An Integrated Systems Approach to Explanation-Based Conceptual Change

Scott E. Friedman & Kenneth D. Forbus

Qualitative Reasoning Group, Northwestern University
2145 Sheridan Road, Evanston, IL 60208-0834 USA
{friedman, forbus}@northwestern.edu

Abstract

Understanding conceptual change is an important problem in modeling human cognition and in making integrated AI systems that can learn autonomously. This paper describes a model of explanation-based conceptual change, integrating sketch understanding, analogical processing, qualitative models, truth-maintenance, and heuristic-based reasoning within the Companions cognitive architecture. Sketch understanding is used to automatically encode stimuli in the form of comic strips. Qualitative models and conceptual quantities are constructed for new phenomena via analogical reasoning and heuristics. Truth-maintenance is used to integrate conceptual and episodic knowledge into explanations, and heuristics are used to modify existing conceptual knowledge in order to produce better explanations. We simulate the learning and revision of the concept of force, testing the concepts learned via a questionnaire of sketches given to students, showing that our model follows a similar learning trajectory.

Introduction

Learning domain theories and changing them over time is a familiar task for humans, but an unsolved problem in Artificial Intelligence. The psychological task of conceptual change is a radical restructuring of knowledge (Carey, 1988), whereby concepts are differentiated (Dykstra et al, 1992), recontextualized, respecified (diSessa et al, 2004), and ontologically reorganized. Current computational models of conceptual change (e.g. Ram, 1993) do not work with automatically encoded stimuli, nor are they capable of modeling developmental trajectories found in the cognitive development literature. This paper describes a model of conceptual change, built on the Companions cognitive architecture (Forbus et al, 2009). Our system integrates sketch understanding to automatically encode stimuli, analogical processing to retrieve and apply relevant qualitative models, truth maintenance to manage explanations, and heuristics to modify conceptual knowledge in the face of anomalies.

We provide the system with a sequence of hand-drawn comic strips as stimuli for learning, automatically encoded using CogSketch (Forbus et al, 2008). Throughout the course of learning, we assess the system's knowledge of force dynamics, using a sketch-based questionnaire from diSessa et al (2004) and Ioannides & Vosniadou (2002). We compare the simulation's answers to those of students from grades K though 9 from the literature. We demonstrate that the system can induce and revise a domain theory of force dynamics from automatically-encoded relational knowledge, and that its concept of force changes with experience similarly to those of students.

We begin by discussing the task of learning a model of a physical domain and summarizing related work. We then discuss the individual components that our model uses, and then explain the unified system. We present the simulation results, and discuss future work.

Learning Physical Domains

Our system learns a physical domain from a sequence of input stimuli. This task has been investigated in cognitive science, machine learning, and computational scientific discovery. Systems such as QMN (Dzeroski & Todorovski, 1995) and MISQ (Richards et al, 1992) compute variable dependencies and qualitative constraints from numerical input data, which is important for learning physical domains. Similarly, Inductive Process Modeling (Bridewell et al, 2008) induces quantitative models from numerical data, which is useful for computational scientific discovery with numerical observations. Other scientific discovery systems such as BACON (Langley et al, 1987) introduced new quantities from observables. Our model is inspired by the experiential learning model of Forbus & Gentner (1986), and operates at the naïve physics level identified in that framework. Our model is closest to Falkenhainer's (1990) PHINEAS, which created qualitative models to interpret qualitative observations. Like PHINEAS, our system creates and revises QP models. Our system uses a cognitive simulation of analogical retrieval instead of the *ad hoc* indexing scheme used in PHINEAS, which it relies on to retrieve relevant domain

Friedman, S. E. and Forbus, K. D. (2010). An integrated systems approach to explanation-based conceptual change. Proceedings of the 24th AAAI Conference on Artificial Intelligence. Atlanta, GA.

Process m_1
Participants:
Entity *e*
Conditions:
nil
Consequences:
hasQuantity(*e*, *rate*)
i+(AxisPos(Horizontal, *e*), *rate*)
rate > 0

Heuristic addConceptualQtyCd
Participants:
CurrentState *s*
ProcessInstance *p*
ProcessType *t*
Conditions:
startsAfterEndingOf(*s*, *p*)
isa(*p*, *t*)
Consequences:
exists(***q***)
ConceptualQuantity ***q***
revise(*t*, addQuantityCond(***q***))

Process m_2
Participants:
Entity *e*
Conditions:
q(*e*) > 0
Consequences:
hasQuantity(*e*, *rate*)
α_{Q+}(*rate*, ***q***(*e*))
i+(AxisPos(Horizontal, *e*), *rate*)

Figure 1: Left: an early QP process model of movement m_1, created by the simulation; Middle: a heuristic that revises process models by adding a conceptual quantity; Right: the result of revising m_1 with the heuristic. Quantity *q*, a placeholder force-like quantity, is respecifiable by other heuristics. Note: α_{Q+} is the QP relationship of qualitative proportionality, and i+ is the QP direct influence relationship.

concepts, in contrast with PHINEAS' use of purely first-principles qualitative simulation.

Computational models of conceptual change should satisfy three constraints: (1) They should learn similar preconceptions as human novices, given similar experience, (2) their concepts should evolve along trajectories similar to those seen in human learners, and (3) their explanations should be comparable to those provided by human learners. An earlier Companions-based simulation by Friedman & Forbus (2009) illustrated that the same combination of techniques described here could satisfy the first and third constraints, using analogical generalization and statistical learning to generate initial models and to provide human-like explanations of its reasoning. Consequently, we focus on the second constraint here, examining the trajectories of conceptual change as anomalous information comes in.

Qualitative Process Theory

We use qualitative process (QP) theory (Forbus, 1984) to formally represent qualitative models. In QP theory, objects have continuous parameters such as position, rotation, and temperature, represented as *quantities*. The *sole mechanism assumption* states that all changes in a physical system are caused directly or indirectly by *processes*. Our model uses this assumption as a criterion for when an explanation is satisfactory. Figure 1 illustrates the representation for processes. *Participants* are the entities which are involved in instances of the process. *Preconditions* and *quantity conditions* describe when the process instance is *active*. *Consequences* are assertions that hold whenever a process instance is active. QP theory can represent a range of models: Forbus (1984) illustrates how QP theory process models can represent Newtonian dynamics, Galilean (medieval impetus) dynamics, and Aristotelian dynamics. Figure 1 (left, right) shows two early QP process models of motion generated by our simulation. The rightmost model resembles a component of the medieval impetus model of motion, similar to preconceptions held by many physics novices (Ioannides & Vosniadou, 2002).

Figure 2: Example learning stimulus.

CogSketch

CogSketch (Forbus et al, 2008) is an open-domain sketching system. CogSketch interprets the ink drawn by the user, and computes spatial and positional relations (e.g., `above`, `rightOf`, `touches`) between objects. Further, CogSketch supports multiple *subsketches* within a single sketch. We use this feature to create comic strips that serve as stimuli, where each subsketch in a stimulus can represent a change in behavior. Figure 2 depicts a stimulus from our simulation. Each subsketch represents a change in the physical system illustrated. Within each subsketch, CogSketch automatically encodes qualitative spatial relationships between the entities depicted, using positional and topological relationships. For example, the wheelbarrow is above and touching the ground in all three states, but the person and the wheelbarrow are not in contact in the first state. The arrows between the subsketches indicate temporal order, via the `startsAfterEndingOf` relation. Physical quantities such as area and positional coordinates are also computed by CogSketch. Using quantity data and temporal relations, the system can identify changes in physical quantities across states, which we refer to as *physical behaviors*. After physical behaviors are identified, they are stored in the scenario case using relations such as `increasing` and `decreasing` to represent the direction of quantity change.

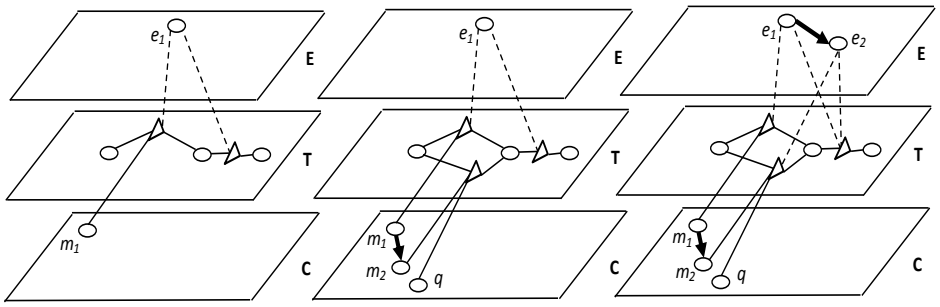

Figure 3: A tiered network. Nodes in the explanation layer (E) represent explanations composed of well-founded TMS support. The TMS layer (T) contains TMS nodes (circles) and justifications (triangles) that support physical behaviors. The conceptual knowledge layer (C) consists of processes and quantities. Left: before a model revision; Middle: after revising m_1 as m_2 using new quantity q (see Figure 1) Right: after computing a preference for e_2 over e_1.

Analogical processing

Our simulation uses similarity-based retrieval to find concepts to use in explaining new behaviors. We use MAC/FAC (Forbus et al, 1995) to model retrieval and SME (Falkenhainer et al, 1989) to model analogical matching. Given a probe case and case library, MAC/FAC efficiently retrieves a case from the case library that is similar to the probe. For scalability, its first stage estimates similarity via dot products on vectors automatically produced from the structured, relational representations used as cases. At most three descriptions are passed to the second stage, which then uses SME to compare their full relational versions to the probe, in parallel, to find the best. SME is based on Gentner's (1983) structure-mapping theory. Given two structured relational representations, SME computes one or two *mappings* which represent how they can be aligned. A mapping consists of *correspondences* which describe "what goes with what" in the two descriptions, *candidate inferences* that can be projected from one description to the other, and a numerical score indicating their degree of similarity. The simulation also uses SME mappings between similar scenarios to find differences in aligned quantities. This allows the system to generate domain knowledge hypotheses such as qualitative proportionalities.

Explanation-based Conceptual Change

Our model of conceptual change is driven by the explanation of stimuli. When a new stimulus is encountered, the system identifies changes in continuous parameters (e.g. position) and changes in relations between entities (e.g. surface contact, or directionality). It then explains these changes with existing models from similar past experiences. If necessary, the system creates new process models (e.g. Figure 1: left, right) and quantities to achieve a coherent explanation using heuristics (e.g. Figure 1: middle), using a simplicity bias to minimize the amount of revisions to its domain theory. This overall behavior involves four main operations: (1) retrieval of cases and concepts, (2) intrascenario explanation, (3) interscenario explanation, and (4) retroactive explanation. We discuss the general knowledge organization, and then describe each of these operations and their contribution to the larger process of conceptual change.

A tiered network model of conceptual knowledge

The organization of domain knowledge is depicted most easily in a tiered network, as in Figure 3. The top tier of the network shows explanations. Each explanation is a set of TMS justifications from the well-founded support for a physical behavior. The justifications that provide the well-founded support reside in a Truth Maintenance System (TMS) (Forbus & de Kleer, 1993) in the middle tier. The TMS contains a persistent record of inferences, including justifications for QP model instantiations. The nodes in the TMS represent facts from the stimuli and inferences made about them. The domain theory, consisting of QP models and hypothesized quantities created by the system, are plotted on the bottom tier, and are used as antecedents of TMS justifications. The directed edge $e_1 \rightarrow e_2$ in the explanation tier represents a preference for e_2 over e_1, and the directed edge $m_1 \rightarrow m_2$ in the domain theory expresses a preference for model m_2 over m_1, which are shown in Figure 1. These preferences between explanations and domain knowledge help guide future learning. Explanations and models are revised throughout the learning process, but the earlier versions remain, as a knowledge trace. Figure 3 shows the tiered network before model m_1 from Figure 1 is revised (left), after it is revised as m_2 (middle), and after a preference is computed for explanation e_2 over e_1 (right). The TMS associates physical behaviors with the conceptual knowledge that was used for explanation. This permits the retrieval and reuse of knowledge in similar scenarios, which drives the incremental process of conceptual change.

The physical behaviors and their supporting justification structure are stored within a set of scenario cases in long term memory. Most scenarios describe more than one physical behavior, such as Figure 2 which describes the translation of the agent and the translation of the

wheelbarrow. The relations linking the justifications to explanations (e.g. e_1 and e_2) are stored in each scenario case as well. A single case library containing all qualitative states the system has previously encountered (e.g. the three in Figure 2) is used for all retrievals.

Retrieving Cases and Conceptual Knowledge

Upon receiving a new stimulus, such as a foot kicking a ball along a surface, the system attempts to retrieve similar previously observed physical behaviors by using MAC/FAC. The state within which the physical behavior occurs serves as the probe. After retrieving a similar state from memory, such as a boy pushing a wheelbarrow to the right along a surface (Figure 2, middle), the system imports the domain knowledge used in the previous explanation into the current logical context, using the preference relations to select the current best explanation.

This method of similarity-based retrieval of past scenarios is the system's only manner of accessing existing conceptual knowledge. This constraint limits the search space of applicable domain knowledge, which reduces processing load. The act of explanation, discussed next, interprets the stimulus using this domain knowledge.

Intrascenario explanation

The process of *intrascenario explanation* explains the physical behaviors in a scenario using retrieved domain knowledge and heuristics. In some cases, first-principles reasoning with previous domain knowledge can explain the behaviors. In other cases, the system must revise its models or generate new domain knowledge. For example, suppose that the boy pushing the wheelbarrow (Figure 2) was the only previous stimulus the system had observed, and it only had the model of rightward movement m_1 (Figure 1, left; Figure 3) as a result. To explain why something slows down and stops moving without collision, the system might hypothesize a conceptual quantity q to mitigate movement, and revise the process model accordingly as model m_2 (Figure 1, right; Figure 3). This stage of intrascenario explanation is complete when all physical behaviors, such as movement, are explained using first-principles reasoning with domain knowledge.

During intrascenario explanation, the system might encounter an anomaly, such as: (1) a physical quantity changes in the scenario but no process model exists which could explain this in the domain theory; (2) a physical behavior starts or stops unexpectedly, due to an existing process model; or (3) conflicting assumptions are made about conceptual quantities. When an anomaly is encountered, domain knowledge is created or modified via *explanation heuristics* (e.g. Figure 1, middle). Syntactically, explanation heuristics have participants and conditions, like QP process models. However, their consequences describe operations on domain theory constructs, including revising process models or introducing new ones. When a physical behavior cannot be explained, explanation heuristics whose conditions are satisfied are found and applied in order of estimated simplicity. The simplicity metric is based on what operations the heuristics suggest, in order of increasing complexity: asserting that a new process instance is active in a state, revising a process model, revising a hypothesized quantity, hypothesizing a new process model, and hypothesizing a new quantity. The system instantiates process models and executes explanation heuristics until all physical behaviors in the state are explained. Heuristics allow incremental revisions for gradual conceptual change. The heuristic in Figure 1 revises a model and introduces a new conceptual quantity. Other heuristics respecify a quantity by changing its conditions for existence (e.g. it may only exist *between* two objects) or various dimensions of its existence (e.g. directional vs. adirectional, static vs. dynamic). The system currently uses 20 heuristics: 16 for intrascenario explanation and four for interscenario explanation.

Revising domain knowledge involves making changes to the formal specification of a quantity or model. Here, the system copies and revises m_1 to create m_2 (Figure 3, center). This preserves the old process model or quantity specification, so that previous explanations that employ m_1 are guaranteed to be valid, albeit outdated. $m_1 \rightarrow m_2$ states that m_2 is preferred over m_1. The result of the intrascenario explanation process is a series of well-founded explanations in the TMS, linking the physical behaviors to the domain knowledge that explains them. This potentially results in new and revised domain knowledge.

Interscenario explanation

Certain domain knowledge, such as qualitative proportionalities, can be induced by comparing two similar scenarios and explaining differences in behavior. This is achieved by the process of *interscenario explanation*. Suppose that the system has just finished explaining the behaviors within a new scenario of a foot kicking a small ball along a surface. Interscenario explanation begins by retrieving one prior scenario with MAC/FAC. If the normalized SME similarity score between the old and new scenario is above a threshold, the system performs interscenario analysis. Suppose that the retrieved scenario is a highly similar scenario of a foot kicking a larger ball along a surface, only a shorter distance. The system aligns the knowledge from the cases as well as QP knowledge from the respective intrascenario explanations to compare the quantity changes due to processes. Suppose these explanations employed the model m_2 in Figure 1. The system would align the rate parameters and the q influences, and could infer that size inversely influences q. Like intrascenario explanation, the inferences and conceptual knowledge revision in interscenario explanation are driven by declarative explanation heuristics.

Retroactive explanation

So far, we have described how the system makes local hypotheses and revisions, and annotates its conceptual and

explanatory preferences. These local changes must be propagated so that previous scenarios are explained using preferred domain knowledge. This is the process of *retroactive explanation*. This involves (1) accessing a previous scenario, (2) determining which domain knowledge currently used is not preferred by the system, and (3) attempting to explain the behaviors using preferred domain knowledge. If retroactive explanation fails to incorporate the preferred domain knowledge into new explanations, the failure is recorded and the old explanation remains the favored explanation. Unlike intrascenario and interscenario explanation, this process does not generate or change domain knowledge.

Simulation

We tested our conceptual change model on the domain of force dynamics. The system was given ten hand-drawn comic strips as training stimuli to learn models of motion. After each training stimulus, we use a sketched questionnaire designed to assess the development of the concept of force in human students, from diSesssa et al (2004) and Ioannides & Vosniadou (2002). We compare the test results of the simulation with results of human students. Though the simulation learns more rapidly than people, we demonstrate that the simulation's concept of force changes along a trajectory comparable to that of human students. We first discuss the results of the original experiments, then describe the simulation setup and compare its results to the trajectory of human models.

The changing meaning of force in students

Ioannides & Vosniadou (2002) conducted an experiment to assess students' mental models of force. They used a questionnaire of scenarios, each of which asked the student about the existence of forces on objects, varying from stationary bodies, bodies being pushed by humans, and bodies in stable and unstable positions. They found that several mental models of force were held by the students:

1. Internal Force (11 students): A force exists on all objects, or only on big/heavy objects. Force is proportional to size/weight.
2. Internal Force Affected by Movement (4 students): Same as (1), but also that moving and unstable objects have less force than stationary objects.
3. Internal and Acquired (24 students): A force exists due to size/weight, but objects acquire additional force when set into motion.
4. Acquired (18 students): Force is a property of objects that are in motion, or have the potential to act on other objects. There is no force on stationary objects.
5. Acquired and Push/Pull (15 students): Same as (4), but a force exists on an object, regardless of movement, when an agent pushes or pulls it.
6. Push/Pull (1 student): A force only exists when objects are pushed or pulled by an agent.
7. Gravity and Other (20 students): Mention of gravity and additional forces.
8. Mixed (12 students): Responses were internally inconsistent, and did not fall within the other categories.

The frequencies of responses by grade are listed in Table 1. These data suggest that young students favor the Internal model of force, and transition, via the Internal/Acquired model, to the Acquired model of force. By grade 9, students tend to adopt the Acquired/Push-Pull and Gravity/Other models. Ioannides & Vosniadou (2002) call these last models *synthetic models*, formed by selectively incorporating instructional knowledge into an intuitive conceptual framework.

Concept of Force	K	4th	6th	9th	Total
Internal	7	4			11
Internal/Movement	2	2			4
Internal/Acquired	4	10	9	1	24
Acquired		5	11	2	18
Acquired/Push-Pull			5	10	15
Push-Pull				1	1
Gravity/Other		3	1	16	20
Mixed	2	6	4		12

Table 1: Frequencies of meaning of force, by grade.

The simulation experiment

As can be seen from Table 1, conceptual change for children in this domain takes place over years, during which time they are exposed to massive amounts of information. Moreover, they are engaged in a wide range of activities, so even when exposed to motion, they may not be attending to it. Providing this amount of input and a similarly varied workload is beyond the state of the art in cognitive simulation. Consequently, we provide a much smaller and much more highly refined set of stimuli. We provided 10 comic strips, for a total of 58 comic strip frames and 22 instances of movement. These were created using CogSketch, as shown in Figure 2. The simulation computes relational and quantity changes between adjacent states in the comic strip as a means of detecting motion. For each training stimulus, the system does intrascenario explanation, and then retrieves a similar scenario to do interscenario stimulation. If models were changed or quantities were respecified, the system does retroactive explanation.

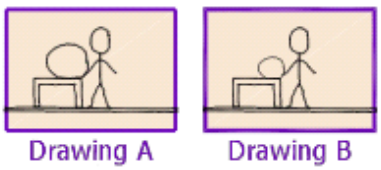

Figure 4: Example testing scenario.

By segmenting the comic strips into qualitative states, the system did not have to find the often-fuzzy boundaries between physical behaviors and is not dealing with noise.

Furthermore, the sketched data conveys relative changes in position, but not relative changes in velocity, so the system cannot differentiate velocity from acceleration, which is difficult for novice students (Dykstra et al, 1992). Finally, our learning stimuli were highly analogous, and there are only a small number of memory items, simplifying anomaly identification and explanation. We believe that this is why our simulation learns much faster than children.

We also used CogSketch to encode all ten comparison scenarios from the diSessa et al (2004) replication of the above experiment. These constitute the testing questionnaire given to the system after each training stimulus. Each comparison scenario contains two sketches that vary over a single variable such as the size of the rock in Drawing A and Drawing B in Figure 4. Like the students, the system is asked whether forces exist on the rock in each sketch, and how the forces differ in type or amount between sketches. This is accomplished by asking the system which conceptual quantities affect the rock in Drawing A, and the same for Drawing B. The system is then asked to compare the conceptual quantities acting on the rocks, and for those that are comparable, to say which is greater. From the system's answers, we can determine (1) the conditions under which a force-like quantity exists, and (2) the effect of factors such as size, height, and affecting agents on the properties of the force-like quantity. We use the same coding strategy as Ioannides & Vosniadou (2002) to determine which model of force the system has learned, given its answers during testing. The system had no initial model of motion or specification of a force quantity, so it relied on intrascenario and interscenario explanation to generate and revise its knowledge of force.

We use an interscenario explanation similarity threshold t = 0.95, so the system only attempts interscenario explanation when the examples are extremely similar. The stimuli in our sketch corpus are highly similar with minimal distracters, so we expected very rapid learning and an abundance of interscenario analyses.

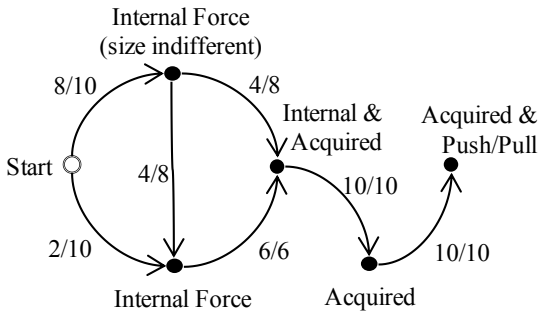

Figure 5: Changes in the simulation's concept of force.

Figure 5 illustrates the transitions in the concept of force across 10 independent trials with different training stimulus order. The simulation starts without any process models or quantities to represent force, and transitions to the Internal Force concept 2/10 times, and a size indifferent internal force model 8/10 times, which was not reported by Ioannides & Vosniadou (2002). We believe that this Internal Force concept without a qualitative proportionality to size/weight is a potential predecessor of the concepts of force identified in the original experiment. The rest of the transitions follow a similar trajectory to the student data in Table 1. Each trial of the simulation completed an average of six model revisions and four respecifications of a placeholder force-like quantity during its learning.

Discussion & Future Work

We have described a model of conceptual change built on the Companions architecture which integrates a variety of AI techniques to learn models of motion from sketched comic strips. Our experiment indicates that the model does indeed learn concepts of force dynamics given reasonable stimuli, and that the trajectories its concepts take are similar to those of human learners. This satisfies our three constraints for modeling conceptual change.

Each individual component of the system contributes to the larger process of conceptual change. Sketch understanding is used to automatically encode scenarios for learning and testing. Intrascenario explanation uses existing conceptual knowledge to explain the scenarios. When existing knowledge does not suffice, declarative heuristics are used to revise or create conceptual knowledge (e.g. hypothesizing and respecifying a force-like quantity). These local conceptual changes are recorded within TMS justification structure, which integrates conceptual and episodic knowledge into explanations. Retroactive explanation incrementally propagates local conceptual changes to other, previous scenarios (e.g. using new meanings of force in previous contexts). Finally, analogical processing is used to retrieve relevant conceptual knowledge based on similarities of behaviors and to hypothesize qualitative relationships between quantities (e.g. hypothesizing that force is qualitatively proportional to size).

A number of extensions will be required to more fully model human conceptual change. First, as noted above, our simulation is given only noise-free relevant data. Based on other experiments involving MAC/FAC, we expect that truly irrelevant data will not be a serious problem, since similarity-based retrieval scales well. Noisy data will require more work, as will incorporating other methods of change detection so that other domains can be modeled. Encoding is historically a serious problem; in early studies of heat, for example, the velocity of the mercury in a thermometer was measured instead of its final state, by analogy with measurements of motion (Wiser & Carey, 1983). Our model currently responds to anomalies by changing its concepts, but people have other responses. Chinn and Brewer (1993) identify a taxonomy of human responses to anomalous data, most of which involve avoiding conceptual revision. Similarly, Feltovich et al (2001) identified strategies called *knowledge shields*

that people employ to avoid conceptual change. Modeling this conservatism is an important next step.

One reason for human conservatism may be that initial concepts are introduced slowly, as local generalizations based on experience (Forbus & Gentner, 1986). Incorporating a model of local causal generalizations (e.g, Friedman & Forbus, 2009) would expand the range of the model and provide a statistical basis for how and when to be conservative.

The explanation heuristics that our system uses to modify its concepts do not exhaust the range of those available to people. Assuming QP models as representations for conceptual structure, we estimate that our current heuristics cover perhaps half of the heuristics that people use. We plan on exploring a wider range of heuristics, and other domains, as a means of finding a sufficient set of heuristics.

Finally, the model so far only deals with experiential learning. But social interaction and language-learning play important roles in conceptual change as well: Calling a phenomena a "flow", for instance, invites particular inferences about it (Gentner 2003), and one source of misconceptions is combining incorrect preconceptions with instruction. Moreover, the learning science literature argues for the importance of discussion, hypothetical situations, and analogies in fostering lasting conceptual change in scientific domains (Stephens & Clement, to appear). We anticipate incorporating these social aspects into our model of conceptual change, as learning from human interaction is a primary goal of Companion cognitive systems.

Acknowledgments

This work was funded by the Cognitive Science Office of Naval Research under grant N00014-08-1-0040.

References

Bridewell, W., Langley, P., Todorovski, L., & Dzeroski, S. (2008). Inductive process modeling. *Machine Learning*, 71, 1-32.

Carey, S. (1988). Reorganization of knowledge in the course of acquisition. In: Sidney Strauss, Ed. *Ontogeny, phylogeny, and historical development*. 1-27.

Chinn, C., and Brewer, W. (1993). The role of anomalous data in knowledge acquisition: a theoretical framework and implications for science instruction. *Review of Educational Research*, 63 (1): 1-49.

diSessa, A., Gillespie, N., Esterly, J. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, 28, 843-900.

Dykstra, D., Boyle, F., & Monarch, I. (1992). Studying conceptual change in learning physics. *Science Education*, 76 (6): 615-652.

Dzeroski, S., Todorovski, L. (1995). Discovering dynamics: from inductive logic programming to machine discovery. *Journal of Intelligent Information Systems*, 4 (1), 89-108.

Falkenhainer, B. (1990). A unified approach to explanation and theory formation. In J. Shrager and P. Langley (Eds.), *Computational models of scientific discovery and theory formation*. 157-196.

Falkenhainer, B., Forbus, K. & Gentner, D. 1989. The Structure-Mapping Engine: Algorithms and Examples. *Artificial Intelligence,* 41, 1-63.

Feltovich, P., Coulson, R., Spiro, R. (2001). Learners' (mis) understanding of important and difficult concepts. In K. Forbus & P. Feltovich, (Eds.) *Smart Machines for Education*. 349-375.

Forbus, K. (1984). Qualitative process theory. *Artificial Intelligence*, 24: 85-168.

Forbus, K. and Gentner, D. (1986). Learning Physical Domains: Towards a Theoretical Framework. In Michalski, R., Carbonell, J. and Mitchell, T. (Eds.), *Machine Learning: An Artificial Intelligence Approach.*

Forbus, K., Lovett, A., Lockwood, K., Wetzel, J., Matuk, C., Jee, B., and Usher, J. (2008). CogSketch. *Proceedings of AAAI 2008.*

Forbus, K. & de Kleer, J. (1993). *Building Problem Solvers*. MIT Press.

Forbus, K., Gentner, D. & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, 19(2), 141-205.

Forbus, K., Klenk, M., & Hinrichs, T. (2009). Companion Cognitive Systems: Design Goals and Lessons Learned So Far. *IEEE Intelligent Systems*, 24 (4): 36-46.

Friedman, S. & Forbus, K. (2009). Learning naïve physics models and misconceptions. *Proceedings of CogSci 2009*.

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

Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and cognition* (pp. 195-235). Cambridge, MA: MIT Press.

Ioannides, C., & Vosniadou, S. (2002). The changing meanings of force. *Cognitive Science Quarterly*, 2: 5-61.

Ram, A. (1993). Creative conceptual change. *Proceedings of CogSci 1993.*

Richards, B., Kraan, I., Kuipers, B. (1992). Automatic abduction of qualitative models. *Proceedings of AAAI 1992*. 723-728.

Stephens, L. & Clement, J. (to appear). The role of thought experiments in science learning. To appear in K. Tobin, C. McRobbie, & B. Fraser (Eds.) *International Handbook of Science Education*, 2. Dordrecht:Springer.

Wiser, M., & Carey, S. (1983). When heat and temperature were one. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp.267-297). Hillsdale, NJ: Erlbaum.