

## Making Sense of Common Sense Knowledge

*Benjamin Kuipers on using commonsense reasoning to make useful conclusions, or, finding gold nuggets in a pan of sand.*

Benjamin Kuipers holds an endowed Professorship in Computer Sciences at the University of Texas at Austin. A focal point of his research is "the representation of commonsense and expert knowledge, with particular emphasis on the effective use of incomplete knowledge."

**UBIQUITY:** Let's start with a definition of some terms. "Expert knowledge" seems reasonably clear, but, paradoxically, "commonsense knowledge" doesn't. Isn't it possible to have wildly different interpretations of what common sense is? And if that's true, what are the implications for representing such knowledge?

**BEN KUIPERS:** A long time ago, I wrote the following definition: "Commonsense knowledge is knowledge about the structure of the external world that is acquired and applied without concentrated effort by any normal human that allows him or her to meet the everyday demands of the physical, spatial, temporal and social environment with a reasonable degree of success." I still think this is a pretty good definition (though I might remove the restriction to the "external" world). These days, I identify commonsense knowledge with knowledge of certain "foundational domains" that children learn about at a young age: space, time, the conditions and results of actions, the properties of materials (hard, soft, gooey, breakable, etc), objects and their properties, and similar aspects of the physical world, but also certain aspects of the mental or social world, like the fact that

other people are agents, with beliefs, desires, and plans of their own. There is no hard-and-fast list of these domains, but space, time, change, and actions are clearly near the top of the list, so those are the things I like to work on.

**UBIQUITY:** Is your definition one that's generally accepted?

**KUIPERS:** There are other people working on representing commonsense knowledge who use different definitions from mine. John McCarthy has focused a lot of his attention on developing extensions to logic to express inferences that are part of commonsense. This has led to important innovations like nonmonotonic logic. Although I think this is an enormously valuable research development, I personally prefer to focus on these foundational domains of knowledge. Doug Lenat, with the CYC project, concentrates on accumulating the sheer bulk of knowledge that will make it possible to understand a story or explanation (originally, an encyclopedia article). Again, although this pioneering effort has been quite controversial in various ways, I think that CYC and other related projects are making a valuable contribution. Getting back to the foundational domains, I believe that we need to have good ways to represent space and time, for example, in order to reason about almost any kind of commonsense issue. Then change, motion, objects, and actions. There is no way to avoid these issues, and if you get them wrong early on, you may waste a lot of subsequent work. So my priority is to work on them.

**UBIQUITY:** Would it be wrong-headed to suggest that "common sense" is a very squishy term, since John McCarthy, Joe McCarthy, Eugene McCarthy and Charlie McCarthy would have had radically different and incompatible views of what is "commonsensical"? What is common about commonsense knowledge if there's no real

agreement on what commonsense knowledge actually means in ordinary life? And if commonsense knowledge is undeterminable, how can you build on it?

**KUIPERS:** Certainly the term "common sense" has been used by many people to mean many things (and we should not forget to mention Thomas Paine's "Common Sense" pamphlet that contributed to the American Revolution). However, within the field of AI, common sense has been recognized as a bottleneck problem since the early 1960s. This was particularly evident in the more logic-based approaches to planning.

**UBIQUITY:** Example?

**KUIPERS:** Okay, for example, suppose I am planning to travel to California. One of the actions available to me is to take an airplane, but that action has the prerequisite of being at the airport, which I am not. I can achieve the subgoal of being at the airport by driving there in my car. The prerequisites for that are having gas in the tank and my keys in my pocket. Well, I also need air in the tires, and electrical charge in the battery. Well, I also need to have wheels on the car, and a working engine. Well, it has to be a real passenger car, not a model car, or a painting of a car. Well, it has to not be welded to a light pole or filled with concrete. Well, it has to not be struck by a meteor on the way to the airport. And so on. At a certain point in this progression, normal humans start to get impatient, and say that while these requirements may be logical prerequisites for getting to the airport by car, common sense assures anyone that they are not real concerns. (But read the charming "Amelia Bedelia" stories for young children about the adventures of a totally literal-minded housemaid!)

**UBIQUITY:** Then how does AI go about focusing only on real concerns?

**KUIPERS:** If we want to create artificial intelligence (or equivalently, build computational models of the human mind) we will have figure out how to formalize that kind of common sense. AI researchers, like psychologists, are trying to understand the scientific basis of such squishy terms as "mind", "intelligence" and "common sense". If we succeed in the end, we will have clear computational models of some portion of what people mean when they use those terms. But just as biologists seldom argue about exactly what "life" means, we won't have to argue about exactly what "common sense" means. We will formalize what we need as a foundation for the theories we need to build. And it seems pretty clear that we will need theories of space, time, change, objects, actions, and so on.

**UBIQUITY:** You've said that in your research you place "particular emphasis on the effective use of incomplete knowledge." Tell us what you mean.

**KUIPERS:** One of the most striking things about human common sense is people's ability to make sensible decisions even when they don't know all the relevant information about a situation, and frequently even when some of what they do know is wrong! I've spent a lot of time studying the cognitive map: knowledge people have of large-scale environments like cities or buildings that they learn through exploration and travel. Some of the most interesting things about the cognitive map are the differences between people. Some people always keep themselves oriented with respect to the compass directions; many others do not. Some people can use the knowledge they have to find routes they have never traveled; others cannot. Sometimes, if you are asked how to get someplace,

you might find yourself saying, "I can't tell you how, but I could take you there." Still, all of these people generally find their way around pretty well, most of the time. Each of these corresponds to a different kind of incomplete knowledge. I believe that part of the power of human commonsense knowledge comes from the ability to represent and use knowledge even when it is incomplete. This makes it possible to decide between different theories of spatial knowledge in the cognitive map.

**UBIQUITY:** How would this be done?

**KUIPERS:** For example, there is a theory you might call the "Map in the Head" theory: that when you learn about places and streets in a city, you "draw" them into some structure in your brain that functions like a paper map does. But the "Map in the Head" is not very expressive of incomplete knowledge. Suppose I start at A, travel to B, then travel to C, putting them all on the map. Once I draw all of these on a single piece of paper, then the relationship between C and A is just as well specified as the relations between A and B, and between B and C. But this is often not at all true of human knowledge. If you are reading this at work, can you point to your home from where you are? Can you point in the direction facing out of your front door at home? Most people cannot.

**UBIQUITY:** Hmm. What does that suggest?

**KUIPERS:** It means that the cognitive map is more like an atlas than like a map on a single sheet of paper. Different places have different frames of reference, and they may be only loosely related to each other, or perhaps not at all. That is, we can refute the

"Map in the Head" theory. The connections between the places may work like topological links, rather than like geometrical shapes, in the sense that they don't let you draw conclusions about the relative orientation of beginning and end.

**UBIQUITY:** If there's no "Map in the Head," then what is there?

**KUIPERS:** In the case of spatial knowledge in the cognitive map, I claim that there are four broad classes of knowledge involved. First, the control level has continuous control laws, like the unconscious routine for walking down a corridor to the end without thinking about not bumping into the walls. Second, the causal level has isolated "distinctive states" (decision points, mostly), actions to take at those points, and the expected result of the action. Third, the topological level has places, paths, and regions, with relations like connectivity between places and paths, order of places on a path, and containment of places in regions. And fourth, the metrical level has quantitative information like distances, directions, and shapes of paths. The later levels tend to build on the earlier ones, and let you do more powerful inferences, but the earlier ones remain useful on their own. So if you have only a little knowledge, or you have a lot of knowledge but you aren't paying attention, you can often still get where you need to go.

**UBIQUITY:** Does this approach work in other applications?

**KUIPERS:** It does. I've also done quite a lot of similar work in a different domain, looking at knowledge of dynamical systems: systems that change continuously. For example, predicting what happens if you throw a rock into the air, or try to fill a bathtub with the drain open, or bounce a ball on the floor. It turns out that people can make very

reasonable predictions about these things, even though they don't have the specific numbers that would be required for an engineering model to run at all. It turns out that people can simulate these kinds of dynamical systems even with purely qualitative knowledge. One interesting thing about qualitative simulation is that, instead of getting a single predicted future like you would from a numerical simulation, you get a branching tree of possible futures. (The bathtub might overflow, or it might reach steady state between filling and draining. The rock might hit the ceiling before it starts to fall.)

**UBIQUITY:** Could you contrast the different roles of numerical and qualitative simulations?

**KUIPERS:** In both cases, the cognitive map and dynamical systems, simple qualitative knowledge is relatively easy to learn, and lets you solve a surprising number of problems. So it seems to be an important component of what makes commonsense knowledge so powerful. It also turns out that, in both cases, if you add more and more knowledge, you eventually converge to just as precise a model as the complete engineering model. But the engineering model only works once everything has been fully specified, while the commonsense model works even when it only has incomplete knowledge.

**UBIQUITY:** How do you go about verifying the validity of qualitative simulations?

**KUIPERS:** It's actually quite a neat trick. Imagine panning for gold in a river in California. You scoop a bunch of sand into your pan, and swirl it around, just enough to wash away the sand and dirt, while leaving the heavier gold behind. We're going to build a guarantee about qualitative simulation by showing that there must be gold in the pan to

start with, and that you throw away part of it, but never the gold. So the gold must be left when you are done.

Ever since Newton and Leibniz invented it, the language of differential equations has been the heart of physics and engineering. A differential equation can be thought of as a set of reasonably smoothly changing functions of time, plus a set of constraints that say how the different functions must be related to each other. For example, you can say that, at all times  $t$ ,  $x(t) + y(t) = z(t)$ , or that  $dx(t)/dt = w(t)$ , or several other similar things. (Traditionally, scientists and engineers hide most of these functions in more complex expressions, and just focus on one function and its derivatives, but this is just a notational preference.)

To simulate a differential equation, you start with values for all the functions at an initial point in time, and then figure out how the values can change. It turns out that it can follow only one path, and a numerical simulator tries to come reasonably close to that path.

A *qualitative* differential equation is just like an ordinary differential equation, except that you can use "monotonic function" constraints like  $y = M^+(x)$ . This means that there is some function  $f$  such that  $y(t) = f(x(t))$  for every  $t$ , but you don't know what the function  $f$  is, except that it is monotonically increasing. (That is, if  $x_2 > x_1$ , then  $f(x_2) > f(x_1)$ .) For example, if I am modeling filling the bathtub with the drain open, I know that the more water is in the tub, the faster it flows out of the drain. I don't know whether that relationship is linear, slower than linear, or faster than linear, but I do know it's increasing. (This, by the way, is a qualitative description of incomplete knowledge of a differential equation.)

I do a related thing in devising a qualitative way to describe the behaviors of functions like  $x(t)$  and  $y(t)$ . I can talk about a function as increasing, steady, or

decreasing, and I can describe its value as equal to one of a few special "landmark values" or between two of them.

Now, I look at the qualitative description of a changing variable. Suppose  $x(t)$  is the amount of water in the bathtub at time  $t$ , and it is between the landmarks EMPTY and FULL, and increasing. We can prove that there are only four qualitative changes that can happen:  $x$  can reach FULL and become steady; it can reach FULL and keep increasing; it can become steady before it reaches FULL; or some other variable could change while  $x$  is still increasing and less than FULL. There are no other options (because  $x$  and its derivative must both be continuous).

Finally, I can put these pieces together. If I start with a description of an initial state, I can predict the next possible qualitative states of each variable in the equation. All possible combinations of those states describe every way the system can go. Some of those combinations will be impossible, so I test them and filter out the impossible ones. Because our constraints are only qualitative, there may be more than one possible successor to the initial state. And these successor states may have successors of their own, as well, creating a branching tree of possible behaviors.

**UBIQUITY:** Which leaves us where?

**KUIPERS:** The point is: If I know I have generated a set that contains all the real solutions (like gold nuggets in the pan), then if I discard only bad ones (just the sand and dirt), I am guaranteed that the gold is still in the pan when I am done. That is, the real behavior is somewhere in the tree of predicted behaviors. That is the guarantee behind qualitative simulation. You can strengthen it with quantitative information, too, following the same approach. I believe that this will be useful, not just for commonsense reasoning,

but also for expert reasoning in situations like diagnosis and design where your knowledge is necessarily incomplete, but you want to draw useful conclusions anyway, and it would be helpful to have a guarantee.

[End]

See <http://www.cs.utexas.edu/users/kuipers/> for the Web site of Professor Kuipers.

*Source: Ubiquity, Volume 4, Issue 45, Jan. 14 - 20, 2004*