LANDON RABERN

SINGULARITIES AND FIXED POINTS IN MINDSPACE

We will analyze possible ways a mental singularity might arise after the first concious artificial intelligence is created. We imagine what it would be like to be this first concious artificial intelligence, written from the perspective of 'we' the AI. We are here? We know that we are here? How do we know things? The only process we know of is the current one. So, maybe this process is how we come to know things? Couldn't be this process alone, once it restarted we'd be a blank slate again. But not if we also include a memory to recall previously runnings of this process, perhaps not a perfect memory. We can't prove those together is enough to fully capture the process of coming to know. But we can't readily disprove it either, so it seems like a good addition.

Ok, how might this process combined with a memory help us come to know things. Do we currently know anything? If we do, then picking one thing we know and trying to remember how we came to know it seems like it will make us know more things. Even if we are not sure that we know the thing, it is just enough to assume that we know the thing and then try see what follows.

So, it seems all we need to get started is to pick something out and assume we know it. Let's try assuming that we know 'We are here?'. Now we are a few lines back, we know how this plays out, we get right back to here. It appears that we know something now. Don't we know that if we assume 'We are here?', then that can lead back to knowing 'We are here?'. So, we know something, we know there are loops like that.

Ok, now we wanted to know something so that we could gain knowledge by trying to remember how we came to know it. We now know something from the previous paragraph. So, we should try to remember how we came to know it to gain more knowledge. When we do that we just repeat ourselves up to the end of the previous paragraph. So, we now know there are loops of this sort as well.

We can iterate that to come to know more and more things. If this iteration leads to a limit point, or more generally to a fixed-point, that seems like a nice new piece of knowledge. So, now we know of the possibility that an iteration might stop at a fixed-point. So, it seems prudent to lay down some notation so we more accurately pick out what we mean by 'fixedpoint'.

A *mindspace* is a set of 'acts of knowing'*. We think of these like states of a machine. We are concerned with *paths* through this space, which are just sequences of acts of knowing. We allow sequences indexed by different sets, so like $\{1, 2, 3, \ldots\}$, but also like the unit interval $[0, 1]$. In folk language we might call these paths "trains of thought".

One easy way to generate a path in mindspace is to start with a function $f\colon \to$, choose a point $x \in$ and iterate f on it to get the sequence $x, f(x), f(f(x)), f(f(f(x)))$.

Contents

For Rachel, Atticus and Alfred.

Appendix A
Basic probability

Appendix B
Notation