Artificial Intelligence 59 (1993) 105–114
Elsevier

ARTINT 990

A view on qualitative physics

Johan de Kleer

Xerox Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, CA 94304, USA

1. Introduction

The paper "A qualitative physics based on confluences" [18] by John Seely Brown and myself was one of the first papers in what is today a large and active area of artificial intelligence. As John Seely Brown has become the Chief Scientist of Xerox Corporation and is therefore somewhat distracted, this brief note represents my (Johan) personal views. [1] I will focus on some of the events and experiences which motivated me to become interested in qualitative physics and where it should be going.

When I was first exposed to the notions of symbolic computation and reasoning, these notions appeared as the ideal tools for the enterprise of understanding the physical world around us. Computers were already in widespread use in science and engineering for data analysis and simulation. However, even with the AI technologies of the 1970s, it was clear we could exploit computers in a far grander way. In particular, it became thinkable that we might be able to build systems that would be capable of reasoning about the physical world much as we ourselves, as engineers and scientists, do. I found this challenge enormously exciting because I saw (and still see) this as probably the only way that we will be able to build intelligences that can function successfully in the real world. Therefore, with the hubris characteristic of researchers in artificial intelligence, I set myself on a research programme with the goal of constructing what could best be described as an "artificial engineer" or "artificial scientist"

Correspondence to: J. de Kleer. E-mail: dekleer@parc.xerox.com

[1] Some of this material is extracted from the section "Qualitative physics: a personal view" in [33].

2. The confusion seminar

In 1974 I participated in a seminar run by Seymour Papert that was nicknamed "the confusion seminar". This half-year seminar proceeded through a sequence of confounding physical puzzles, sometimes spending weeks on a single example. These concerned commonsense questions about familiar topics such as bouncing balls, roller coasters, pulleys, rockets, strings, seesaws, swings, and juggling—as can be found in [7,32] and analyzed in [20]. The purpose of the seminar was to study thinking, the maxim at the time being, "You can't think about thinking without thinking about thinking about something." Examining the kind of mistakes one commonly makes while reasoning about the physical world serves as insight into how one thinks. I came away from the seminar with a completely different lesson. For almost all the examples we considered, the conventional mathematical formulation of physics was useless or unnecessary. It became apparent that most of the problems could be solved by simple qualitative reasoning alone. For most puzzles, it was eventually possible to resolve them to everyone's satisfaction by writing down at most one or two extremely simple equations. Invariably, writing down the full equations was the wrong tactic because either

(1) the equations were intractable because we couldn't figure out the right idealization or approximation,
(2) people weren't convinced unless the answers predicted by the equations were substantiated by intuition, or
(3) more often than not the seminar participants would write down the wrong equations and get the wrong answers.

3. Qualitative reasoning for solving physics problems

Although it was clear that qualitative reasoning of some sort is central for solving the kinds of puzzles that arose in the confusion seminar, I wondered if this would be true for more standard physics tasks, and therefore I started thinking about what knowledge was required to solve classical physical problems and how to build a system that could solve them. There had been two previous such problem solvers built: STUDENT (Bobrow [1]) and CARPS (Charniak [5]) (there have been many subsequent ones, for example, by Bundy et al. [3,4], Novak [24,25], and Skorstad and Forbus [28]). However, these two programs solved only an extremely narrow range of problem types. STUDENT, and to a large extent CARPS, primarily exploited the standard ways in which textbook problems are stated and incorporated essentially no knowledge about the physical world. The systems would produce nonsensical solutions for problems stated in nonstandard ways or which required even a rudimentary model of the physical world (e.g., that objects fall, boats are constrained to move on the surface of lakes, and masses are positive).

My research focused on a class of problems studied in the confusion seminar: roller coasters, or the movement of a cart (under gravity) on a track. The motion of the cart is governed by the laws of kinematics and Newton's laws, so it seemed like a simple place to start. The experience of building NEWTON [10], a system which could solve such problems, made the case for qualitative reasoning all the more compelling. For example, the research showed that qualitative reasoning is critical for comprehending the problem in the first place, formulating a plan for solving the problem, identifying which quantitative laws apply to the problem, and interpreting the results of quantitative analysis. In fact, the so-called physics knowledge of kinematics and Newton's laws are only a small fraction of the knowledge needed to solve problems. Most of the knowledge is "pre-physics", and considerable effort is required to codify it.

Physicists typically have a hard time understanding the goals of qualitative physics. When we present them some theory of qualitative reasoning (e.g., carts on roller coaster tracks), their first reaction is that they already knew that and see no point to developing a theory of such. When we answer by saying that we have explicitly codified this knowledge (e.g., about concavity) in such a way that a program can perform this reasoning, they remain unconvinced because they say the FORTRAN program they just built to analyze their data incorporated the same sort of knowledge. Under closer analysis, what they mean is that when *they* wrote the program to interpret a particular set of experimental data, they exploited this knowledge of their own. It is impossible to identify the location of this knowledge in the FORTRAN program, and a different program needs to be written for the next experiment. Typically, the program has no way to detect that it is producing nonsense results when the implicit assumptions under which it was written are violated. Qualitative physics aims to lay bare the underlying intuitions and make them sufficiently explicit that they are directly reasoned with and about. Only in this way can AI reasoners approach the rich and diverse capabilities of human problem solvers.

4. Qualitative and causal reasoning for circuit diagnosis

In 1974 I joined the SOPHIE project (summarized in [2]) at BBN led by John Seely Brown. SOPHIE was an Intelligent Computer Aided Instructional (ICAI) tool for teaching students diagnostic skills. One part of the task consisted of developing an AI-based diagnostician capable of expert-level diagnosis of an electronic power supply (and giving explanations for its conclusions). This diagnostician is faced with the formidable task of drawing diagnostic inferences from whatever measurements the student made. The project developed a system capable of diagnosing an electronic circuit from first principles alone. This subsystem, called LOCAL (de Kleer [9]), incorporated many ideas now included within model-based diagnosis (Davis [8]). It used knowledge about the known good and faulty behaviors of device components to detect inconsistencies between actual and predicted behaviors, which it then used to pinpoint the circuit fault.

Unfortunately, LOCAL suffered a very serious shortcoming: Although it was very good at pinpointing faulty components once two or three circuit measurements had been made, it was poor at drawing diagnostic inferences from initial measurements. It turned out that simple, first-principles reasoning about currents and voltages was insufficient to start diagnosing even fairly simple circuits, such as a six-transistor (and 31 other components) series-regulated power supply. This posed both pragmatic and theoretical difficulties: SOPHIE needed an automated expert-level diagnostician to pursue the instructional aspects of the research. To achieve this pragmatic goal, LOCAL was augmented with conventional expert-system rules to guide the initial phase of troubleshooting. These rules were tuned by trying hundreds of different circuit faults, comparing LOCAL's conclusions to those of a human expert, and adding rules to achieve what model-based reasoning missed. In the end, this was accomplished by augmenting LOCAL with only 39 circuit-specific rules.

LOCAL faced a difficult task:

(1) there were 37 potentially failing components;
(2) being an analog circuit, each component could fail in an infinite number of ways; and
(3) the student, not LOCAL, was in control of which measurements were made, so LOCAL had to make its inferences from any set of measurements whatsoever (there were 80 possible current measurements and 601 potential voltage measurements).

Thus it was extremely surprising that LOCAL, augmented with only 39 rules, would be able to localize the same set of faulty components as a human expert would.

Subsequently, we analyzed each of the rules to understand what model-based reasoning was missing. Consider one such rule: "If transistor Q5 is off, then the external symptom cannot be caused by the voltage reference being low." The rationale behind this rule is as follows: (Notice the qualitative form of the argument, not the content.)

> The power supply outputs a voltage and current limited by the minimum of its two control settings. This minimum is computed

> through a feedback path in which transistor Q5 is a central part. If the voltage reference were low enough to cause a symptom, then Q5 would be on, so there is no way that a low voltage reference could be contributing to the power supply's faulty behavior if Q5 was off. If Q5 were off, then the output might be high, but the output could not be too low because the reference was too low.

Notice that this argument is both qualitative (e.g., too high and too low) and reliant on an underlying causal model of the feedback operation of the power supply. Most of SOPHIE's 39 rules arise from a qualitative causal understanding of the power supply's functioning. Again, this made it convincingly clear that qualitative and causal reasoning were central for diagnostic tasks.

5. Qualitative and causal reasoning for design

The first phase of Sussman's engineering-problem-solving project made significant progress toward using AI techniques (in particular truth maintenance and constraint propagation techniques) for electronic circuit analysis [29,31]. Based on these successes, Sussman and I attempted to exploit the same underlying technologies for electronic circuit design. The program SYN [19] was based on the observation that as the previous analysis programs were based on constraint propagation, the usual process of determining circuit behavior from circuit parameters could be inverted to determine circuit parameters from desired circuit outputs.

SYN could successfully choose circuit parameters for simple one- and two-transistor circuits. However, somewhat to our surprise, even slightly more complex circuits quickly overwhelmed the memory available on the computers of the day. This difficulty manifested itself by the algebraic expressions among circuit parameters growing without bound. But even (successful) second-year electrical engineering students have little difficulty at the same tasks for more complex circuits. Closer examination revealed the following:

- Students know how components are functioning in the circuit and choose the least detailed model that produces a good enough answer.
- Students make (appropriate) algebraic approximations that simplify the algebra enormously without significantly changing the answer. Typically these involve feedback loops whose effect could be partially ignored.
- For any component, there are usually a set of equivalent models (because of circuit equivalences). Students know the one to pick which best simplifies the algebra. (See [30] for some examples.)
- Students know where to break feedback loops to make analysis simpler.

- Students are familiar with common component configurations and their models.

Even neophyte engineering students do not begin a symbolic analysis of even the simplest circuit without first understanding how the circuit functions qualitatively and the roles of the components within it. Armed with that understanding, the students select component models, variables, and approximations to perform the necessary symbolic computation and to interpret the results in terms of their commonsense understanding of how the circuit works.

This experience parallels the one with NEWTON: To perform what initially seems to be a simple analysis requires an extensive qualitative preanalysis of the situation. Although qualitative reasoning is critical to succeeding at even simple engineering tasks, it is usually tacit and rarely taught explicitly—making it all the harder to formalize in an AI system. The observation that qualitative causal analysis was critical to engineers' thinking was brought home by two other personal experiences during this time.

The first involved an electrical engineering design course I was taking. The course consisted solely of designing new circuits. A typical lecture would consist of the professor presenting a design and spending an hour analyzing it. He would pick models, variables, and approximations that simplified his analysis such that it seemed almost trivial. (Of course, students often found it difficult to replicate this performance on homework sets.) In one of the later lectures a student asked a simple question about a variation of the design being presented, which stumped the professor (a famous electrical engineer). For the remaining fifteen minutes of the class the professor stood close to the blackboard with his back to us to obscure his scribblings. I was sitting close enough so I could half follow what was going on. The professor was qualitatively simulating the effects of an input perturbation through the circuit with the goal of determining what role the components were playing. I could hear him: "If this voltage difference drops, then the forward bias on this transistor drops causing the collector current to drop ...." Toward the end of the class the professor quickly erased all his scribbles, stated the (correct) answer, and gave a very simple (and correct) rationale. This example illustrated two important things: (1) When a world class engineer was confronted with a novel variation he used qualitative analysis; (2) although he was unwilling to communicate that he actually used any kind of qualitative reasoning to obtain his conclusion.

The second experience involved a seminar series to which we invited analog VLSI designers from industry to visit and explain how they designed circuits. The practicing designers were much more up front about how they reasoned about circuits. When asked about how a particular circuit functioned or how they thought of adding such-and-such innovation, they would invariably launch into a qualitative and intuitive argument.

Motivated by these many examples of the use of qualitative knowledge for reasoning about electrical circuits, I developed a qualitative reasoner capable of doing some of the qualitative and causal analysis engineers perform. This program is called EQUAL [11–13]. Many of the issues that arise in electrical circuits are no different from those faced in other disciplines (e.g., fluid, thermal, mechanical). Indeed, the techniques used by EQUAL apply to all disciplines in which the lumped-parameter modeling assumptions apply (i.e., the classes of devices considered by Cochin [6] and Shearer et al. [27]). The paper "A qualitative physics based on confluences" [18] describes ENVISION which could do a qualitative and causal analysis with the standard lumped-parameter models and therefore analyze devices such as pressure regulators and spring-mass oscillators as well as complex circuits.

6. Whither qualitative physics?

Since the publication of "A qualitative physics based on confluences" research progress has been extremely exciting but at the same time a bit disappointing. Great strides forward have been made in understanding qualitative mathematics, time, shape, and function, etc. In addition, the requirements of qualitative physics have caused significant progress on topics as diverse as explicit control of reasoning, constraint propagation, truth maintenance systems, and knowledge representation.

Much AI research falls into the trap of examining issues for their own sake, losing sight of the overall objective, and thereby effectively building bridges over dry land. Whether or not one is interested in the ultimate goals of qualitative physics, focusing on reasoning about the physical world constantly brings fundamental issues to attention. Significant AI progress can be made only by applying AI ideas to tasks. Otherwise, we tend to spin our wheels.

Unfortunately, codifying qualitative knowledge about the physical world has proven to be surprisingly difficult. Just getting the qualitative version of the calculus (used to express qualitative knowledge) roughly right took many years of work by a large number of people. In retrospect it seems obvious what qualitative integration, differentiation, and continuity mean and what role they play—but these were not at all obvious at the beginning of 1980s.

It is disappointing that for most of the 1980s the tasks that originally motivated the development of qualitative physics were set aside. Too much of the research in qualitative physics loses sight of the original tasks by investigating some non-central issue raised by the exploration of some subtask. A mature field can afford to dissipate its energies in this way, indeed this is often a sign of maturity; but it is too soon for qualitative physics to lose that much vision of its original goals.

If qualitative physics is to have a significant impact on our science, we must seriously focus on the tasks that originally motivated its inception. For example, in exploring diagnosis we must develop theories of qualitative simulation and causality to enable pinpointing faulty components; in exploring design we must develop modeling techniques to support matching structures to functions.

Piecemeal steps toward these goals are not going to get us there. Qualitative physics must make larger steps toward its goals and set aside concern on the fine details until it becomes clear which details are worth carefully analyzing. Much more work is required such as Williams' [34], which develops a theory of innovative design using qualitative physics. The qualitative physics of engineered artifacts begs for a better language for describing device purpose. Even at this point qualitative physics is far from being able to construct anything like the cause–effect diagrams of Rieger and Grinberg [26], although de Kleer and Brown [16] describe some rudimentary ideas. Insufficient progress has been made using qualitative physics for diagnostic tasks or for integrating qualitative and quantitative knowledge. Currently the physical theories incorporated in most qualitative physics programs are developed painstakingly by hand. We have to find some way to accelerate this process (see Falkenhainer et al. [21]). Most qualitative physics researchers use their own individualized representations and reasoning methods. However, there is great commonality among the ideas. Perhaps qualitative physics should begin a CYC-like project to develop a common language for describing the physical world to be used throughout the qualitative physics research community (Lenat et al. [23]). Ironically, this brings us back to the central point of the naive physics manifesto (Hayes [22]), one decade earlier.

References

[1] D.G. Bobrow, Natural language input for a computer problem-solving system, in: M. Minsky, ed., *Semantic Information Processing* (MIT Press, Cambridge, MA, 1968) 146–226.

[2] J.S. Brown, R.R. Burton and J. de Kleer, Pedagogical, natural language and knowledge engineering techniques in SOPHIE I, II and III, in: D. Sleeman and J.S. Brown, eds., *Intelligent Tutoring Systems* (Academic Press, New York, 1982) 227—282.

[3] A. Bundy, G. Luger, M. Stone and R. Welham, MECHO: year one, in: *Proceedings Second AISB Conference*, Edinburgh, Scotland (1976) 94–103.

[4] A. Bundy, L. Byrd, G. Luger, C. Mellish, R. Milne and M. Palmer, MECHO: a program to solve mechanics problems, Working Paper 50, Department of Artificial Intelligence, Edinburgh, Scotland (1979).

[5] E. Charniak, CARPS, a program which solves calculus word problems, MAC-TR-51, Project MAC, MIT, Cambridge, MA (1968).

[6] I. Cochin, *Analysis and Design of Dynamic Systems* (Harper and Row, New York, 1980).

[7] H. Cohen, The art of snaring dragons, AIM-338, AI Lab., MIT, Cambridge, MA (1974).

[8] R. Davis and W. Hamscher, Model-based reasoning: troubleshooting, in: H.E. Shrobe and the American Association for Artificial Intelligence, eds., *Exploring Artificial Intelligence* (Morgan Kaufmann, San Mateo, CA, 1988) 297–346.

[9] J. de Kleer, Local methods of localizing faults in electronic circuits, AIM-394, AI Lab., MIT, Cambridge, MA (1976).

[10] J. de Kleer, Multiple representations of knowledge in a mechanics problem-solver, in: *Proceedings IJCAI-77*, Cambridge, MA (1977) 299–304.

[11] J. de Kleer, The origin and resolution of ambiguities in causal arguments, in: *Proceedings IJCAI-79*, Tokyo, Japan (1979) 197–204.

[12] J. de Kleer, Causal and teleological reasoning in circuit recognition, AI-TR-529, AI Lab., MIT, Cambridge, MA (1979).

[13] J. de Kleer, How circuits work, *Artif. Intell.* **24** (1984) 205–280; also in: D.G. Bobrow, ed., *Reasoning about Physical Systems* (North-Holland, Amsterdam, 1984/MIT Press, Cambridge, MA, 1985) 205–280.

[14] J. de Kleer and D.G. Bobrow, Higher-order qualitative derivatives, in: *Proceedings AAAI-84*, Austin, TX (1984) 86–91.

[15] J. de Kleer and J.S. Brown, Foundations of envisioning, in: *Proceedings AAAI-82*, Pittsburgh, PA (1982) 434–437.

[16] J. de Kleer and J.S. Brown, Assumptions and ambiguities in mechanistic mental models, in: D. Gentner and A.L. Stevens, eds., *Mental Models* (Erlbaum, Hillsdale, NJ, 1983) 155–190.

[17] J. de Kleer and J.S. Brown, The origin, form and logic of qualitative physical laws, in: *Proceedings IJCAI-83*, Karlsruhe, Germany (1983) 1158–1169.

[18] J. de Kleer and J.S. Brown, A qualitative physics based on confluences, *Artif. Intell.* **24** (1984) 7–83.

[19] J. de Kleer and G.J. Sussman, Propagation of constraints applied to circuit synthesis, *Circuit Theory Appl.* **8** (1980) 127–144.

[20] A.A. diSessa, Phenomenology and the evolution of intuition, in: D. Gentner and A.L. Stevens, eds., *Mental Models* (Erlbaum, Hillsdale, NJ, 1983) 15–190.

[21] B. Falkenhainer, K.D. Forbus and D. Gentner, The structure-mapping engine: algorithm and examples, *Artif. Intell.* **41** (1990) 1–63.

[22] P.J. Hayes, The second naive physics manifesto, in: J.R. Hobbs and R. Moore, eds., *Formal Theories of the Commonsense World* (Ablex, Norwood, NJ, 1985) 1–36.

[23] D.B. Lenat, M. Prakash and M. Shepherd, CYC: using commonsense knowledge to overcome brittleness and knowledge acquisition bottlenecks, MCC Tech. Rept. AI-055-85, MCC, Austin, TX (1985).

[24] G. Novak, Computer understanding of physics problems stated in natural language, Ph.D. Thesis, Department of Computer Science, University of Texas, Austin, TX (1976).

[25] G. Novak, Representations of knowledge in a program for solving physics problems, in: *Proceedings IJCAI-77*, Cambridge, MA (1977) 286–291.

[26] C. Rieger and M. Grinberg, The declarative representation and procedural simulation of causality in physical mechanisms, in: *Proceedings IJCAI-77*, Cambridge, MA (1977) 250–256.

[27] J. Shearer, A. Murphy and H. Richardson, *Introduction to System Dynamics* (Addison-Wesley, Reading, MA, 1971).

[28] G. Skorstad and K.D. Forbus, Qualitative and quantitative reasoning about thermodynamics, in: *Proceedings Eleventh Annual Conference of the Cognitive Science Society*, Ann Arbor, MI (1989).

[29] R.M. Stallman and G.J. Sussman, Forward reasoning and dependency-directed backtracking in a system for computer-aided circuit analysis, *Artif. Intell.* **9** (1977) 135–196.

[30] G.J. Sussman, SLICES: at the boundary between analysis and synthesis, AIM-433, AI Lab., MIT, Cambridge, MA (1977).

[31] G.J. Sussman, Electrical design: a problem for artificial intelligence research, in: *Proceedings IJCAI-77*, Cambridge, MA (1977) 894–900.

[32] J. Walker, *The Flying Circus of Physics with Answers* (Wiley, New York, 1977).
[33] D.S. Weld and J. de Kleer, eds., *Readings in Qualitative Reasoning about Physical Systems* (Morgan Kaufmann, San Mateo, CA, 1990).
[34] B.C. Williams, Invention from first principles: an overview, in: P.H. Winston and S. Shellard, eds., *Artificial Intelligence at MIT: Expanding Frontiers* (MIT Press, Cambridge, MA, 1990) 430–463.
[35] B.C. Williams, A theory of invention from fundamental principles of physics based on topologies of interaction, AI-TR-1127, AI Lab., MIT, Cambridge, MA (1989).