Indeed, a new observation by Alan Dressler of the Mount Wilson and Las Campanas observatories, one of the original Seven Samurai, suggests that the trans-Hydra-Centaurus region may be richer than it seems. In a recent nine-night observing run he obtained redshifts for some 600 galaxies in this part of the sky. Plotting the data along with 300 redshifts already available in the literature, he found that the distribution had two peaks. One, as expected, was at the distance of Hydra-Centaurus. But the other was 50% further away, at roughly 150 million light years. Moreover, it contained just as many points as the first.

Dressler, for one, argues that this second peak is exactly what it seems to be: the Great Attractor. To begin with, he says, galaxies are fainter when they are farther away, and we therefore see fewer of them. Correcting for that effect, one can argue that the second peak actually represents about four times as many galaxies as the first. Next, he says, one can convert the estimated number density of galaxies into an estimated mass density, and thereby come out with a mass for the region as a whole. The result works out to about ten times the mass of the Local Supercluster—which is just about what is needed for the Great Attractor.

Dressler's argument is clearly intriguing. But it is hardly proof—especially considering that it depends upon a series of assumptions about numbers and masses that other observers may want to question. Moreover, as Dressler himself points out, "to prove that the overdensity is doing the pulling you have to study other nearby mass distributions to make sure they aren't equally massive," he says. "Also, you have to study the velocities within this cluster to make sure *it* isn't moving." Nonetheless, he says, finding a major group of galaxies sitting in roughly the right place is an important step.

Assuming for the sake of argument that the Great Attractor does exist, then astronomers have to face a final question: how could such a thing form? It may not be an easy question to answer. The uniformity of the 2.7 K microwave background radiation implies that the universe was quite homogeneous when the radiation was emitted, about 100,000 years after the Big Bang. And yet, as observers have mapped our present-day universe on larger and larger scales, they have continued to find that matter is clustered on larger and larger scales. The theorists have been having enough trouble trying to explain the formation of clusters and superclusters of galaxies. The existence of structure on the scale of the Great Attractor may only make the challenge that much tougher. ■

M. MITCHELL WALDROP

Causality, Structure, and Common Sense

Ordinary common sense turns out to be far too subtle for conventional theories of logic; at a minimum it demands a much better accounting of such everyday notions as causality

IN their efforts to teach computers how to show "common sense," artificial intelligence (AI) researchers in recent years have found themselves paying more and more attention to such everyday notions as causality, structure, process, and time. These are the notions that underlie our intuitive understanding of the world. They seem easy and straightforward. And yet they turn out to be surprisingly difficult to pin down in any theoretical sense. Indeed, the struggle to capture these concepts in a computer-usable form proved to be one of the strongest undercurrents in the work presented at the recent annual meeting of the American Association for Artificial Intelligence (AAAI), which was held this July in Seattle. The very profusion of names for this research—qualitative process theory, naïve physics, and temporal logic, to mention just a few—was a testament to just how broad-ranging, how pervasive, and how unsettled the problem really is.

Two survey talks presented at the meeting gave some of the flavor of these efforts. The first, which addressed the critical role of causality in common sense reasoning, was given by Judea Pearl of the Cognitive Systems Laboratory at the University of California, Los Angeles (UCLA).

Consider the following situation, said Pearl: you go outside in the morning and you notice that the grass is wet. The obvious inference is that it rained during the night. In fact, you are almost sure that it rained. However, suppose that you now learn that someone left the lawn sprinkler on during the night. Suddenly, said Pearl, your confidence in the rain goes down considerably. In other words, upon receiving a new fact, you withdraw your original conclusion.

This kind of logical flip-flop, which is known in the AI community as "nonmonotonic" reasoning, is the epitome of common sense. Unfortunately, it is also a blatant violation of the conventional theory of logic, which is the first and most obvious place that one might look for a theory of automated computer reasoning. As formulated by generations of philosophers and mathematicians, the standard formalism of logic does offer an elegant way to represent facts about the world—as axioms—and it does provide a well-defined method for drawing conclusions from those facts: state each conclusion as a theorem and then prove that theorem. Indeed, it is for precisely this reason that AI researchers have spent so much time devising fast and powerful algorithms for computer theorem-proving. However, the standard theory of logic also implies that a new axiom (a new fact) can never change the validity of a previously proved theorem (a previous conclusion); the most it can do is allow the computer to prove new theorems that it could not prove before. In a word, conventional logic is "monotonic."

Clearly, then, something more flexible is needed for common sense reasoning. The question is what? Fool around with the rules of logical inference and it is all too easy to prove that grass is simultaneously green and purple.

This question of nonmonotonic reasoning has become something of a cause célèbre in the AI community during the past decade, and not only because of its deep theoretical significance. Those same 10 years have also seen a sharp rise of commercial interest in the so-called expert systems, which are programs that are supposed to give expert-level advice in fields such as medical diagnosis or tax planning. The expertise in these systems is ultimately provided by human specialists in consultation with the programmers. But because the knowledge is often uncertain ("If the patient has symptoms X, Y, and Z, then he *most likely* has disease D"), the program will almost always arrive at conclusions that are tentative—just as the human experts do. And for that very reason, expert systems have to be able to revise their conclusions ("The patient's nausea is caused by something she ate") on the basis of new evidence ("The patient is pregnant"). In other words, some approximation of common sense is absolutely critical.

To get a feel for the difficulty, said Pearl, consider the aforementioned lawn: after walking absent mindedly through the grass one morning, you look down and notice that your shoes are wet. The obvious inference is that the grass is wet. In addition, suppose that you now glance back and notice that the grass seems cold and shiny—that is, wet-looking. Your obvious impulse is to be more certain than ever that the grass is wet. After all, it is only common sense.

But therein lies the problem, said Pearl. In the first example, a new fact led to a retraction of the previous conclusion. Yet in this second example, the new fact *reinforces* the previous conclusion. In other words, two arguments that seem to have exactly the same logical form lead to radically different results. So how can this kind of reasoning ever be captured in a formal theory? And how can a computer ever reproduce it?

Pearl's answer, reduced to its essence, is to pay careful attention to the difference between cause and effect—certain aspects of which, he said, "have not received due treatment by logical formalisms of common sense reasoning."

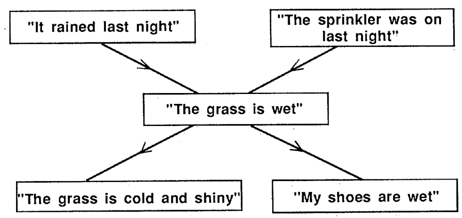
In the first anecdote about the soggy lawn, for example, the wet grass was an observed phenomenon—that is, an effect—whereas the rain and the sprinkler were two potential causes competing to explain it. Indeed, said Pearl, causes almost always compete: support for one tends to undermine our belief in any others.

In the second anecdote, however, the wet grass was the cause, whereas the wet shoes and the cold and shiny appearance of the grass were effects. That is, they could both be taken as evidence for the wet grass. Thus the asymmetry in the two cases, said Pearl: although causes compete, evidence cooperates. The more clues we have to support a given hypothesis, the more confident we are that the hypothesis is true.

These relations are illustrated in the accompanying diagram, where the arrows run from cause to effect. The diagram, in turn, is a simple example of a sophisticated construct that Pearl and his UCLA colleagues call "belief networks," which are intended as models of the way human beings draw causal inferences about the world.

In the full theory, which is still under development, these networks can become very complicated indeed, in keeping with the fact that the world—and our understanding of it—is also very complicated. As in real life, any given event can be the result of any number of possible causes, and can likewise produce a multiplicity of effects. Consider that many different medical conditions can produce a fever, for example, and that conversely, many different symptoms can be produced by a cold virus. (Statisticians and social scientists will recognize the resemblance between belief networks and path diagrams based on Bayesian probability theory. This is no accident: Pearl's work was originally inspired by Bayesian analysis and can be seen as an attempt to unify probability-based theories of common sense reasoning with theories based on classical logic.)

Nonetheless, said Pearl, no matter how complex a given belief network may be, the process of reasoning with those beliefs will still be governed by the same simple principles that illuminated the wet-grass example. The distinction in the way we reason about causes and consequences can be codified into a few straightforward rules of inference, rules that allow a computer to reach plausible conclusions on the basis of available evidence—and, of course, to revise those conclusions when new facts arise.



AAAI

Reasoning from causality. *The arrows in this diagram run from cause to effect. According to AI researcher Judea Pearl of the University of California, Los Angeles, such "belief networks" allow computers to exhibit many of the subtleties of common sense reasoning.*

Shortly after Pearl's talk, the AAAI participants heard a second survey talk that dealt with the issues of causality and process, albeit from a very different perspective. The speaker was University of Illinois–Urbana psychologist Dedre Gentner, and her topic was analogy.

Analogy is clearly a powerful tool in human thought. We reason by analogy. We learn by analogy. ("A motorcycle is like a bicycle, but with a gasoline engine.") And our language is rife with analogy in the form of metaphors. ("His face was set like granite.") So it is hardly surprising that this subject has become a major area of research within AI: if computers are ever going to exhibit common sense in the way that humans do, they are going to have to master analogy.

Nor is the researchers' interest purely theoretical. When programmers get together with human specialists to create a new expert system, the specialists will often articulate their knowledge as a series of anecdotes and specific cases. ("I remember one patient who . . . .") Indeed, this is typical of the way humans think: when confronted with a new situation, we often recall a similar situation that we have experienced in the past and then reason by analogy to apply that experience to the present. Unfortunately for the programmers, however, current-generation expert systems cannot do that. Instead, the programs have to have a set of clear-cut, general-purpose rules that can be used in a wide variety of situations. This means in turn that the programmers and the experts have to go through a long, painstaking process of trial and error to pin down what those rules really are. It would be far easier if the computer could just take the experts' anecdotes and cases as they are given, and then use analogical reasoning to apply them.

However, said Gentner, although analogy is obviously a very important tool in human thought, it is also a very tricky and dangerous tool. An analogy can all too easily be misunderstood—especially by a computer. When a machine is told that *John is a bear of a man*, for example, it may very well conclude that John is covered with thick black fur. And therein lies one of the central problems of analogy: how can computers (and for that matter, humans) sort out useful analogical inferences from those that are trivial, or misleading, or simply wrong?

Gentner's answer is her "structure mapping theory," which she developed in the early 1980s from extensive psychological studies of human analogizing. More recently she has implemented it as a computer program—the Structure Mapping Engine—in collaboration with Illinois AI researcher Kenneth D. Forbus and graduate student Brian Falkenhainer. (Forbus, as it happens, is also her husband.)

In essence, Gentner argues that the soundest analogies tend to be those that take account of the causal and structural parallels between two situations, as opposed to the superficial details. This is true even if the corresponding objects are utterly different in appearance or character; what matters is the roles they play.

A simple example is the analogy between fluid flow and heat flow, which is illustrated in the diagram on page 1299. To begin with, said Gentner, the computer (or the human learner) is given a description of the two situations and a set of facts about each, as shown in the bottom half of the diagram. Then the computer is told that heat flow is analogous to fluid flow; its task is to make the correct analogical mapping between the two situations and to draw the correct inferences.

To accomplish that task, said Gentner, the computer has to do four things, not necessarily in order. First, it has to find correspondences between the entities in the two domains. For example, *heat* corresponds to *water*, and *large beaker* corresponds to *warm coffee*. This is trickier than it seems, because a priori, *heat* might just as easily correspond to *pipe*. In practice, said Gentner, the computer is given a handful of heuristic rules that help it home in on the most reasonable choices.

Second, the computer discards isolated attributes such as LIQUID(water)—that is, "Water is a liquid"—and FLAT-TOP(coffee). By definition, these are the attributes that bear little relation to anything else in the situation; moreover, they tend to describe the superficial appearance of things. For both these reasons, said Gentner, they are probably not important for making the analogy.

Third, the computer looks in the water-flow domain for relations such as GREATER-THAN[PRESSURE(beaker), PRESSURE(vial)] —that is, "The pressure in the beaker is greater than the pressure in the vial." Then it maps them onto corresponding relations in the heat-flow domain: GREATER-THAN [TEMPERATURE(coffee), TEMPERATURE(ice cube)]. Once again, a number of different mappings may be possible.

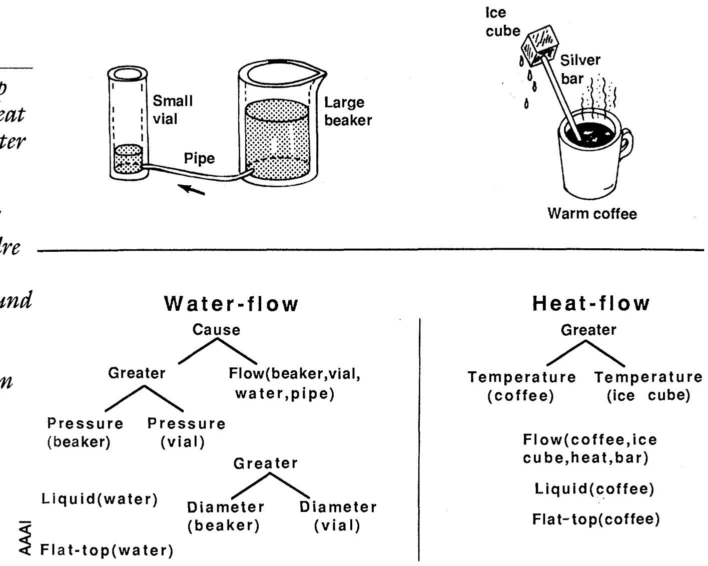
Finally, the computer resolves any remaining ambiguity by what Gentner calls the "systematicity principle": it looks for higher order relations, and keeps only those mappings that fit them. In the case of water flow, the GREATER-THAN relation between diameters is an isolated one and is therefore discarded. The only higher order structure is one governed by a CAUSE relation between the pressure differential and the water flow, which can be paraphrased in English as "The pressure differential between the beaker and the vial causes water to flow through the pipe."

As it happens, the computer has been given no such CAUSE relation in the heat-flow situation. However, this is not a problem, but an opportunity. The computer has been told about temperature differentials, and it has likewise been told about the flow of heat along the silver bar. So it makes an analogical inference: it conjectures that a causal relation *does* exist, and is identical in structure to the one in the water-flow example. In short, said Gentner, the computer uses the causal and structural relations in the problem to learn a new and potentially valuable fact: that the flow of heat is caused by a difference in temperature.

To date, the Structure Mapping Engine has successfully been applied to more than 40 different examples; these range from an analogy between the solar system and the Rutherford model of the atom to analogies between fables that feature different characters in similar situations. It is also serving as one module in a larger program known as PHINEAS, which is being developed by Falkenhainer as a model of scientific discovery. Basically, PHINEAS tries to explain newly observed physical phenomena in analogy with previously understood phenomena. To do so it couples the Structure Mapping Engine to another program module based on Qualitative Process Theory, which was devised by Forbus several years ago to provide a more formal account of structure and causality in such physical processes as motion, boiling, and liquid flow.

Reasoning by analogy

As illustrated in the top half of this diagram, heat flow is analogous to water flow. According to the Structure-Mapping Theory of University of Illinois psychologist Dedre Gentner, the essence of that analogy can be found in the causal and structural relations illustrated in the bottom half of the diagram.



Finally, said Gentner, the Structure Mapping Engine has been used as a model of human cognition. In one recent experiment, Gentner and her colleagues first asked their subjects to read a set of 18 short fables featuring characters such as Karla, the hawk who gave some feathers to a hunter, and Zerdia, the country that sold some of its computers to a neighboring country. A week later, the subjects were given 18 new stories featuring similar characters and situations; any time that the subjects were reminded of one the old stories, they were to write it out in as much detail as they could remember. Next, the subjects were given pairs of stories and asked to rate the soundness of the analogy in each case.

As the theory predicts, said Gentner, the subjects agreed that the soundest analogies were between fables that shared many underlying relations, even though their surface attributes were quite different. (The stories about Karla the hawk and Zerdia the country were one such pair.) Moreover, the Structure Mapping Engine reproduced the subjects' rankings quite closely.

Surprisingly, however, the subjects did just the opposite in the reminding task: they tended to recall stories that shared many superficial attributes with the current one, even if the analogy itself was weak. Gentner called such comparisons "mere appearance" matches (example: "Her eyes were as blue as the sky") and conjectured that they play an important role in identifying analogous situations in the first place.

In retrospect, of course, such a result is not so unreasonable. The superficial attributes of an object or situation tend to be its most obvious and easily recognized features; thus, they may be the most natural features to look for when one is searching through memory.

Nonetheless, said Gentner, this result does suggest that human performance in this area is governed by two very different set of rules: one concerning access to analogies, and the other concerning inference from analogies. Indeed, the Structure Mapping Engine can reproduce the human subjects' performance during recall, but only if the researchers turn off its ability to search for deep structural analogies and instead tell it to look only at the superficial details. On the other hand, intelligent programs are ultimately going to have to work at every level: accessing analogies, creating the proper map from one situation to the other, and then evaluating the quality of the map. The Structure Mapping Engine has already shown that it can model human performance in at least some of these areas. Perhaps these new experiments provide a clue for how such programs can go even further. ■ **M. MITCHELL WALDROP**

ADDITIONAL READING

J. Pearl, "Embracing causality in formal reasoning," *Proceedings of AAAI-87, The Sixth National Conference on Artificial Intelligence*, (1987), p.369.

J. Skorstad, B. Falkenhainer, D. Gentner, "Analogical processing: A simulation and empirical corroboration," *ibid.* (1987), p.322.