

Viewpoint

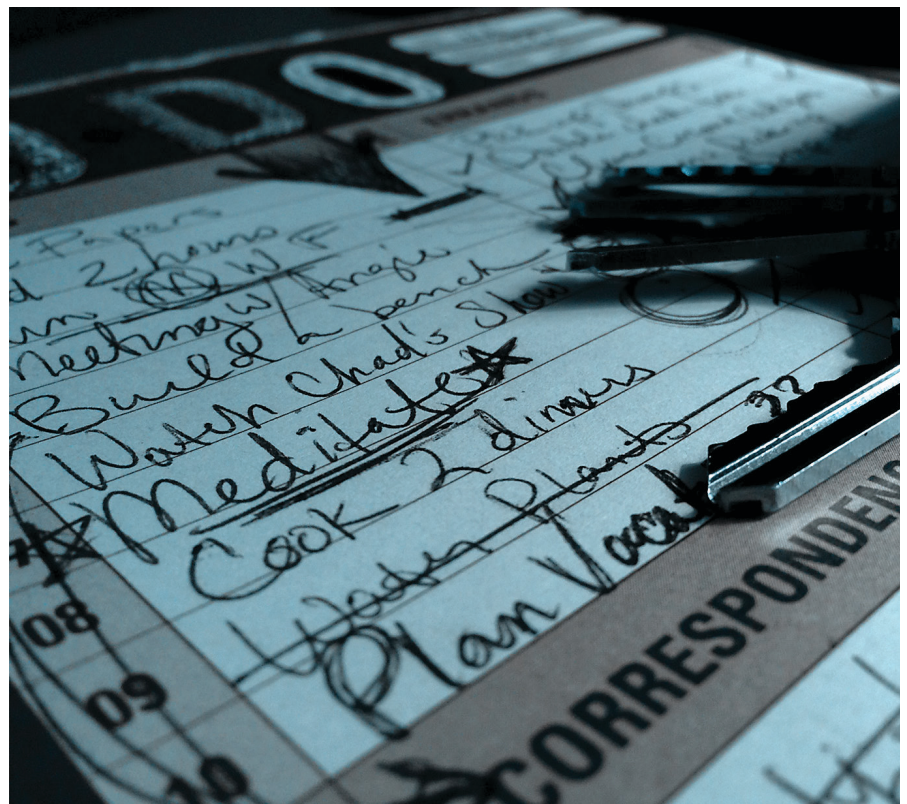
Why Knowledge Representation Matters

A personal story: From philosophy to software.

THERE IS A big difference between the attention artificial intelligence (AI) is currently receiving and that of the 1990s. Twenty years ago, the focus was on logic-based AI, usually under the heading of knowledge representation, or KR, whereas today's focus is on machine learning and statistical algorithms. This shift has served AI well, since machine learning and stats provide effective algorithmic solutions to certain kinds of problems (such as image recognition), in a way that KR never did. However, I contend the pendulum has swung too far, and something valuable has been lost.

Knowledge representation is not a single thing. While I think an argument could be made about KR as a whole, I will be focusing on the “applied philosophy” aspect of it—the logical representation of commonsense notions, with an emphasis on clear semantical underpinnings.

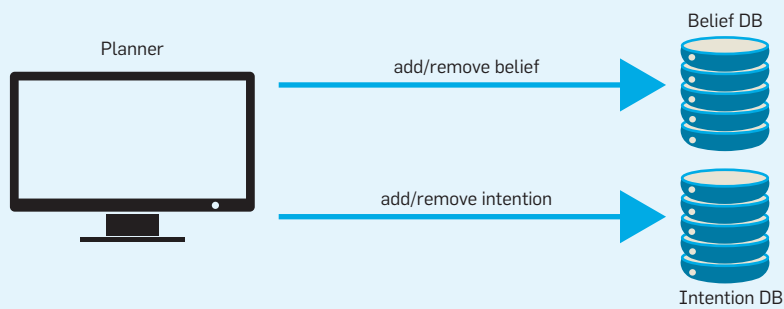
I will make the case for the most part through a personal story. The story starts with a paper I published in 2009 in the *Journal of Philosophical Logic*, continues with a research project at Stanford and Duke, later with a company called Timeful, and concludes with Timeful being acquired by Google in 2015. The point of the story is there is a direct link between the original journal paper and the ultimate success of the company.



The journal paper was “Logics of Intention and the Database Perspective.”⁶ This paper followed an important though thin strand of papers in AI on the logic of intention, spawned by Cohen and Levesque’s seminal “Intention is Choice + Commitment.”⁴ This literature in turn was inspired by the less formal litera-

ture in philosophy on rational agency, such as Bratman’s “Intentions, Plans, and Practical Reason.”³ My own paper took inspiration from the Cohen and Levesque paper but questioned its foundations, and proposed an alternative approach. Although my approach was computationally motivated (as indicated by the title),

Proposed database perspective.



the arguments were theoretical and philosophical in nature.

Following that journal paper I sought some funding to continue the research, as professors tend to do. And as funders tend to do, my would-be funder requested that I include some potential applications of this work. Then several things occurred to me. The first was I deal with intentions all the time—in my personal calendar. The second was these were intentions of a very specific kind—rigid events and meetings. And the third was that my personal calendar was not all that different from that of my late grandfather, which is odd given how the demands on people’s time have changed, and how technology has advanced. This led to an obvious question: What would happen if I enhanced the calendar with richer and more flexible intention types, and the calendar had the intelligence to help deal with the resulting complexity?

To understand this point better, it is worth discussing the ideas in the journal paper a bit further. The proposed database perspective is encapsulated in the accompanying figure, which can be thought of a generalization of the AGM scheme for belief revision,² the latter being restricted to the “belief” part of the picture. In the AGM framework, the intelligent database is responsible not only for storing the planner’s beliefs, but also ensuring their consistency. In the enriched framework there are two databases, one for beliefs and one for intentions, which are responsible for maintaining not only their individual consistency but also their mutual consistency. In the journal paper I laid out the main consistency conditions, and in a subsequent paper with Icard and Pacuit⁵ we gave a logical formalization

of it, which is a conservative extension of the AGM framework. It is not appropriate in this Viewpoint to go into more technical details, and indeed many of them are not relevant here. What is important to take away is the view of an intention database that performs intelligent functions on the part of the agent.

Returning to the storyline, the funder was persuaded, and we started a small project to explore these ideas. The next two years were fun but there is not much to say about them that is relevant to the story here, except: the project was soon led by a new Ph.D. student, Jacob Bank; it was also joined by my longtime friend and colleague Dan Ariely, a renowned behavioral economist; and by the beginning of 2013 we decided to start a company, which eventually came to be called Timeful. We were not so much driven by the specifics of our joint research up to that point, as by the realization of how acute the problem of time management was in society, and how ill suited current tools were to deal with it.

When Timeful 1.0 came out in July 2014, the reaction from both users and press was very favorable. Some 2,000 user email messages poured in during the first month, many of them emotional. Timeful had clearly struck a nerve, even if the product still had a way to go. Very soon the company attracted interest from major players, leading to the eventual acquisition by Google. None of this would have happened were it not for KR; here is why.

Intention Objects as the Basic Data Model

Timeful developed the concept of the Personal Time Assistant (PTA), whose role it is to help manage time, the resource that is both the scarcest and

the most difficult to manage. The approach rested on three main pillars. The first was allowing the user to naturally represent in the system everything that vied for their time. The second was the application of machine learning and other algorithms to what is inherently a hard optimization problem. The third pillar was behavioral science, which meant crafting an environment that subtly helps correct for natural time-management mistakes we all make (such as procrastinating, and overestimating our future availability). Of these, it is the first pillar I want to focus on; it was the most fundamental of the three, and the one based directly on KR.

Consider all the things that vie for our time: meetings, events, errands, projects, hobbies, family, health maintenance, sports, or just time to think and recharge. They are all superficially different, and historically reside in different applications (meetings and events in the calendar, errands in a to-do list, projects in a project-management system) or simply stay in our head. But they all vie for the same resource—time—and if you are to make intelligent trade-offs, they ought to reside in the same place. And indeed, they are all intentions, albeit with different properties. Following the vision of the intelligent intention database, the first fundamental decision was to develop a data model rich enough to encompass all these intention types. The result was a data model called the intention object (IO). An IO is a feature vector that includes a textual description, temporal attributes (when it can be executed, when it should be, its duration—all specifiable at various degrees of precision), conditions for executing the intention (such as location, or tools needed), and other attribute types.

Intention objects became the foundation for the system, and everything—including the algorithmic scheduling and the behavioral nudges—hinged on them. Of course, the user was not presented with a feature vector, but rather with several pre-packaged classes of intentions. As of April 2015, there were four classes: Events (such as meetings); tasks (such as making a phone call); habits (such as jogging three times a week); and projects (such as writing a

long report). But under the hood, for the system they all broke down to feature vectors.

More Product Decisions

Knowledge representation as not only the original impetus for Timeful and the inspiration for its data model. The team repeatedly found itself seeking guidance from the philosophical literature when making specific product decisions. It is difficult to fully convey this, but here are two concrete examples.

The first example has to do with the modest checkmark. Every to-do list allows you to check off tasks accomplished. Timeful had this feature too, but it bothered us that tasks had checkmarks and events did not, even though they were both IOs. It was not so much the aesthetic asymmetry, but more the underlying principle, and how that principle should be applied to other IOs, such as habits and projects. Then we went back to our roots and realized it had to do with tracking one's commitments. If there is one principle the philosophical literature agrees on it is that intention involves commitment (as reflected in the very title of the Cohen and Levesque article). When I intend to do something, it is not that I merely make a note of it; I am committed to tracking it and making it happen. When seen in this light, we realized events do not require tracking; a meeting is accomplished by being scheduled (there are exceptions, such as when the meeting has a goal that may not be achieved, but those were handled by specifying a separate task associated with the meeting). All other intention types require explicit monitoring, and so we ended up attaching checkmarks to all IOs except events.

The second example has to do with the temporal scope of an intention. Most to-do systems are “lists of shame”—things you write down but never do. We wanted to avoid that, and did it via strict time scoping. This early decision traced back to a mini-debate in the literature. In the Cohen and Levesque formalism, statements such as “I intend to read this book” are the basic concept. But in my journal paper I argued this is problematic, and it goes back to the issue of commitment. If I am committed to an intention that is

When you build a product you want it to be beautiful on the inside.

not anchored in time, what exactly am I committing to, and how does it actually drive action? (If you have a teenager at home you know what I mean.) Instead, I argued, the basic construct should be statements such as “I intend to read the book from 2 P.M. to 4 P.M. on Saturday.” You can then relax those by existential quantification, and say things such as “I intend to read the book for two or three hours sometime this weekend.” But you are always explicit about the time scope. Timeful adopted this philosophy; the implicit contract with the user was that she should be serious about her intentions, and in return the system would help her accomplish those by placing them on her calendar and prodding her to get them done (the tagline when the app launched was “get it scheduled, get it done”). Thus every task required either a specific “do on” or “do by” date. The task then appeared on the time grid, alongside the events. (In the case of “do by,” the system selected a time before the due date, which the user then could change if needed. Indeed, if an event later displaced the task, the system would move the task automatically.) The same logic applied to habits and projects, in more involved ways.

Conclusion

The story of Timeful is a happy one, and much of the credit goes to KR. Could one have arrived at the same insights without KR or philosophy? Possibly, but the fact is no one did, and I do not think that is an accident. When you build a product you want it to be beautiful on the inside. What I mean by this is that often, when you set out to design a great user experience, you either do not have the conceptual vocabulary with which to do it well, or, worse yet, you are fighting an existing conceptual framework and data model. And if the

internal structure is not right, you will never have a truly beautiful user experience. Philosophy and KR encourage you to think rigorously about your conceptual architecture, and provide guidance when designing specific features.

This does not lessen the importance of machine learning or statistics. But machine learning requires a feature space, and stats require an event space. Even the most avid deep learning aficionado will not argue those will always arise *ex machina*, unaided by human insight (unless you work for Google, and are only interested in cats^a).

Does this mean every philosophical conundrum and logical puzzle has a direct practical implication? Of course not. But if you are designing a car you do need wheels, so you might as well not reinvent them, especially if yours would end up not quite round.

There are reasons to be optimistic. There are signs researchers are becoming increasingly leery of the “machine learning and stats will solve everything” viewpoint, and are seeking to integrate the (fantastic) achievements of machine learning into a broader AI approach. For example, a recent AAAI symposium¹ brought together leading researchers from knowledge representation, machine learning, linguistics and neuroscience to discuss interactions among these areas. My sense is the pendulum is beginning to swing back ever so slightly, and that if we as a community encourage the trend, AI will be better for it. □

a Here, I am counting on the sense of humor of my Google colleagues.

References

1. AAAI Spring Symposium on KRR: Integrating Symbolic and Neural Approaches, Stanford University (Mar. 2015); <https://sites.google.com/site/krr2015/>
2. Alchourrón, C.E., Gärdenfors, P., and Makinson, D. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic* 50 (1995), 510–530.
3. Bratman, M. *Intentions, Plans, and Practical Reason*. Harvard University Press, 1987.
4. Cohen, P. and Levesque, H. Intention is choice + commitment. *J. Artificial Intelligence* 42 (1990), 213–261.
5. Icard, T., Pacuit, E., and Shoham, Y. Joint revision of belief and intention. In *Joint Proceedings of the 12th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 2010.
6. Shoham, Y. Logics of intention and the database perspective. *J. Philosophical Logic* 38 (2009), 633–647.

Yoav Shoham (shoham@stanford.edu) is Principal Scientist at Google and Professor (emeritus) at Stanford University.

Copyright held by author.