Automating Commonsense Reasoning for Elementary Physical Science

Proposal Submitted to the NSF, “Robust Intelligence.”

Ernest Davis, Courant Institute, New York University, New York, New York

Project Description

1. Automated Commonsense Physical Reasoning: Significance and Applications

A fundamental part of human experience is the interaction with physical materials of many different kinds: rigid solids, non-rigid solids of many different kinds (e.g. cloth, string, rubber bands, springs, and so on), liquids and gasses. The rich pre-scientific understanding of these materials acquired by every child at an early age is critically important, both because it enables a person to deal effectively, in daily life and in more specialized activities, with a world full of these materials, and because this fundamental physical understanding serves as a grounding for more sophisticated knowledge of many kinds. The implementation of this body of knowledge in an intelligent knowledge base and the analysis of the knowledge that would be entailed in this implementation would have immense value for many different purposes:

Robotics: An autonomous robot that needs to deal flexibly and sensibly with an uncontrolled environment needs to understand how the environment behaves and how it reacts to his actions. This includes robots that work in a household; in complex and unpredictable industrial setting; in unusual and distant environments (undersea, Mars, etc.); and in hospitals.

Automated instruction: Interactive instructional computer programs could be made much more powerful and effective if they could draw on this body of commonsense physical knowledge. These could potentially be instructional programs for consumers (e.g. for using a sewing machine or for applying first aid); for students, especially in physical science courses; or for workers.

Product and process design: The knowledge base could be used to design or validate tools and techniques for using these tools.

Science knowledge base: An ambitious long-term project, called Project Halo, is underway to encode scientific knowledge in a knowledge base, the Digital Aristotle [40, 41] In the first stage of this project, knowledge-based systems were developed that achieved about the mean human score on questions about balancing chemical equations and pH levels from the high-school AP chemistry test. The second stage, currently nearing completion, has focused on developing tools to aid domain experts to develop knowledge bases for answering questions from the AP physics, chemistry, and biology tests [4]. However, the subject matter in both stages was carefully chosen to avoid issues of spatial reasoning and of commonsense reasoning ([40] and personal communication). But these issues obviously cannot be evaded forever if the project is to attain any reasonable coverage.

Text/diagram understanding: Descriptions of physical processes and mechanisms written for human readers use a sophisticated combination of natural language text and pictorial diagrams. Automated systems that could interpret these at a deep level would be of great value, both for applications which process these texts for the benefit of a human reader, such as document retrieval or machine translation, and for those that use the texts as a knowledge source for a computer-centered activity, such as robotics or CAM. But achieving deep understanding of these documents requires a correspondingly deep understanding of the domain.

Grounding for science and mathematics: Experiences of simple physical interactions are a major epistemological grounding for the understanding of basic physical science and mathematics, directly (concepts of physical science correspond to elements of experience); as explicanda (scientific theories explain commonsense experience); as metaphors (e.g. imagining atoms as hard balls); and as contrasts (e.g. understanding the differences between atoms and hard balls).[26]

Testbed for knowledge representation research: Elementary physical reasoning is a testbed for research in knowledge representation and automated reasoning that is appealing, easily understood, and remarkably rich. It involves many of the central issues of knowledge representation, including taking knowledge that is "intuitive" and hard to verbalize, and representing it explicitly; integrating commonsense and expert knowledge, spatial and symbolic knowledge, deterministic and probabilistic knowledge; and reasoning at multiple levels of abstraction.

Thus, a fully-developed theory of commonsense physical reasoning and its relation to spatial reasoning would be a central conceptual infrastructure in the achievement of robust intelligence in programs for many different kinds of applications and in the development of a cognitive theory of the understanding of science and mathematics; and a important source of insight and experience that can be applied to the development of automated reasoners for other domains.

2. The SOPHY project

In 2005 we began work on a long-term research project, called SOPHY, to develop a knowledge-based system for automated reasoning about simple high-school level chemistry experiments. We have thus far developed a representation and a theory that can support qualitative reasoning about a number of basic physical substances and processes: loading solid objects, or pouring liquids into an open container and carrying them in the container [22, 21]; burning fuel in a closed container; and passivizing a metal by exposure to oxygen [29]. (In passivization, a thin layer of oxide is formed on the surface of a metal; since the oxide is chemically inert, the rest of the metal remains unchanged.)

The next stage of the project will focus on extending these theories to further kinds of physical materials and processes and developing methods for plan verification and plan expansion (section 3). The direction of theoretical analysis will be informed by and grounded in the domain-knowledge needs of the task of of interpreting natural language and diagrammatic descriptions of lab experiments in course assignments. (section 4).

3. Integrating Commonsense and Scientific Reasoning

Consider the experiment shown in figure 1. Potassium chlorate ($KClO_3$) is heated in a test tube, and decomposes into potassium chloride (KCl) and oxygen. The gaseous oxygen expands out of the test tube, goes through the tubing, bubbles up through the water in the beaker, and collects in the inverted bottle over the the water. Once the bubbling has stopped, the experimenter raises or lowers the bottle until the level of the top of water inside and outside the bottle are equal. At

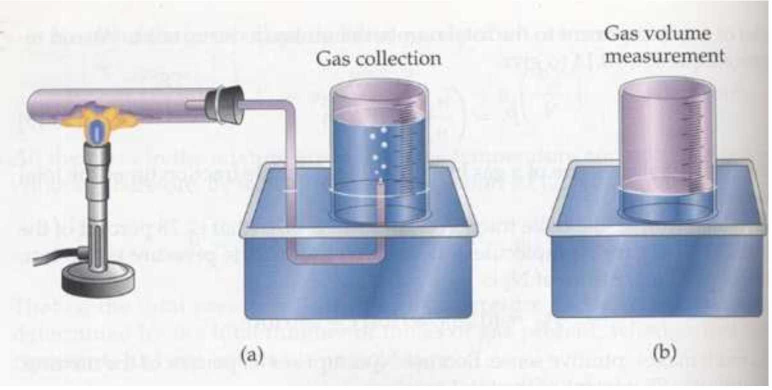

Figure 1: Collection of Gas over Water. From [3], fig. 10.15, p. 372.

this point, the pressure in the bottle is equal to atmospheric pressure. Measuring the volume of the gas collected over the water, and correcting for the water vapor which is mixed in with the oxygen, the experimenter can thus measure the amount of oxygen released in the decomposition.

This experiment is summarized in the formula $2KClO_3 \rightarrow 2KCl+3O_2$, but that, clearly, is only the tip of the iceberg. A real understanding of the experiment involves understanding how the success of the experiment corresponds to the structure of the set-up. Achieving this understanding requires having some basic knowledge about solids, liquids, and gasses, and their interactions; how these interactions are affected by the shapes and spatial relations of the objects involved; and how the physical set-up reflects the abilities of the human experimenter to manipulate the objects involved and to perceive the progress of the experiment.

Our general objective is to develop a knowledge base encoding the basic knowledge of the dynamics of solids, liquids, and gasses needed to understand simple physics and chemistry experiments such as this one. This includes the experimental set-up shown in figure 1 and various devices for controlling the movement of solids, liquids, and gasses, ranging from simple devices such as open and closed boxes, open and closed fluid containers, and pipes, to somewhat more sophisticated devices such as doors, valves, stopcocks, and simple pumps, manometers, and barometers.

3.1. Variants and directions of inference

The adequacy of our representation and our inference techniques will be demonstrated by its ability to understand *variants* of the system and to make inferences in different *directions*.

Variants are greater or lesser modifications of the system which may make no change in its behavior, a small change, or completely alter its behavior [16]. The ease with which a theory can be extended to deal with simple variants is its *elaboration tolerance* [55]. For instance, an elaboration tolerant theory of the decomposition experiment might be expected to be able to answer such questions as:

> What would happen: If the bottle had a hole in its bottom? If it had a hole in the side, below the level of the water in the pan? If it were right-side up and the tubing entered the bottle through the open top? If it were cubical rather than cylindrical? If

> it were not graduated? If it were painted black? If it were the size of a thimble? If the end of the tube were below the opening of the bottle rather than inside the bottle? If the end of the tube were in the pan but outside the bottle? If the tube were blocked? If it were miles long? If the pan were empty? If the water in the pan and the bottle were replaced by milk, honey, loose gravel, or solid granite?

A knowledge-based system should also be able to employ the same knowledge to carry out different kinds of inferences, where different kinds of information are given and demanded. This is indeed one of the major objectives in using a knowledge-based system rather than an task-specific algorithm. In this domain, directions of inference include:

- **Plan projection.** Given the starting state of a physical system and containing an agent (or agents) and a plan of action for that agent, determine whether the plan is executable and what its effect will be. An important special case is **autonomous projection**, in which an inanimate physical system evolves on its own, with no agent. A variant is **comparative projection**: how a change in the conditions of a problem affects the outcome. For example: the denser the liquid in a barometer, the lower the column of liquid will be. The smaller the hole in a container, the more slowly liquid will leak through.
- **Planning, plan completion, plan modification**. Given the starting state of a physical system containing an agent, and a task to be carried out, find a plan of action. Given a partial plan specification or a buggy plan, modify it to be a correct, complete plan.
- **Explanation/diagnosis.** Given the behavior of a system, determine its physical structure. For instance, if liquid is observed to be leaking slowly from the bottom of a container, infer that there is a small hole. If, in the experiment, the substance has melted but no gas is bubbling through the water, infer that there is either a leak or a blockage.
- **Safety (unattainability)**. The purpose of many physical system is to ensure that specific states cannot be attained. An object in a closed box cannot come into contact with objects outside the box. An agent outside a locked door cannot go through the door.

This project will focus primarily on plan projection and plan completion; however, the knowledge base will be designed so as to support other directions of inference.

3.2. Physical theory

We will use a simplified physical theory — it will be complex and challenging enough! Our theory models solids, liquids, and gasses. Solid objects are modelled as strictly rigid in shape. Liquids are taken to be incompressible. We ignore such issues as viscosity, surface tension, adhesion, and cohesion. We assume that a closed system of gasses in a container rapidly attains a uniform equilibrium which satisfies ideal gas law and the law of partial pressures. Our focus, at least in the short term, is on dynamics; other aspects of the physics involved will be dealt with in an *ad hoc* way, sufficient to deal with the particular example.

Our model of action is that one or more solid objects are directly controlled by the agent, and that the agent can either move them along a desired trajectory, or use them to exert a force against an object that is in contact. The objects being controlled may either be the robot's own manipulators or, more abstractly, the tools he is working with; for example, a robot can either think of moving its hand so as to pick up a beaker, or, more abstractly, it can think simply of moving the beaker, without envisioning its hand.

3.3. Partial Geometric Specification

The most important difference between the SOPHY project and all but a handful of the many existing programs that do physical projection is that we focus on cases where the problem characteristics, especially the geometry, are specified only partially or abstractly. That is, almost all physical reasoning programs that deal with geometry at all assume that precise geometric and material characteristics are given in the problem specification. Given such precise boundary conditions, the programs can then make precise predictions. In many important applications, the geometric knowledge is not complete; it derives either from inexact and partial perception (e.g. significant parts of the objects involved are occluded); or from natural language text, which is rarely geometrically complete; or from schematic diagrams, which often show topology but are often (deliberately) incorrect otherwise; or from inference (e.g. from previously observed physical behavior). Partial geometric specifications also arise when one is reasoning inference about all objects in a class rather than a single object, or about a system that is in the process of being designed and has not yet been fully specified.

3.4. Abstracting transitory states

The behavior of physical systems often goes through highly complex transition states before setting into a state of equilibrium. For instance, when liquid is poured into a container, its shape and motion while being poured is extremely complicated and unstable with respect to small perturbations, whereas its final state – sitting at rest at the bottom of the container under a horizontal top surface – is extremely simple. There are three standard methods of projection for this problem: first, to ignore these transitional states altogether, and posit atomic transitions from one equilibrium state to the next; second, to view the pouring as a process that continuously reduces the volume in one container and increases it in the other, with no account of how this is mediated; and third, to use partial differential equations or simulation to trace the way in full detail through all the intermediate stages. The first two, though often useful abstractions, are too abstract for many kinds of commonsense reasoning (e.g. predicting what happens if an object is placed into the stream of liquid); the third is unusable in the context of commonsense reasoning, since it requires precise information of geometry, timing, and material characteristics, and it gives output that is too detailed, unreliable, and unstable. What we need is a representation and a mode of reasoning somewhere between the second and the third; that constrain the behavior during transitions, so that one can be sure that nothing too strange is happening, while not requiring that this behavior be worked out in full detail

We have already developed a number of different physical domain theories of this kind, that support making qualitative predictions over extended time to be made based on qualitative spatial and physical specifications: loading objects into a box, and carrying objects in an open box, even if the objects inside the box may shift while this is happening [22]; carrying liquids in an open container and pouring liquids from one container to another [21]; burning a fuel in a closed container, and passivizing a metal by exposure to oxygen [29].

4. Deep domain knowledge for text understanding

Developing an automated system with a deep understanding of texts that describe physical systems and processes will require a correspondingly deep domain theory. We plan to study this issue by carrying out a careful analysis of a small selection of such texts and determining how difficulties in their interpretation could be addressed using domain knowledge. For instance, a preliminary examination of the online experimental description “The Decomposition of Potassium Chlorate” [45] reveals many places where correct interpretation depends on domain knowledge. (Note: the experimental set up here is not the same as in figure 1 above.) The following instances are typical:
- Step 1 of the procedure is “Record the atmospheric pressure from the laboratory barometer.” In fact, this step need not be first; it can be carried out at any stage.
- Step 2 states, “Caution, $KClO_3$ is a very strong oxidizing agent. . . . Do not let this substance contact paper or the rubber stopper in the test tube of the apparatus.” In fact, this constraint applies throughout the duration of the experiment, but is primarily a concern during step 6, in which the substance is poured into the test tube, and step 7, in which the test tube is manipulated.
- Step 10 states, “Place the glass tube (connected to the hose) back into the beaker”. In fact this means, “Take the glass tube that is connected at one end to the hose, and put the other end into the beaker, maintaining the connection to the hose at the first end.”

Examples of this kind occur on practically every line of the document [27]. Moreover, the diagram is, of course, a two-dimensional projection of a temporal snapshot of the experiment; reconstructing the actual three-dimensional structure and the relation of the snapshot to the temporally extended execution of the procedure calls on both physical reasoning and the text.

These do not at all reflect actual flaws in the write-up, which, after all, is addressed to human students and not to robots. The point is that making this text usable by an automated system requires a lot of plan elaboration, plan repair, and disambiguation that cannot be met using superficial techniques but inescapably requires deep reasoning.

Our short-term objectives for this part of the project are to analyze and categorize as completely as possible the interpretive cruxes of this kind that arise in a few sample texts and diagrams; to characterize the kinds of domain knowledge and reasoning that would be needed to resolve these cruxes; to show how this domain knowledge, or part of it, can be inferred from the knowledge base; and to analyze the issues involved in accessing this knowledge in the process of text interpretation

I do not want to put forward unrealistic expectations for the relevance of this domain knowledge analysis for a practical interpretation system; the histories of KR and NLP are littered for proposals for knowledge-based text analysis that seemed great on paper but never materialized. Nonetheless, there has recently been significant progress in this direction, drawing on the existence of very large knowledge bases such as CYC, the immense textual corpuses that can be drawn from the Web, and progress in information extraction techniques; see for example [5, 6] on the use of Cyc for question answering and for disambiguation, and [1] for an overview of the use of knowledge-based system in question answering.

5 The state of the SOPHY project

We began the SOPHY project five years ago, building on previous work on the representation of commonsense physical knowledge. We have thus far made substantial progress on formalizing commonsense knowledge about the dynamics of solids, liquids, and gasses including phase change and chemical reactions. The next steps are to extend the scope of the domain theories; to use proof verification software to verify sample inferences; and to begin work on the task of interpreting lab assignments.

5.1. Work completed: Boxes

We developed a logical theory capable of supporting the inference that an agent can move a collection of objects from one location to another by loading them one by one into a box and then carrying the box to the destination [22]. The theory consists mostly of first-order axioms, augmented by two non-monotonic rules that serve to exclude certain scenarios that are highly implausible but not absolutely impossible. Some notable features of this theory are:

- We define a semantics of plans, and prove the validity of the plan,
  ```
  { repeat until (all the cargo objects are in the box)
            load some object into the box;
    carry the box to the destination }
  ```
  The plan semantics required here differs from standard semantics for planning languages, because the space of actions includes all possible motions satisfying some basic smoothness constraints.

- The specifications for the starting state give only qualitative information about the shapes of the box and the cargo objects, and the number of cargo object. The plan specifications give only qualitative information about the motions involved in loading the objects and in carrying the box.

- As described in section 3.4, the theory supports the inference that the objects remain in the box while being loaded and while the box is being carried, without a detailed analysis of their motions or the forces on them. This can be done even if the objects settle into place while being loaded, or rattle around in the box while it is carried.

- The theory covers a number of significant variants, including using a box with a lid, carrying objects in a milk crate (a box with small holes) and putting one box inside another.

5.2. Work completed: Liquids

We developed a logical theory for qualitative reasoning about carrying liquids in closed and open containers and pouring liquids from one container to another [21]. Like the theory of boxes, this supports reasoning from qualitative specifications of shape and motion; it supports reasoning about extended time without detailed specification of behavior over differential time; and it supports a number of different variants including carrying a liquid in a closed container; carrying liquids in an open container; using a container formed by the combination of several solid objects; causing a liquid to overflow by dropping pebbles into a pitcher; and, to some extent, pouring liquid over arbitrarily shaped solid objects.

5.3. Work completed: Ontology of matter

We carried out an extensive comparison of seven ontologies for matter in terms of their suitability for representing eleven different kinds of physical laws and physical behavior [28]. Specifically, the ontologies we considered were the model of atoms and molecules with statistical mechanics; models of spatio-temporal fields, with either points, regions, or histories; models of continuous moving material in terms of chunks of matter, with or without point particles; and a hybrid theory that combines atoms and molecules, chunks of matter, and continuous fields, using each where appropriate. The physical concepts and scenarios were part/whole relations among bodies of matter; additivity of mass; motion of a rigid solid object; continuous motion of fluids; fixed mass proportions and spatial continuity at chemical reactions; conservation of mass at chemical reactions; gasses in a container attaining equilibrium; the ideal gas law and the law of partial pressures; liquid at rest in an open container; carrying liquid in an open container; the constant availability of oxygen for reactions in an atmosphere; and surface passivization of metals. Our conclusion was that overall, the field model with histories and hybrid model are the best suited to these kinds of problems though neither is unproblematic. Of course, the hybrid model must be carefully constructed to make sure that the different viewpoints are mutually consistent.

5.4. Work completed: Chemical experiments

We have developed a logical theory in which two simple chemical experiments can be represented and reasoned about qualitatively: the burning of solid fuel in a closed container, and the passivization of the surface of a metal by exposure to the atmosphere [29]. In this theory, we use the hybrid representation of the ontology of matter described above.

The theory supports inferences such as the following. In the first experiment:

- If there is too little oxygen in the container, then some of the fuel will remain unburned. If there is little fuel and plenty of oxygen, then some of the oxygen will remain unused.
- If the contact between the fuel and the atmosphere is broken at the points where combustion is taking place — for instance the fuel is doused with water — then the combustion will stop.
- If a chunk of fuel remains connected and solid, then its constituent atoms maintain a constant relative position. (Note, by contrast, that if a chunk becomes split, then the parts may move relative to one another.)
- The gaseous chemical products of the reaction spreads rapidly throughout the container. Once combustion stops, they soon become uniformly spread throughout the atmosphere in the container.
- The combustion process cannot generate a new internal cavity inside the fuel. It can expand an internal cavity, either by burning through to the cavity from the outside, or by burning it on the inside; but the latter is possible only if there is oxygen in the internal cavity and the combustion can be ignited.

In the second experiment, suppose that the part of the surface that is exposed to the atmosphere changes over time; for example, the bottom of metal bar is in an oil bath while the top part is exposed to the atmosphere, and the bar rotates during the course of the experiment. Then:

- If a part of the surface is exposed to the atmosphere, then it immediately is passivized and remains passivized.
- Any part of the interior of the metal, and any part of the surface that is never exposed to the atmosphere, is not passivized.
- The passivization of the surface of a metal bar does not measurably change the concentration of oxygen in the container. However, if the container is filled loosely with metal filings, which

passivize, then that may measurably reduce the concentration of oxygen.

5.5. Work in progress: More complex scenarios

The next major step is to extend our analysis to cover a broader range of physical processes. For example, representing the experiment shown in figure 1 requires at least a partial theory of heat; a theory of pressure; and theories of gas movement of various kinds including bubbling, evaporation, and the behavior of an inverted beaker of gas over liquid.

One issue that arises very often in this kind of reasoning is the problem of ignoring negligible quantities. For instance, the evaporation of water into the gas in the beaker is negligible as regards the water level in the basin, though it is significant in computing the vapor pressure inside the beaker. The evaporation of water from the basin into the open air is negligible over the time period of the experiment, but would be important if the apparatus were left for a month. The qualitative calculus of the kind of order-of-magnitude reasoning is well understood (e.g. [61]); but there are few formal attempts to integrate it with a rich physical theory.

5.6. Future work: Verification of reasoning

The object-level proofs of inferences in these domains that we present in [22, 21, 29] are long and hand-constructed, and draw on powerful mathematical theories of geometry and real analysis. Constructing these proofs automatically is far beyond the state of the art. Therefore, as a first step toward general automated reasoning, we will attempt the much easier task of verifying the proofs, using a high-powered proof-verifier such as Isabelle/HOL [60]. The hope here is that a large library of lemmas can be built up that can be substantially reused from one variant problem to another, and from one direction of inference to another. Conversely, this would provide a new and quite different testbed for proof verification which might be of interest to that community. Another potential application of the proof structures developed here could be explanation generation; a natural language explanation could, in part, follow the structure of the formal proof.

5.7. Future Work: Interpretation of Laboratory Assignment

As a specific application of the domain theory, we will address the problem of automating the interpretation of laboratory assignments, elaborating the texts and images in the assignments to a fully fleshed-out form that could be executed by a robot. Over the short term, we certainly do not plan to address either the issues of natural language processing faced in interpreting the text, the issues of vision faced in interpreting the image, or the issues of robotics faced in physically carrying out the experiment. Rather, we will take as input logicized translations of the information in the text and image and deliver as output an abstracted high-level specification of the robot control. At a later stage of research, it may be possible to interface this to actual NLP and vision front ends, and a robotic or animated back end.

Of course, even this truncated task involves many difficult problems:

- **Plan completion and plan repair.** As discussed above, the plans as stated in the assignments are always incomplete, and, taken literally sometimes buggy. In order to be actually executed, they need to be completed and repaired.
- **Diagrammatic interpretation.** The diagrams are both over-specified, in that they show

specific dimensions and spatial relations that need not be met exactly, and schematic, in that they omit actual features that are necessarily present in the real set-up but would be merely distracting if shown in the diagram. In using the diagram to guide the physical set up, it is therefore necessary to incorporate both information from the text and from domain knowledge.
- **Disambiguation.** Ambiguities in the text and in the diagram can be resolved using domain-specific knowledge.

6. Evaluation

Numerical evaluation, based on percentage of success over a preselected corpus of example tasks, is not very meaningful for a project like SOPHY, which is really a first exploration of issues in representation and reasoning over a largely unstudied domain. One could perhaps preselect measures of success associated with the specific task of interpreting lab assignments, but formulating such measures is a substantial task in itself, and seems unproductive, for a number of related reasons. First, the formulator would necessarily be the researcher himself, so it could not be an independent measure. The project does not address a specific end-user application, so these measures do not reflect any immediate real-world payoff (e.g. dollars saved). SOPHY is not in competition with any other project on this task, so there is no issue of comparative degree of success. Finally, at this stage of our understanding, it is impossible foresee what aspects of the task are feasible or most important, so evaluation in terms of a predefined measure could easily either miss the progress that is made, or exaggerate its significance.

Rather, at this stage of research, meaningful evaluation can only be a qualitative and retrospective consideration of the progress made in the conceptual analysis of the domain and the task.

The chief criteria for progress in the representational analysis are:
- **Scope and extensibility.** How broad is the range of physical phenomena, of qualitative information, and of the kinds of inference that the theory will support? Does the theory seem to be easily extensible to broader ranges?
- **Coherence and clarity.** Is the viewpoint of the theory coherent? Is the meaning of the symbols clear? Can other researchers use the representation correctly?
- **Fruitfulness.** Will the conceptual framework be useful in other domains and applications?

The chief criteria for progress in the automated verification project are:
- **Scope:** What kinds of inferences are able to be verified automatically?
- **Ease of use:** How much human labor is involved in verifying an inference?
- **Multiple use:** Can proof structures developed for one inference be used in another, or must each inference be worked out from scratch?
- **Insight:** The domain here is rather different from most of the applications of proof verifiers (mostly mathematics or program verification). Does this work make any useful suggestions as to the design of proof verifiers?

The chief criteria for progress in the interpretation task are:
- **Scope:** What aspects of the interpretation task are in principle characterized by the theory developed, and what lies outside the theory?
- **Implementation:** What parts of the interpretation task are implemented, and how accurate

and efficient is the implementation?

• **Front and back end interface:** How large and how difficult is the gap between the real text and image input and to the real robotic control output and the specifications used and output generated by the theory or implementation?

7. Related Work

The previous research relevant to this proposal falls mostly into two categories, which have not previously interacted much: work on rule-based approaches to commonsense physical reasoning; and work on algorithmic approaches to physical reasoning. For a more extensive review see [20]. There is also some relevant recent work in the ontology of chemistry and the philosophy of chemistry.

7.1. Rule-based physical reasoning

This is a comparatively small body of research. It was initiated by Pat Hayes' "Naive Physics Manifesto" [43], which proposed a large-scale project of encoding naive physics in an axiomatic system, deferring the problems of effective implementation. Hayes provided an extensive instance of his project in the "Ontology for Liquids" [44], which constructed an ontology and an axiomatic system for reasoning about liquids.

Prior to beginning work on SOPHY, I developed a number of theories of this kind: an axiomatic system supporting the inference that a small marble dropped inside a large funnel will come out the bottom [7]; axiomatizations [8, 9] of the qualitative reasoning in ENVISION [33] and QP [38]; a pair of ontologies, with axiomatizations, for reasoning about cutting solid objects [11]; and an ontology for continuous branching time, needed for modelling an agent that controls a manipulator [13]. I also wrote a couple of methodological papers: a discussion of the infinitary problems that arise in axiomatizing physical theories [10] and a review of methodology in this field, with a discussion of the pros and cons of various approaches and an analysis of the major difficulties that are encountered [16].

Sandewall [64] developed a logical description of a microworld of point objects moving along surfaces. The chief focus of this work was integrating non-monotonic logic with temporal logic. Three researchers [54, 57, 65] worked on the problem, posed by me [56], of axiomatizing reasoning about an egg being cracked into a bowl The papers are independent; each presents a separate axiomatization. The Halo project [40] is discussed in the introduction above.

7.2. Solid Object Physics

There are many AI systems that carry out computations of one kind or another about the physical interactions of solid objects.

The use of configuration spaces for kinematic analysis of mechanisms was first put forward by Faltings [36] and Joskowicz [48]. Subsequent work has improved the algorithms and found a wide range of applications [62, 52]. Joskowicz and Sacks [50] determined that a large fraction of the mechanisms enumerated in a standard encyclopedia of mechanisms can be explained purely in kinematic terms.

Simulators for the behavior of solid objects using a full dynamic theory have been developed in the context of computer-aided engineering [69] and AI [42] These carry out a exact simulation of behavior given exact geometric specifications of the objects involved. Sacks and Joskowicz [63] present an algorithm that efficiently carries out dynamic simulation for two-dimensional assemblies using configuration spaces to expedite the problem of collision detection.

The works in solid object physics most relevant to SOPHY are the studies of qualitative reasoning about the behavior of physical systems. The first projects of this kinds were NEWTON [32], which gave a qualitative analysis of the behavior of a point object on a roller-coaster track, and FROB [37], which analyzed the behavior of point objects in among two-dimensional obstacles.

Programs such as [36, 59, 62] took a configuration space that had been computed exactly from exact shape descriptions, divided the configuration space into significant regions, and thus were able to compute qualitative properties of the systems from the relations between these regions. [39] and [67] extend this approach with a qualitative representation of forces and motions, and thus producing a system for qualitative dynamic prediction.

Other directions for the study of qualitative reasoning include studies of kinematic reasoning in cases where the shapes of the objects involved are known only to within a given tolerance [51, 15], and studies of abstraction techniques that can be applied to solid object kinematics [59, 14].

7.3. Reasoning about Fluids

The major study of qualitative reasoning about fluents is by Kim [53]. The program described here does qualitative prediction ("envisionment") for physical systems of liquids and gasses in solid containers; it can handle basic instances of many of the systems described in section 1. The spatial language combines topological predicates with order relations on vertical height. The envisionment algorithm begins by dividing a given situation into separate physically significant regions (a "place vocabulary") and then using a form of qualitative reasoning to carry out a qualitative prediction.

Johnston and Williams [46, 47] have developed a simulation program for the egg-cracking problem discussed above; this requires integrating models of rigid objects, of cracking as a process operating on rigid objects, and of liquids (the internals of the egg).

7.4. Ontology and Philosophy of Chemistry

A major effort at constructing an ontology (in the technical sense) of chemicals and their properties is being undertaken in the ChEBI project [35, 34, 2]. This project and ours encounter some of the same representational problems in characterizing chemicals and their behavior. As more knowledge of chemistry is built into SOPHY, it may be possible to draw on the information in the ChEBI database.

Some work in philosophy of chemistry considers formal analysis of chemistry at the mesoscopic level and its relation to chemistry at the molecular level, in ways that bearing on our conceptual analysis (e.g. [58]).

8. Strengths of the Proposal

I have been working in the areas of commonsense physical reasoning and qualitative spatial reasoning for twenty years, and have published extensively in these areas. I have written a textbook on commonsense reasoning [8] with an extensive discussion of spatial and physical reasoning. I have also supervised one doctoral thesis [49] in the area of physical reasoning.

The project will be much enriched by ties and communication with other ongoing projects at NYU and elsewhere. The large research endeavors at Courant in modelling and simulation and in computational geometry, which involve both the computer science and the mathematics departments, are sources of much inspiration and information as to more mainstream approaches to physical and spatial reasoning. I can also draw on the expertise of the NYU research group in program verification — Clark Barrett, Ben Goldberg, and Patrick Cousot — in dealing with automated logical inference of all kinds.

I have also had very helpful email discussions with Vinay Chaudhri, who is leading the Halo project at SRI. He has generously made available to me the content and the interface for their knowledge base, which should be an extremely valuable resource in the development of SOPHY.

9. Broader Impact

As discussed in the introduction, the chief impact of this work is in the long term. The project is a first step toward a general knowledge base that could be used in autonomous robotics; automated science instruction; tools for planning, design, and verification of laboratory experiments and tools; and a knowledge base of scientific knowledge. It is also an exploration in a new and rich domain of automated proof verification, plan completion, and plan repair techniques.

The short term impact of the project includes:

- The training of doctoral students. Over my career I have supervised nine doctoral students, who are now working at such places as IBM Watson Labs, Microsoft Research, Hebrew University, and ISI. I certainly hope to continue this.
- Dissemination of knowledge beyond the immediate research community. I am working on a book [26] addressed to a semi-popular audience, on the subject of how commonsense physical and spatial knowledge interacts with formal mathematical calculations. This will draw on many of issues that I have studied in my research.

10. Results from Recent NSF-supported Research

Grant: "Automating Commonsense Reasoning for Elementary Physical Science," NSF IIS-0534809, $328,877, 2/06-1/09.

During the past five years of NSF support, my associates and I have carried out in-depth studies in three general areas of commonsense reasoning: physical reasoning; qualitative spatial reasoning; and reasoning about multi-agent planning and communication. Our research group has also carried out research in automated planning in domains with continuous change; and a number of techniques for for improved retrieval of web documents.

Another substantial educational project is that I developed a new course, "Mathematical Techniques for Computer Science Applications", an introductory course in linear algebra, probability, and statistics for computer science masters students, and I am currently writing a textbook [25].

Physical Reasoning

Our studies of physical reasoning have led to the following results, described in section 5:

1. The analysis of commonsense reasoning about loading objects into boxes and carrying objects in boxes.

2. The analysis of commonsense reasoning about carrying liquids in containers and pouring liquids between containers.

3. The analysis of two simple chemical experiments: burning a fuel in a closed container, and the passivization of a metal.

4. Detailed comparison of a number of different ontologies for matter in terms of their suitability for inference about simple physical and chemical processes.

Spatial Reasoning

5. An analysis of the expressivity of the first-order language allowing quantification over regions, and containing the one predicate, "Closer(x,y,z)" (region x is closer to y than to z). We have shown that any relation that is analytical and invariant under orthogonal transformations can be expressed in this language. Roughly speaking, the language is capable of expressing essentially all the concepts in standard mathematical geometry and analysis. [19].

6. An analysis of a number of techniques for reconstructing spatial regions from sample points, and a proof that, under specified conditions, the reconstructed region is "close" to the true region, under a number of different definitions of "closeness." [24]

7. An analysis of the use of transition graphs in reasoning about continuous spatial change. We give general definitions of different categories of transition graph for a partition of a topological space. We prove that the class of paths through the graphs is elementary equivalent to the class of continuous paths through the space, relative to a specified first-order language. We show how this theory can be applied in real-world domains such as rigid objects, strings, and liquids [23].

Multi-agent plans and communication

8. A new, highly expressive, language of informative communications, that allows an agent to communicate almost any property of the current situation, defined in terms of logical combinations of conditions that hold currently, held in the past, or will hold in the future, including facts about agents' knowledge and facts about other communicative acts [17, 18]. We have proved that this theory is consistent with a large class of physical theories, and that it avoids both the liar paradox and the "unexpected hanging" problem.

9. Leora Morgenstern and I have developed a new, highly expressive, theory of multi-agent planning. This extends the very general notion of a plan and of plan correctness developed in [12] and the theory of informative communications described in (5) above to include a similarly expressive theory of requests. We have shown that this theory is supports validation of a simple sample multi-agent plan, and that it is consistent [31].

Planning with Continuous Time

10. Ji-Ae Shin [66] implemented a planner that solves problems in domains that involve continuously changing numerical fluents. Her technique is to compile these problems into Boolean

combinations of linear constraints and propositional atoms; to use LPSAT [70] an existing constraint engine, to find a solution to the constraints; and then to interpret the solution as a plan. We have proved that the planner is sound and complete.

Web Search Engines

11. Ziyang Wang developed and tested a system that monitors a local web site for new information and presents it to the user [68].

Development of human resources

One student has completed a doctorate under my advisement within the last five years: Ziyang Wang, "Incremental Web Search: Tracking Changes in the Web." May 2006.

I am currently advising another student, Paul Bethe, who is writing a master's thesis on aspects of computer bridge.

Publications:

E. Davis. "Knowledge and Communication: A First-Order Theory." *Artificial Intelligence,* vol. 166 nos. 1-2, 2005, pp. 81-140.

E. Davis. "Mathematics as Metaphor: Review of *Where Mathematics Comes From,* by George Lakoff and Raphael Nuñez." *Journal of Experimental and Theoretical Artificial Intelligence,* vol. 17, no. 3, 2005, pp. 305-315.

E. Davis. "The Expressivity of Quantifying over Regions." *Journal of Logic and Computation,* vol. 16, 2006, pp. 891-916.

E. Davis. "Physical Reasoning." In *The Handbook of Knowledge Representation,* F. van Harmelen, V. Lifschitz, and B. Porter (eds.), Elsevier, Oxford, 2008, to appear.

E. Davis. "Pouring Liquids: A Study in Commonsense Physical Reasoning." *Artificial Intelligence,* vol. 172, 2008, pp. 1540-1578.

E. Davis. "How Does a Box Work? A Study in the Qualitative Dynamics of Solid Objects." *Artificial Intelligence.* To appear.

E. Davis. "Preserving Geometric Properties in Reconstructing Regions from Internal and Nearby Points." Submitted to *Discrete Computational Geometry.*

E. Davis. "Ontologies and Representations of Matter." In preparation

E. Davis "The Logic of Fire: Representing the Kinematics of Chemical Reactions." In preparation.

E. Davis and L. Morgenstern. "A First-Order Theory of Communication and Multi-Agent Plans." Vol. 15, No. 5, 2005, pp. 701-749.

J. Shin and E. Davis. "Processes and Continuous Change in a SAT-Based Planner." *Artificial Intelligence,* vol. 166 nos. 1-2, 2005, pp. 194-253.

Z. Wang, "Incremental Web Search: Tracking Changes in the Web." NYU Ph.D. thesis, May 2006.

References

[1] M. Balduccini, C. Baral, and Y. Lierler, 2008. Knowledge representation and question answering. In *The Handbook of Knowledge Representation,* F. van Harmelen, V. Lifschitz, and B. Porter (eds.), Elsevier, Oxford, pp. 779-819.

[2] C. Batchelor. 2008. An Upper-Level Ontology for Chemistry. *FOIS-2008.*

[3] T.L. Brown, H.E. LeMay, and B. Bursten. 2003. *Chemistry: The Central Science,* Prentice Hall.

[4] V. Chaudhri et al. 2009. "AURA: Capturing Knowledge and Answering Questions on Science Textbooks". Tech. Report, SRI International.

[5] J. Curtis, G. Matthes, and D. Baxter. 2005. On the effective use of Cyc in a question answering system. *Proceedings of the Knowledge and reasoning for Answering Questions Workshop, IJCAI-05*, pp. 61-70.

[6] J. Curtis, J. Cabral, and D. Baxter. 2006. On the application of the Cyc ontology to word sense disambiguation, *Proc. 19th Intl. Florida Artificial Intelligence Research Society Conference*, pp. 652-657.

[7] E. Davis. 1988. A Logical Framework for Commonsense Predictions of Solid Object Behavior. *AI in Engineering,* vol. 3 no. 3, pp. 125-140.

[8] E. Davis. 1990. *Representations of Commonsense Knowledge.* Menlo Park, Calif.: Morgan Kaufmann.

[9] E. Davis. 1992. Axiomatizing Qualitative Process Theory. *Third International Conference on Knowledge Representation and Reasoning.*

[10] E. Davis. 1992. Infinite Loops in Finite Time: Some Observations. *Third International Conference on Knowledge Representation and Reasoning.*

[11] E. Davis. 1993. The Kinematics of Cutting Solid Objects. *Annals of Mathematics and Artificial Intelligence* 9(3,4): 253-305.

[12] E. Davis. 1994. Knowledge Preconditions for Plans. *Journal of Logic and Computation,* vol. 4, no. 5, Oct. 1994, pp. 721-766.

[13] E. Davis. 1994. Branching Continuous Time and the Semantics of Continuous Action. *Second Intl. Conference on AI Planning Systems.*

[14] E. Davis. 1995. Approximation and Abstraction in Solid Object Kinematics. NYU Computer Science Tech. Report #706, September 1995

[15] E. Davis. 1995b. Approximations of Shape and Configuration Space. NYU Computer Science Tech. Report #703, September 1995.

[16] E. Davis. 1998. The Naive Physics Perplex. *AI Magazine*, Winter 1998, Vol. 19. No. 4. pp. 51-79.

[17] E. Davis. 2004. A First-Order Theory of Communicating First-Order Formulas. *KR-2004.*

[18] E. Davis. 2005. Knowledge and Communication: A First-Order Theory. *Artificial Intelligence,* vol. 166 nos. 1-2, 2005, pp. 81-140.

[19] E. Davis. 2006. The Expressivity of Quantifying over Regions. *Journal of Logic and Computation,* vol. 16, pp. 891-916.

[20] E. Davis. 2008. Physical Reasoning. In *The Handbook of Knowledge Representation,* F. van Harmelen, V. Lifschitz, and B. Porter (eds.), Elsevier, Oxford, pp. 597-620.

[21] E. Davis. 2008. Pouring Liquids: A Study in Commonsense Physical Reasoning. *Artificial Intelligence,* Vol. 142, pp. 1540-1578.

[22] E. Davis. 2010. How Does a Box Work? A Study in the Qualitative Dynamics of Solid Objects. *Artificial Intelligence.* To appear.

[23] E. Davis. 2010. Qualitative Reasoning and Spatio-Temporal Continuity. In S. Hazarika (ed.) *Qualitative Spatio-Temporal Representation and Reasoning: Trends and Future Directions,* IGI Global Press, to appear.

[24] E. Davis. Preserving Geometric Properties in Reconstructing Regions from Internal and Nearby Points. Submitted to *Discrete Computational Geometry.*

[25] E. Davis. *Linear Algebra, Probability, and Statistics for Computer Science Applications.* In preparation.

[26] E. Davis. *Commonsense Knowledge and Elementary Mathematics.* In preparation.

[27] E. Davis, Domain knowledge in interpreting a chemistry experiment assignment, In preparation.

[28] E. Davis, Ontologies and representations of matter. In preparation.

[29] E. Davis, The logic of fire: Representing the kinematics of chemical reactions. In preparation.

[30] E. Davis, N.M. Gotts, and A.G. Cohn. 1999. Constraint Networks of Topological Relations and Convexity. *CONSTRAINTS,* Vol. 4 No. 3, pp. 241-280.

[31] E. Davis and L. Morgenstern (2005). A First-Order Theory of Communication and Multi-Agent Plans. *Journal of Logic and Computation,* Vol. 15, No. 5, 2005, pp. 701-749.

[32] J. de Kleer. 1977. Multiple Representations of Knowledge in a Mechanics Problem Solver. *Proc. IJCAI-77* pp. 299-304.

[33] J. de Kleer and J.S. Brown. 1985. A Qualitative Physics Based on Confluences. In D. Bobrow (ed.) *Qualitative Reasoning about Physical Systems.* Cambridge, Mass.: MIT Press.

[34] P. de Matos et al. 2009. Chemical entities of biological interest: an update. *Nucleic Acids Research.*

[35] K. Degyarenko et al. 2008. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research* **36**, D344UD350.

[36] B. Faltings. 1987. Qualitative Kinematics in Mechanisms. *IJCAI-87,* pp. 436-442.

[37] K. Forbus. 1980. Spatial and Qualitative Aspects of Reasoning about Motion. *AAAI-80.*

[38] K. Forbus. 1985. Qualitative Process Theory. In Bobrow, D. (ed.) *Qualitative Reasoning about Physical Systems,* Cambridge, Mass.: MIT Press.

[39] K. Forbus, P. Nielsen, and B. Faltings. 1991. Qualitative Spatial Reasoning: The CLOCK project. Technical Report 9, Northwestern University, Institute for the Learning Sciences.

[40] N. Friedland et al. 2004a. Toward a quantitative platform-independent quantitative analysis of knowledge systems, *KR-2004.*

[41] N. Friedland et al. 2004b. Project Halo: Toward a Digital Aristotle, *AI Magazine,* Winter 2004, vol. 25 no. 4, pp. 29-47.

[42] A. Gelsey. 1995. Automated Reasoning about Machines. *Artificial Intelligence,* vol 74, pp. 1-53.

[43] P. Hayes. 1979. The Naive Physics Manifesto. In Michie, D. (ed.) *Expert Systems in the Microelectronic Age.* Edinburgh: Edinburgh University Press.

[44] P. Hayes. 1985. Naive Physics 1: Ontology for Liquids. In Hobbs, J. and Moore, R. (eds.) *Formal Theories of the Commonsense World.* Norwood, N.J.: Ablex Pubs.

[45] A. Jircitano, "The Decomposition of Potassium Chlorate". Online at
http://chemistry.bd.psu.edu/jircitano/KClO3decomp05.pdf

[46] B. Johnston and M.A. Williams. 2007. A generic framework for approximate simulation in commonsense reasoning systems. *Commonsense-07,* pp. 71-76.

[47] B. Johnston and M.A. Williams. 2009. Autonomous learning of commonsense simulations. *Commonsense-09* pp. 73-78.

[48] L. Joskowicz. 1987. Shape and Function in Mechanical Devices. *IJCAI-87,* pp. 611-615,

[49] L. Joskowicz. 1988. Reasoning about Shape and Kinematic Function in Mechanical Devices. NYU Ph.D. Thesis.

[50] L. Joskowicz and E. Sacks. 1991. Computational Kinematics. *Artificial Intelligence,* 51: 381-416.

[51] L. Joskowicz, E. Sacks, and V. Srinivasan. 1997. Kinematic Tolerance Analysis, *Computer-Aided Design,* Vol. 29, No. 2.

[52] L. Joskowicz and R. Taylor. 1996. Interference-free insertion of a solid body into a cavity: an algorithm and a medical application. *International Journal of Robotics Research.* Vol. 15, No. 3.

[53] H. Kim. 1993. *Qualitative Reasoning about Fluids and Mechanics.* Ph.D. thesis, Institute of Learning Sciences, Northwestern University.

[54] V. Lifschitz. 1997. Cracking an Egg: An Exercise in Formalizing Commonsense Reasoning. Fourth Symposium on Logical Formalizations of Commonsense Reasoning.

[55] J. McCarthy. 1997. Elaboration tolerance. Fourth Symposium on Logical Formalizations of Commonsense Reasoning.

[56] R. Miller and L. Morgenstern. 1998. Commonsense Problem Page.
http://www-formal.stanford.edu/leora/cs/

[57] L. Morgenstern. 2001. Mid-Sized Axiomatizations of Commonsense Problems: A Case Study in Egg Cracking. *Studia Logica,* vol. 67 pp. 333-384.

[58] P. Needham. 2002. Chemical Substances and Intensive Properties. In J. Earley Sr. (ed.) *Chemical Explanation: Characteristics, Development, Autonomy*, Annals of the New York Academy of Sciences, Vol. 988, pp.99-113.

[59] P. Nielsen. 1988. A Qualitative Approach to Mechanical Constraint. *Proc. AAAI-88*, pp. 270-274.

[60] T. Nipkow, L.C. Paulson, and M. Wenzel. 2002., *Isabelle/HOL: A Proof Assistant for Higher-Order Logic,* LNCS 2283: Tutorial, Springer-Verlag.

[61] O. Raiman. 1990. Order of Magnitude Reasoning. In D. Weld and J. de Kleer (eds.) *Readings in Qualitative Reasoning about Physical Systems*, Morgan Kaufmann. pp. 422-434.

[62] E. Sacks and L. Joskowicz. 1993. Automated modelling and kinematic simulation of mechanisms. *Computer-Aided Design.* Vol. 25, No. 2. pp. 106-118.

[63] E. Sacks and L. Joskowicz. 1998. Dynamical Simulation of Planar Assemblies with Changing Contacts Using Configuration Spaces. *Journal of Mechanical Design,* Vol. 120.

[64] E. Sandewall. 1989. Combining Logic and Differential Equations for Describing Real-World Systems. In Proceedings of the First International Conference on Knowledge Representation and Reasoning, 412-420. Menlo Park, Calif: Morgan Kaufmann.

[65] M. Shanahan. 2004. An Attempt to Formalise a Non-Trivial Problem in Common Sense Reasoning. *Artificial Intelligence,* Vol. 153, 2004, pp. 141-165.

[66] J. Shin and E. Davis. 2005.“Processes and Continuous Change in a SAT-based Planner, *Artificial Intelligence,* to appear.

[67] T. Stahovich, R. Davis, and H. Shrobe. 2000. Qualitative rigid-body mechanics. *Artificial Intelligence* Vol. 119, pp. 19-06.

[68] Z. Wang, Incremental Web Search: Tracking Changes in the Web. NYU Ph.D. thesis, May 2006.

[69] R.A. Wehage and E.J. Haug. 1982. Dynamic analysis of mechanical systems with intermittent motion. J. Mechanical Design. Vol. 104, pp. 784-788.

[70] S. Wolfman and D. Weld. 1999. The LPSAT Engine and its application to resource planning. *IJCAI-99*.