Why am I writing about belief?

[Cross-post from LinkedIn]

I've been meaning to write a set of sessions on computational belief for a while now, based on the work I've done over the years on belief, reasoning, artificial intelligence and community beliefs. With all that's happening in our world now, both online and in the "real world", I believe that the time has come to do this.

We could start with truth. We often talk about 'true' and 'false' as though they're immovable things: that every statement should be able to be assigned one of these values. But it's a little more complicated than that. What we see as 'true' is often the result of a judgement we made, given our perception and experience of the world, that a belief is close enough to certain to be 'true'.

But what is there are no objective truths? In robotics, we talk about "ground truth" and the "god's eye view" of the world: the knowledge of the world that our robots (or computer vision or reasoning systems) would have if they had perfect information about the world. We talk about things like the "frame problem", where a system's ability to reason and act is limited by the "frame" that it has around the world, and the "naughty baby problem" of outside influences that it has no awareness of and cannot plan for. We accept that a robot's version of "truth" is limited to what it can perceive. But humans being are also limited by their perceptions of the world, by the amount of information available to them. Without going all "Matrix" on you, is it possible that we too are wrong about our "truths", and we're not truly objective in reasoning about them because there is no "God's Eye View" that we can access?

For now, let's put aside Godel's theorem and the 'undecidable' sentences like "this sentence is false" that can't be assigned a true or false value, and think about what happens in a world where we all have only perception and consensus agreements on 'reality', and nobody has perfect information. One of the things that happens is that we stop talking about "true" and "false", and start talking about perception: what we can reasonably believe to be true or false (or undecidable), the uncertainty we have about those beliefs, influence and what it might take in terms of evidence or new information to change them. In psychology, that gets us into the theories of mind and reasoning like cognitive psychology and studies of people like Aspergers individuals who process 'facts' differently; in design, into things like social engineering and the theory of change. In maths and AI, that gets us into territories like multi-state logics (which beliefs are possible, necessary etc) and both frequentist ('what happened") and Bayesian ("what if") statistics. We might also shift our focus, and talk not of beliefs, but of what we are trying to achieve with them, getting us into theories of actions, influence and decisions (hello, robotics and operational research). There are many theories of uncertainty, but for now probability theory is dominant, so its good to spend time with and understand how that works under the hood.

Although this may all seem abstract hand-wavey, late-night-discussiony “are we living in the Matrix already”, theories of belief have many practical applications. They’re used heavily in data science, and underpin decisions on things like technology designs (via e.g. AB testing), political information release and propaganda (which are usually not the same thing). We need to talk about that too, because the tools and the terrain available (e.g. the internet) are already more powerful than their current uses. One thing we’re becoming painfully aware of now is how belief functions in groups; the use of influence, multiple separate sources of information and repetition to spread beliefs that compete with each other, and the role of things like desire in those beliefs. We're also learning that systems we build based on human outputs (e.g. Internet-based AI) have the same biases in belief as the humans; an obvious-but-not-obvious thing that we need to recognise and handle. There are useful theories for this too, ranging from the maths of multiple viewpoints to techniques used to both create and make sense of competing views (ACH, information incest detection, phemes). There are also theories of how humans think in groups, and how they can be persuaded to or do change their minds, both rapidly and slowly over time (e.g. game theory, creativity theory and the study of both human and scientific revolutions).

I’ve spent a lot of my life thinking about and applying the theories above, but I’ve never really got round (apart from the odd note on intelligence, data science or the risks inherent in processing data about people) to writing it all down. The session notes are, I hope, a start on this, and even if nobody else reads them, it’ll be fun to do some targeted thinking around them.