Mapping Meaning in Computation

I was hooked on AI quite early on. I was four years old, in fact, and I was reading a rather unusual book. The book was spiral bound at the top like a pad of paper. Its pages were made of heavy cardboard and they had been cut through into wavy segments. The effect was to tie four little books together side by side. Each of the little pages showed a crude picture and offered a phrase of text as a caption. You could, of course, turn the flaps separately, and make any selection from among the four sets of words and images. But all the flaps in each booklet agreed in *syntax*—in their formal structure as characterized by the grammar of English. So, amazingly, no matter what you did, each selection made a meaningful sentence, and narrated a comprehensible, if sometimes absurd, state of affairs. One "page" might read this way: *The firemen — rescued — the black cat — in the tall treetop*. And another might read this way: *Grandma — cooked — roast beef — for Sunday dinner*. If so, it was only a matter of time before you considered this alternative, cruel meal: *Grandma — cooked — the black cat — for Sunday dinner*. Or before you confronted this new, and perhaps even more transgressive, picnic: *The firemen — cooked — roast beef — in the tall treetop*!

Playing with this book was an electric experience. I pushed at the book to find a bad sentence or an impossible story, and I invariably failed; the book always presented me with a scenario that (with effort) I could make sense of. I may or may not have absorbed the uplifting '70s message about the fallibility of stereotypes. But I did understand that this book highlighted something profound and new about *myself*. This book embodied the fact that our statements and thoughts—the most intimate parts of ourselves—are constituted by computational rules. I'm still awed, in a way, that it is possible to experience this so directly. But we can. We can readily build mechanical tools that allow us to rediscover and reinvent ourselves by manipulating the fundamental units of which our minds are made. At their best, these tools take us by surprise and explain what and who we are in new ways that we can not only understand but grow from. For me, these possibilities epitomize the aims and rewards of Artificial Intelligence.

I think of Artificial Intelligence as the discipline we practice whenever we use computation as a conceptual framework to think more carefully about the mind. Computation is so closely associated with specific practical technologies that one could easily fail to appreciate how radically abstract this enterprise is. That's part of why I opened by talking about a book—not a computer system at all. It's all AI.

Computation is in fact just the mathematics of finite physical systems. These systems can be in only one one of a number of finite states; that means simply that you can write down a sentence that characterizes the input that is driving the system, the memory the system is maintaining, and the output the system is exhibiting. So in fact, we could characterize my little book in a computational framework, by writing down a sentence that says which four flaps are facing up. A laptop computer is engineered with however many megabytes of memory: enough to fix the state of the operating system, the user interface, and all the various programs—calendar and scheduling, email and communication, document preparation and information access—that you typically interact with. A laptop is bigger, but qualitatively it's the same: we've built it to behave discretely, like the simple book. When you put your laptop to sleep, it creates the sentence that names its state, and writes that to persistent storage; by reading that sentence in, it can start up again where it left off. Our own minds offer an even richer repertoire of discrete possibility. But we usually think of ourselves as finite. With a long enough report, you could exhaust what you know, what you remember, what you want and hope, your intentions and incomplete deliberations. It would make a vast tome—but ultimately, it too must come to an end. Fundamentally, computation is an abstraction that fits all three cases equally.

As time passes, computational systems pass from one finitely-specified state to another finitely-specified state, according to patterns that can themselves be described in a finite way. There is no magic; nothing infinitely large, nothing infinitesimally precise. We can say what is happening, in a sentence. The study of computation focuses specifically on these descriptions of evolving discrete systems. These descriptions are interesting, it turns out. They are rules, programs, insights: patterns for creating new behavior and clues to understanding our own. They are at the heart of Artificial Intelligence.

When I do AI, I'm particularly concerned—still—to think more clearly about the process of communication. In my research, I design, implement and analyze models that explain how intelligent agents, including ourselves, can come to share ideas using natural language. The phenomenon of communication shows how essential computational thinking is to any sensitive account of our own mental life. To make herself understood, a speaker must plan a finite message and thereby bring the state of her own thought and that of her audience into correspondence. If speakers can succeed in making meaning, and if listeners can succeed in recognizing it, we can *only* map the mutual understanding they achieve within a computational framework.

Computational thinking is just the starting point for getting clear on the processes that make up our everyday conversations with one another. The challenge really begins once we try to reconcile our apparent abilities to communicate effectively with the underspecification and vagueness we actually find in the rules of language. For example, our knowledge of language doesn't say once and for all what animal is *the black cat* or what man is *him*; that is left unspecified and depends on context. Our knowledge of language also doesn't say exactly how tall something *tall* is; that's left vague, and it's up to the context to supply some relevant standard for which things will count as *tall*. When we talk, we fill in these gaps by connecting our words creatively to our understanding of the shared world we inhabit, and we work together to ensure that we succeed in communicating with each other. Meaningful language use thus reflects the complex coordination of an agent's perceptual abilities, social skills, linguistic knowledge and rational decision-making. So even though our success in making and recognizing meaning must have the same kind of computational explanation as our success in making and recognizing the syntactic forms of grammatical sentences—the skill my little book exploited way back when—a theory of meaning will require us to come to grips with a profoundly richer and broader interplay of discrete structures and computational mechanisms.

Indeed, we are just beginning to have computational accounts that genuinely help us in our reflections on communication. But the accounts we do have hint at untold ways that a computational description of our communicative competence will enrich our understanding and our experience of ourselves. For instance, we can expect increasingly computational research unfolding in fields like literature, anthropology, sociology and art. As we succeed in framing precise and general enough analyses of particular meanings we value, we can start to build computational tools to help us explore, appraise and even create these meanings. At the same time, we can expect computational insights to ramify throughout our commonsense explanations of our everyday skills and behaviors. For instance, we can hope to teach children to be more proficient readers and writers by engaging them with simple implementations of meaning-making and calling their attention to the computations the mind uses to link words to their interpretations in context. Talking robots, and the other ingredients of the science fiction sometimes associated with triumph in Artificial Intelligence, will come last. They presuppose our ability to frame these kinds of computational research and insights. Natural language technology is exactly as far off as computational cultural theory or algorithmic elementary school language arts curricula. In all cases it's the same ideas we're missing: a precisely-articulated understanding of the finite basis of the rich and productive knowledge and inference we use to make and recognize meaning.

It's no surprise we're a long way off. After all, today's AI researchers were lucky to have books in their kindergarten classrooms that embodied the insight that language even has rules in the first place. When we turn to everyday practice, the insights AI offers continue to seem small and hard-won. I once spent a year working on a program called "animated conversation" that portrayed a couple of talking robots.[1] The robots overacted absurdly. They signaled whenever they were triggered to signal by the rules we had programmed—regardless of whether their partner needed the signal or not. It was usually not. I associate that year's effort principally with a clear statement of our mistake: you simply cannot formulate a coherent natural language utterance without considering how it will be *interpreted* in context by your audience. I spent another few years on a system called "SPUD" that aimed to fix the bug.[2] SPUD made sure that it distinguished its intended meaning from salient alternatives in context; once it had done this it stopped adding words in. But even SPUD was too verbose; I had neglected to factor in the role the ongoing task plays in disambiguating utterances. Everyone knows you don't have to consider an interpretation if it wouldn't have made any sense to say it, but I was shocked at how often people actually rely on this kind of background to rule out bad interpretations! Even simple models need this principle urgently. My graduate student David DeVault and I have just spent a few more years writing a program called "RUBRIC" that uses both the context and the ongoing task to dis-

---

[1] Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost and M. Stone. *Proceedings of SIGGRAPH*, 1994, pages 413–420.

[2] Microplanning with communicative intentions: The SPUD system. M. Stone, C. Doran, B. Webber, T. Bleam and M. Palmer. *Computational Intelligence* 19(4):311–381, 2003.

ambiguate utterances in dialogue.[3] We have, at last, got qualitatively different behavior. RUBRIC, it turns out, is too terse. So far we haven't been inclusive enough in our model of the different ways that dialogue can unfold. Because RUBRIC thinks there are fewer possibilities than there actually are, RUBRIC thinks it's easier than it really is to disambiguate what it says, and it uses fewer words than it really should. Again, this experience culminates in a simple and awfully hard-won insight: people are not only proficient in the ways they coordinate with each other in conversation, they are eclectic and flexible. Conversational systems need a lot more social knowledge, so they can map out these eclectic possibilities, and we are still a long way from getting clear about what's involved.

I keep doing AI. In fact, I think the AI programs I have worked on have been great successes. These caricatures of my work don't convey the unique significance of computational research. Insights in AI are realized in representations, algorithms and architectures that offer unambiguously better ways to think about the phenomena that make up mental life. New researchers who adopt these new frameworks need not even *avoid* the mistakes that others have previously made: these errors are simply *unthinkable*. I'll *never* make the same mistakes again. In AI, we don't just build on one another's work; we really do stand on one another's shoulders.

Neither do the caricatures I have offered convey the unique freedom of AI research. AI admits speculative work more readily than the other disciplines of cognitive science; if you can dream it and

[3]See for example: An information-state approach to collaborative reference. D. DeVault, N. Kariaeva, A. Kothari, I. Oved and M. Stone. *ACL Proceedings Companion Volume—Interactive Poster and Demonstrations*, pages 1–4, 2005.

think it through and build it, you can count on your work seeing serious consideration. If something really *works*—if it really does something interesting and new—nobody can contest the achievement by haggling over unexplored counterarguments, conflicting intuitions or missing experimental controls. Rather, we are all confronted with the rich, collaborative effort of explaining why.

My own projects have challenged my understanding of conversation, sharpened the distinctions I use when I categorize events in conversation, and discovered new principles and relationships among these categories. When I present my work, it is read not only by AI researchers, but also by linguists, philosophers, and psychologists—by anyone who cares about meaning. I happen to use computational modeling to study the phenomenon, whereas others study it by introspection, critical analysis, and human-subjects experiments. The fact remains that *none* of us *really* understand meaning well enough to say *precisely* how it works. That's the only real reason why there are as yet no talking robots. But just you wait. We're all accumulating our separate, little hard-won insights. We *can* get it right. And it will be amazing.

— Matthew Stone
Edinburgh and Rutgers