

The Physiology of the Noosphere: Segment 6

David Sloan Wilson: Is there an interesting comparison between the autonomic and sympathetic nervous system, something that's relatively automated, as opposed to something that's more reflective? We know that our nervous system is specialized in that way. Can we say something similar about AI in the Noosphere?

Terrence Deacon: Let me start with the neuroscience of it, and that is the autonomic and the parasympathetic, sympathetic nervous system, which run digestion, breathing, heart rate, basic metabolic functions. These are things that are run in background, we're not conscious of them, we feel them when they're out of sync. In the evolution of nervous systems, there are nervous systems that were basically just that. So if you look in particular organisms that are radially symmetric, they mostly don't have heads and tails and eye spots and mouths and so on. And as a result, their nervous system is quite distributed and things are locally maintained. When animals developed, they had a head end in which most of the work was done, in which most of the prediction, because once you start moving around, you have to predict the future.

You have to have organs that can see into the distance, can hear in the distance and so on, or not just immediate. So that oftentimes the nervous system in animals is broken up into this sort of, you might say the predictive nervous system and the maintenance nervous system, they're linked together for obvious reasons. One of the questions right now, I think that largely we're dealing with this linkage of AI into human interactions, as a more autonomic maintaining, runs the stock market these days, and things like that, background stuff, keeping track of traffic and that sort of thing.

The question is, are we more like a starfish with a distributed nervous system, or are we developing a head? A head end is one that predicts the future, and the two have to be the linked together in a complicated way. In the course of evolution, it was the metabolic system that came first and it became more and more sophisticated, and only when it began to have to deal with prediction, did heads develop, did brains develop. My guess is that the state we're in now is still autonomic, but it's moving away from that. And this is the noospheric question, carried out to this analogy to a global brain, I think the analogy is too simple for all of these reasons, even in terms of talking about what a brain is, brains are very different in different species for different reasons.

But one of the questions I'm trying to pursue is, and I think this is hidden in Ben's interests as well, to what extent will we be able to rely on AI to do these autonomic functions, so that we don't have to pay attention to them? Run our countries, run our energy systems, run our fuel systems, our food systems, our distributing systems, and so on, in a way that we don't have to think about it, it's just handed off. And we have to have then this other part of the system, which is the look ahead, the experiment, the trial and the error that we've been talking about, that's what brains are mostly about.

DSW: That's great. But there are forecasting systems, so I think they both exist to a degree. I think that there are the forward looking predictive algorithms. Even before technology, there were forecasting decision making processes and so on, scenario planning and that kind of thing. So it does exist to degree, but I love the distinction, Terry, I love the way you described it because it juxtaposes, especially when we get to Stuart, basically the biological realm and now the human technological realm and how they're so similar to each other. So Stuart, what do you have to say about this?

Stuart Russell: So I think that you could ask all these questions in the complete absence of technology. You could look at the whole human social and economic system and ask, how do we draw analogies to the autonomic nervous system or the sympathetic nervous system. And there are lots of parts of a system that do sort of work like the metabolic system, they just function in the background. When you think about the system of food production and distribution, it's incredibly complex and yet incredibly effective. That somehow we manage to deliver 20 odd billion meals a day to seven odd billion people.

And it does break down a little bit here and there, but overall, despite its huge complexity, it functions pretty well, and a lot of that has to do with the way markets operate. So to a large extent, we have turned over a lot of the management of our human system, to markets, which are in a sense algorithms, and interestingly, we did a little bit of work in the 18th century onwards, to try to reassure ourselves that they would actually function, but they were functioning for thousands of years before that, with no theory, they just emerged and they seem to work well enough to fulfill those autonomic roles.

And I think we haven't done the due diligence for what happens when we allow AI systems, first of all, to function as participants in those markets, and then even to replace the market mechanism altogether with some new kind of decision making or control system.

And I think with a flash crash, we saw that, oh yeah, it doesn't work, when you replace enough of the human actors in the market, with AI actors, you can get these, just like with the magnetic domains, you get this failure because of systems being too similar to each other, or somehow causing each other to operate in sync, in ways that human beings don't, and you've got these massive failures occurring. And if we look at where else are we putting AI systems in charge of these autonomic functions, employment markets are increasingly, both in telling people where to apply, and on the corporate side, deciding who's going to get employed.

That's a huge function that's been working moderately successfully, although different countries have different systems. Some countries actually disallow employment agencies, that's an illegal function in some countries, but not in others. And so there are different designs, but putting AI in charge of that, we've already seen with Amazon that, oh, they didn't even realize that their algorithm was rejecting all female applicants out of hand for technical positions. Well, not all female, but if your résumé mentioned the word women's, like I was a member of a women's choir or a women's rugby team, or I support this organization for women in technology or something, it would just reject your application and they didn't even realize that was going on.

DSW: Was that a feature or a bug perhaps? Was that a feature or a bug? Is that intentional-

SR: No, it was completely unintentional, it was just a result of correlations that the algorithms noticed in the previous decisions that were made by Amazon's hiring managers. I think it was over-emphasized there and ended up being a disaster, both in public relations terms and in the effect on applicants and Amazon's business. So I think again, you see this problem that you move from a large number of heterogeneous agents, to the possibility of these coincidental alignments, where the things start to happen in synchrony and the stability of the system goes out the window. And I think other areas, a lot of eCommerce now, particularly business to business, is becoming more and more automated in terms of selection of vendors and agreement of prices, and so on, is moving towards full automation.

And again, we could start to see large sectors of the economy where it's actually AI algorithms that are deciding how that sector evolves, its structure, the flows of money and goods, and so on. And just like the employment market, that could go wrong in ways that we might regret.

And I think social networks are another thing, to a large extent, AI algorithms are telling people who they should talk to and who they should get to know and who they should mate with, and who knows what effect that could have. We used to have human matchmakers and they had certain defects as well, but we've relied on systems that evolve lots of heterogeneity, lots of serendipity, and so on.

And I use this analogy actually for talking about whether we should edit the genes of our offspring, but I think it applies here too, which is that we're in this casino, we're all playing roulette and then someone just comes along and picks up the ball and puts it in the zero, and the element of chance is now gone and they collect all the money and leave. And is that a good thing? Do you want to get rid of the element of chance and the systems that are built up around it, to actually benefit from all the chance interactions of all these heterogeneous elements. Our society's partly organized around that and to benefit from it,

and if you take that away, it might be the way our society is organized, it no longer gets to benefit from that, and is not well adapted to this new kind of circumstance.