences of actions; for example, LoadUprightContainer(ox,ob) is the sequence of moving to grasp object ox, which lies outside open container ob; carrying ox so that it is inside ob, and then withdrawing the hand so that it is again outside ob.

Necessary and sufficient conditions can be given for the occurrence of these, and stronger qualitative axioms can be associated with these than with the general actions MoveTo and Travel. For example, we have an axiom that states that, if a collection of objects is in an upright container, and the container is carried upright in such a way that the interior is always large enough to contain the collection, then the objects remain inside the container.[5] The formal statement of this is given in section 5.1 below.

---

[4] In a more complete theory, the argument should specify a trajectory through configuration space but the theories we are developing are too crude to take any advantage of that.

[5] One can construct cases where this does not hold, but those are exceptional.

4.3 Dynamic containers and cavities

In a container made of flexible material, cavities can split and merge; they can open up to the outside world or close themselves off from the outside world.[6]

To characterize cavities dynamically, we use *histories*; that is, functions from time to regions (Hayes, 1979). The place occupied by an object, or by a set of objects, over time is one kind of history. We say that a history C is a *dynamic cavity* of history H from time Ta to time Tb if it satisfies these two conditions:

- At all times Tm strictly between Ta and Tb, C is spatially a cavity in closed container H.
- C is *weakly continuous*. That is, for any time Tm there exists an interval (Tc,Td) and a region R such that throughout (Tc,Td) R is part of C. Intuit*ively, a cavity i*s weakly continuous if a small marble that can foresee how C will evolve and can move arbitrarily quickly can succeed in staying inside C.

We distinguish three categories of dynamic cavities (figure 5):

- C is a *no-exit cavity* of H if there is no way to escape from C except through H.
- C is a *no-entrance cavity* of H if there is no way to get into C except through H.
- C is a *persistent cavity* of H if it is both a no-exit and a no-entrance cavity.

5. The knowledge base

The full details of the knowledge base, as far as it has been developed, are in the supplementary paper, "Radically Incomplete Reasoning about Containers: A First-Order Theory" which can be found at http://www.cs.nyu.edu/faculty/davise/containers/containerAxioms.pdf
Space does not permit the inclusion of a full account here. We will here just summarize the characteristics of the knowledge base, and illustrate with some examples.

The theory is by no means complete or definitive. It is essentially a sketch, to illustrate the kinds of issues that arise in formulating this kind of reasoning and some techniques that might be useful in automating it.

The knowledge base is written in a sorted (typed) first-order logic with equality. We use symbols in lower case typewriter font for variables, such as u,v; symbols in typewriter font, starting with an upper case character, for constants, function, and predicates symbols; such as Lt or Union; and symbols in italics for sorts, such as *Time* or *Region*.

The sorting system is simple.

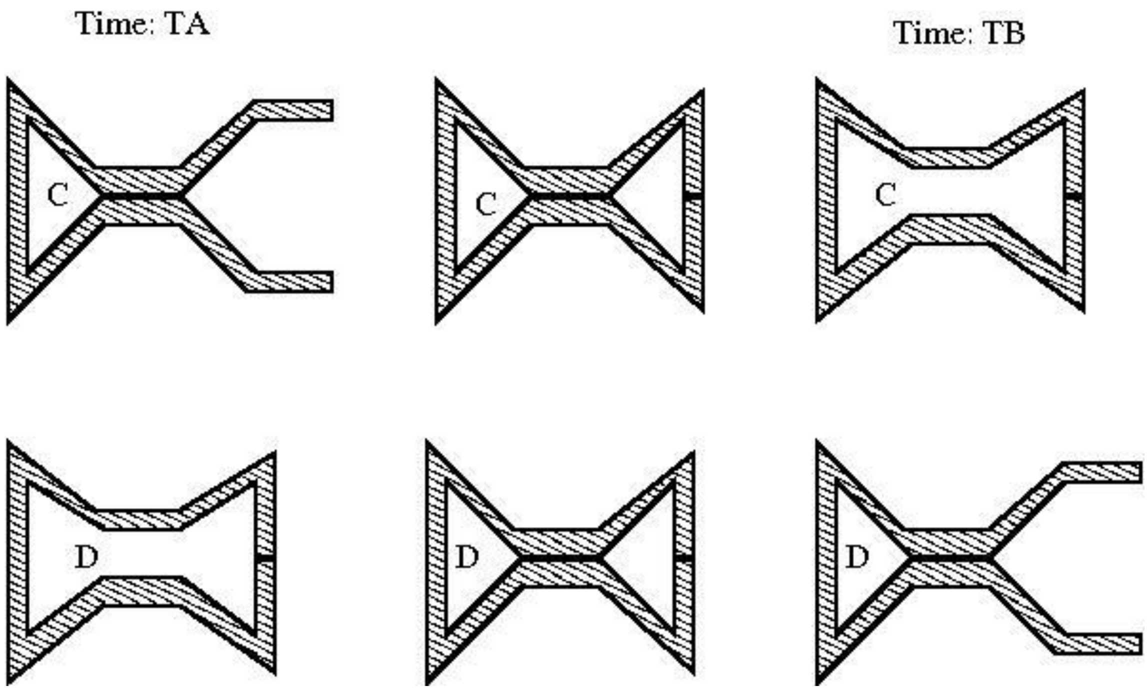
- A sort is equivalent to a monadic predicate.
- The space of entities is partitioned into 6 disjoint sets: *Time, Region, History, Object, ObjectSet* and *Event*.
- Every non-logical symbol (constant, function, predicate) has a unique sortal signature. We do not use overloading or polymorphism.

Therefore any theory in the sorted logic is easily translated into an equivalent theory in an unsorted logic, using a monadic predicate for each sort.

In addition to the general theory, the knowledge base also includes some specifications of specific scenarios and examples of inferences that the theory supports. These are discussed in section 5.2.

---

[6] The classic discussion of cavities and in particular the individuation of cavities is (Casati & Varzi, 1994)



C is a no-exit cavity from TA to TB.
D is a no-entrance cavity from TA to TB.

Figure 5: Dynamic Cavities

5.1 Some examples of axioms

We show a few characteristic axioms to illustrate the flavor of the theory and the knowledge base.

$\forall_{t:Time;\, o:Object}$ FeasiblePlace(o,Place(t,o)).
Every object always occupies a feasible place.

$\forall_{p,q:Object;\, t:Time}$ $p \neq q \Rightarrow$ DR(Place(t,p), Place(t,q)).
Any two objects are spatially disjoint.

$\forall_{t1,t2:Time;\, hc,hb,hs:History}$
NoExitCavity(t1,t2,hc,hb) $\wedge$ Continuous(t1,t2,hs) $\wedge$
P(Slice(t1,hs),Slice(t2,hc)) $\wedge$ ~P(Slice(t2,hs),Slice(t2,hc)) $\Rightarrow$
$\exists_{tm:Time}$ Lt(t1,tm) $\wedge$ Lt(tm,t2) $\wedge$ O(Slice(tm,hs),Slice(tm,hc)).

Let hb be the history of a container (box or bottle or bag); let hs be the history of some stuff; and let hc be a no-exit cavity of hb. If hs is inside hc at time t1and is not inside hc at time t2, then it must overlap with hb at some time tm in between t1 and t2.

$\forall_{t1,t2:Time;\ o:Object}$
$Lt(t1,t2) \wedge Place(t1,o) \neq Place(t2,o) \Rightarrow$
$[\ o = Agent \vee$
$[\ \exists_{tc,td,oc,rx}\ TimeIntervalOverlap(tc,td,t1,t2) \wedge GoesWith(tc,o,oc) \wedge$
$[\ Occurs(tc,td,MoveTo(oc,rx)) \vee Falling(tc,td,oc)]\ ]\ ].$

This is the frame axiom for change of position, discussed in section 4.2. If object o changes its position between times t1 and t2 then either o is the agent, or o moves along with some object oc which is either directly manipulated or falls during some subinterval of [t1,t2].

$\forall_{t1,t2:Time;\ o:Object;\ s:ObjectSet;\ r:Region}$
$Occurs(t1,t2,CarryUpright(o,s,r)) \Leftrightarrow$
$Occurs(t1,t2,MoveTo(o,r)) \wedge$
$\exists_{hc}\ s = Occupants(t1,Slice(t1,hc)) \wedge Continuous(t1,t2,hc) \wedge$
$\forall_{tm}\ Leq(t1,tm) \wedge Leq(tm,t2) \Rightarrow$
$UprightContainer(tm,o,Slice(tm,hc)) \wedge Fits(s,Slice(tm,hc)).$

This is the definition of the event of carrying an open container o upright to region r with contents s (a set of objects). This event occurs if object o is moved to r and there is a dynamic upright cavity hc of o that contains all of s at ta and that is always large enough that all of s fits inside.

$\forall_{t1,t2:Time;\ o:Object;\ s:ObjectSet;\ r:Region}$
$Occurs(t1,t2,CarryUpright(o,s,r)) \Rightarrow$
$UContents(t2,o) = UContents(t1,o).$

If you carry a collection of objects s in an upright container o, then the objects remain in the container. (One can construct cases where this does not actually hold, but those are exceptional.) This is a frame axiom for this specific category of action.

5.2 Sample Inferences

We are now in a position to begin to characterize broad classes of radically incomplete reasoning patterns in terms of inference from the axioms.

Example 1: If Ob is a rigid object and it is a closed container containing object Os, then Os remains inside Ob.

Example 2: If at one time, object Ob contains object Os and at a different time it does not, then Ob is not a rigid object. Note that, logically, this is just a contrapositive of example 1; however, since it is not formulated as a prediction problem, this would be difficult for many forms of automated inference other than knowledge-based reasoning.

Example 3: If Ob1 is a rigid object and a closed container containing Ob2 and Ob2 is a closed container (not necessarily rigid) containing Os3, then Os3 will remain inside Ob1.

Example 4: If the situation depicted in figure 3 is modified so that the red region is flush against the walls of the maze, then the ball must reach the red region before it can reach the green region. (Our theory is not powerful enough to support the same inference in the case where there is a small gap, as shown in figure 3.)

Example 5: If object Ox is outside upright container Ob, and the current contents of Ob together with Ox are much smaller than the interior of Ob5, and the agent can reach and move Ox5 and can reach into Ob, then

a. The agent can load Ox into Ob; and
b. If the agent does load Ox into Ob, then Ox will be contained in Ob.

These inferences have been formally verified to varying degrees. Example 1 has been verified in the formal theorem prover SPASS (Weidenbach, et al., 2009). SPASS is not capable of finding a complete proof from beginning to end; however, we divided a hand-constructed proof into four parts, and SPASS was able to find a proof of each part. Thus, we are using SPASS as a proof-verification system rather than as an end-to-end theorem prover. Complete proofs for examples 2-5 have been hand written in a slightly informal natural deduction system. All these proofs, including the input and output from SPASS, can be found on the project web site: http://www.cs.nyu.edu/faculty/davise/containers/

In short, the knowledge base supports a broad range of inferences from highly underspecified scenarios.

6. Related Work

There is much previous AI work on physical reasoning with partial information, especially under the rubric of "qualitative reasoning" (QR) in the narrow sense (Bobrow, 1985). This work has primarily focussed on qualitative differential equations (Kuipers, 1986) or similar formalisms (Forbus, 1985) (de Kleer & Brown, 1985). The current project is broadly speaking in the same spirit; however, it shares very little technical content, because of a number of differences in domain. First, our theory uses a much richer language of qualitative spatial relations than in most of the QR literature. Systems considered in the QR literature tend to involve either no geometry (e.g. electronic circuits (de Kleer and Brown, 1985); highly restricted geometry (e.g. the geometry of linkages (Kim, 1992)); or fully specified geometries (e.g. (Faltings, 1987)). Second, the problems considered in the QR literature involved primarily the internal evolution of physical systems; exogenous action was a secondary consideration. In our domain, almost all change is initiated by the action of an agent. Finally, the QR literature is primarily, though not exclusively, focused on prediction; our theory is designed with the intent of supporting inference in many different directions.

More directly relevant to our project is the substantial literature on qualitative spatial reasoning, initiated by the papers of Randell, Cui, and Cohn (1992) and of Egenhofer and Franzosa (1991) and extensively developed since (Cohn & Renz, 2008). In particular, we use the RCC-8 language of topological relations between regions as the basis of our spatial representation; the concept of a closed container can be defined in that language. However, the theory of open containers and of open upright containers, and the theory of temporally evolving containers are largely new here.

In previous work (Davis, 2008) (Davis, 2011) we developed representation languages and systems of rules to characterize reasoning about loading solid objects into boxes and carrying objects in boxes, and pouring liquids between open containers. The theory discussed in this paper differs in two major respects. First, the earlier work developed moderately detailed dynamic theories of rigid solid object and of liquids. In this paper, the dynamic theories apply across a much wider range of materials, and therefore are much less detailed. Second, the previous work made arbitrary use of set theory, geometry, and real analysis in constructing the proofs; that is, considerations of both effective implementation and cognitive plausibility were entirely ignored in favor of representational and inferential adequacy. The current project aims at a theory that is both effectively implementable and cognitively realistic, sacrificing expressive and inferential power where necessary.

A study by Smith, Dechter, Tenenbaum, and Vul (2013) studies the way in which experimental participants reason about a ball bouncing among obstacles. They demonstrate that, though in many circumstances, subjects' responses are consistent with a theory that they are simulating the motion of the ball, under some circumstances where qualitative reasoning easily supplies the answer, they can answer much more quickly than the simulation theory would predict. For example, when presented with the situation shown in figure 3 and asked, "Which region does the ball reach first: the red region or the green region?" they can quickly answer "the red region". As it happens, all of the instances of qualitative reasoning they discuss can be viewed as some form of reasoning about containment. Our knowledge base supports this particular inference (assuming that the ball is itself the agent or is being moved by the agent.)

7. Conclusions and future work

Human commonsense physical reasoning is strikingly flexible in its ability to deal with radically incomplete problem specifications and incomplete theories of the physics of the situation at hand. We have argued that an appropriate method for achieving this flexibility in an automated system would be to use a knowledge-based system incorporating rules spanning a wide range of specificity. As an initial step, we have formulated some of the axioms that would be useful in reasoning about manipulating containers, and we have shown that these axioms suffice to justify some simple commonsense inferences.

In future work on this project, we plan to expand the knowledge base to cover many more forms of qualitative reasoning about containers, and to expand the collection of commonsense inferences under consideration. We will also attempt to implement the knowledge base within an automated reasoning system that can carry out the inferences from problem specification to conclusion. As discussed, the SPASS inference engine can only deal with these when they are broken into bite-sized pieces. However, we expect that, if the inference engine can be supplemented with appropriate search heuristics, it may be possible to find the proofs from beginning to end.

References

Battaglia, P., Hamrick, J., & Tenenbaum, J. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences,* to appear.

Bobrow, D. (1985). *Qualitative Reasoning about Physical Systems.* Cambridge, Mass.: MIT Press.

Casati, R., & Varzi, A. (1994). *Holes and Other Superficialities.* Cambridge, Mass:: MIT Press.

Cohn, A., & Renz, J. (2008). Qualitative Spatial Representation and Reasoning. In F. van Harmelen, V. Lifschitz, & B. Porter, *Handbook of Knowledge Representation* (pp. 551-596). Amsterdam: Elsevier.

Davis, E. (1998). The naive physics perplex. *AI Magazine, 19*(4), 51-79.

Davis, E. (2008). Pouring liquids: A study in commonsense physical reasoning. *Artificial Intelligence, 172*, 1540-1578.

Davis, E. (2011). How does a box work? *Artificial Intelligence, 175*(1), 299-345. doi:http://dx.doi.org/10.1016/j.artint.2010.04.006

Davis, E., & Marcus, G. (2013). The Scope and Limits of Simulation in Cognition and Automated Reasoning. *Under submission*.

de Kleer, J., & Brown, J. (1985). A qualitative physics based on confluences. In D. Bobrow, *Qualitative Reasoning about Physical Systems* (pp. 7-84). Cambridge: MIT Press. doi:http://dx.doi.org/10.1016/0004-3702(84)90037-7

Egenhofer, M., & Franzosa, R. (1991). Point-set topological spatial relations. *Intl. Journal of Geographical Information Systems, 5*(2), 161-174.

Faltings, B. (1987). Qualitative Kinematics in Mechanisms. *IJCAI*, (pp. 436-442).

Forbus, K. (1985). Qualitative process theory. In D. Bobrow, *Qualitative Reasoning about Physical Systems* (pp. 85-168). Cambridge, Mass.: MIT Press.

Hayes, P. (1977). In defense of logic. *IJCAI*, (pp. 559-565).

Hayes, P. (1979). The naive physics manifesto. In D. Michie, *Expert Systems in the Microelectronic Age.* Edinburgh: Edinburgh University Press.

Hayes, P. (1985). Ontology for Liquids. In J. Hobbs, & R. Moore, *Formal Theories of the Commonsense World.* Ablex.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Science, 8*(6), 280-285. doi:http://dx.doi.org/10.1016/j.tics.2004.04.001

Herdagdelen, A., & Baroni, M. (2012). Bootstrapping a game with a purpose for commonsense collection. *ACM Transactions on Intelligent Systems and Technology, 3*(4), 59:1-24.

Hespos, S., & Baillargeon, R. (2001). Reasoning about containment events in very young infants. *Cognition, 78*, 207-245.

Hespos, S., & Spelke, E. (2007). Precursors to spatial language: The case of containment. In M. Aurnague, M. Hickman, & L. Vieu, *The Categorization of Spatial Entities in Language and Cognition* (pp. 233-245). Amsterdam: Benjamins Publishers.

Inhelder, B., & Piaget, J. (1958). *The Growth of Logical Thinking from Childhood to Adolescence.* Basic Books.

Kim, H. (1992). Qualitative Kinematics of Linkages. In B. Faltings, & P. Struss, *Recent Advances in Qualitative Physics.* Cambridge, Mass.: MIT Press.

Kuipers, B. (1986). Qualitative simulation. *Artificial Intelligence, 29*(3), 289-338.

Newell, A. (1981). The Knowledge Level. *AI Magazine, 2*(2), 1-20.

Randell, D., Cui, Z., & Cohn, A. (1992). A spatial logic based on regions and connection. *Proc. 3rd Int. Conf. on Knowledge Representation and Reasoning* (pp. 165-176). Morgan Kaufmann.

Reece, J., Urry, L., Cain, M., Wasserman, S., Minorsky, P., & Jackson, R. (2011). *Campbell Biololgy.* (9th, Ed.) Boston: Benjamin Cummings.

Reiter, R. (1995). *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems.* Cambridge, Mass.: MIT Press.

Sandewall, E. (1995). *Features and Fluents: The Representation of Knowledge about Dynamical Systems.* Oxford: Oxford University Press.

Schubert, L. (1990). Monotonic Solution of the Frame Problem in Situation Calculus. In H. Kyburg, R. Loui, & G. Carlson, *Knowledge Representation and Defeasible Reasoning* (pp. 23-67). KLuwer.

Siegal, M. (1991). *Knowing children: Experiments in conversation and cognition.* Mahwah, N.J.: Erlbaum.

Smith, K., Dechter, E., Tenenbaum, J., & Vul, E. (2013). Physical predictions over time. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society.* From http://www.academia.edu/3689251/Physical_predictions_over_time

Teglas, E., Girotto, V., Gonzalez, M., & Bonatti, L. (2007). Intuitions of probability shape expectations about the future at 12 months and beyond. *Proc. Nat. Academy of Sciences, 104*(48), 19156-19159.

Teglas, E., Vul, E. G., Gonzalez, M., Tenenbaum, J., & Bonatti, L. (2011). Pure Reasoning in 12-Month-Old Infants as Probabilistic Inference. *Science, 332*, 1054-1059.

Weidenbach, C., Dimova, D., Fietzke, A., Kumar, R., Suda, M., & Wischnewski, C. (2009). SPASS Version 3.5. *22 Intl. Conf. on Automated Deduction (CADE), LNCS 5563*, pp. 140-145.