Review of Landgrebe and Smith (2022)

Jonathan Simon

Review of Landgrebe and Smith (2022)

Jonathan Simon

In this formidable book, Jobst Landgrebe and Barry Smith argue that no AI will ever attain human-level intelligence. The book is a challenging read, but it is full of important insights, its master argument is original, it is informed by an impressive array of sources, and it is timely. It merits philosophical attention. The book is also noteworthy because it is the collaboration of an engineer and a philosopher. Landgrebe runs an AI software company: he has an M.D. and a Ph.D. in biochemistry (although notably, his grandfather Ludwig was a famous phenomenologist). Smith is an expert on the Austrian phenomenological tradition and in formal and applied ontology (where, notably, he has pioneered a way to apply philosophical theory and method to data engineering). This kind of collaboration is vital but hard to achieve.

In very broad strokes, the book's master argument is this: Human-level intelligence requires coping with and getting on in environments that are complex dynamical systems—that is, environments that are open and chaotic and subject to feedback effects, with trends and statistics that change over time (think: the weather or the stock market). Data models of such complex dynamical systems are always mere approximations not good enough to enable long-term prediction in a complex world in constant flux (think: why you don't know whether it will rain next Saturday and why you can't reliably beat the stock market). But AI systems are just data models. So, in principle, they can't enable the sort of coping that humans are capable of. What about full-on emulations of the human neuro-cognitive system? That system, too, is a complex dynamical system, so no ensemble of algorithms based on data models can approximate it well enough for full-on emulation (in Landgrebe and Smith's terminology: well enough for a model that is both "adequate" and "synoptic"—where this means a model that enables predictions that are accurate enough for the task at hand).

Here is how they proceed. Chapters 1–6 present their general picture of intelligence and mindedness, as well as language and sociality, and why they think that having these capabilities entails that we can cope *with* and also that our neuro-cognitive systems *are* complex dynamical systems. Highlights from these chapters include fresh insights on the mind-body problem in chapter two, the authors' breakdown of human intelligence into "primal" and "objectifying" intelligence and their critique of reward-optimization conceptions of intelligence in chapter three, and detailed analyses of language and sociality informed by both phenomenology and empirical work in chapters 4–6. Chapters 7–8 deliver the linchpin of the master argument: the claim that AI systems cannot adequately or synoptically model complex dynamical systems. Chapters 9–12 argue that it follows from what they argue in chapters 1–8 that AGI is impossible, that machines will not master human language or sociality, and also that mind uploading is a waste of time, as are attempts to create digital minds to carry on our civilization. Finally, chapter thirteen makes positive recommendations, discussing what Landgrebe and Smith think AI is good for and how they think it should be used.

Let's look closer at the linchpin of the master argument. In section 7.5.2, Landgrebe and Smith enumerate seven key features of complex systems:

> CHANGE AND EVOLUTIONARY CHARACTER (pp. 126–128). Complex systems evolve in various ways: the system's boundaries can shift, new elements come, and old elements go. In many cases, complex systems can undergo changes in the *types* of elements they contain or interactions they participate in.

> ELEMENT-DEPENDENT INTERACTIONS (pp. 128–129). Complex systems typically have different kinds of functionally individuated elements, e.g., the different roles played by proteins, kinases, and ATP in phosphorylation (contrasted with the way that mass and velocity are all you need to chart all of the interactions of a Newtonian system). Elements of a system can also change their functions over time.

> FORCE OVERLAY (pp. 130–131). Complex systems typically involve interactions between all four of the basic physical interactions (EM, gravity, strong and weak).

NON-ERGODIC/COMPLEX PHASE SPACES (pp. 131–132). We cannot predict the trajectory of a complex system over its phase space by averaging over volumes of that phase space.

DRIVENNESS (pp. 132–136). A driven system is a system that does not generally converge to equilibrium because it has access to a reliable energy source.

CONTEXT-DEPENDENCE (p. 137). The interface between a complex system and its environment is constantly changing, e.g., which elements are part of the system vs. part of the environment, or what states the system can occupy.

CHAOS (p. 137–138). Chaotic systems are unpredictable because small differences in initial conditions may lead to large differences down the road.

Say that systems having all seven of these features are *fully complex.* Landgrebe and Smith's master argument is that AGI would only be possible if fully complex systems could be adequately and synoptically modelled (either the ones in the environment or the ones in the brain), but that fully complex systems cannot be adequately and synoptically modelled.

I am not going to say much here about whether they are correct that fully complex systems cannot be adequately and synoptically modelled. It is intractable to find exact solutions to the dynamical equations for most complex systems (even ones that are not *fully* complex, like three-body gravitational problems). Approximation is thus the name of the game. The more chaos a system exhibits, the more its distribution changes over time, etc., the harder it can be to find approximations that are both tractable and accurate enough for the problem at hand. This much is beyond dispute. However, Landgrebe and Smith are arguing for something extremely ambitious: not just that suitable approximations are sometimes or even typically very costly, but that they are, in principle, unavailable for a wide range of cases and will continue to be, even with the increases in computing power that we can expect the future to bring. This is less clear. It is hard not to look at, for example, NASA's recent successes on missions like DART or OSIRIS-REx and come away with the impression that, when there is a will to find suitably accurate approximations, there is a way.

For the remainder, though, I'll focus on the more philosophical questions that arise in Landgrebe and Smith's defence of their claim that AGI would only be possible if fully complex systems could be adequately and synoptically modelled. They pursue two routes to this conclusion. I'll call these the argument from coping and the argument from emulation.

According to the argument from coping, there are fully complex systems in our environments; we cope with them, and AGI is only possible if you can achieve this coping by means of adequate and synoptic modelling.

According to the argument from emulation, our neuro-cognitive systems are fully complex systems, and AGI is only possible if you can emulate them by adequately and synoptically modelling them.

The coping argument is mainly developed in an earlier work, Landgrebe and Smith (2021), but it serves as background for the emulation argument, which is the focus of the present book.

On coping: here, I worry that there is an equivocation. I'll grant unequivocally that there are fully complex systems in our environments, like weather or the stock market, but I'm not sure what it means to allow that we cope with them. Do individuals really cope with hurricanes or stock market crashes? Arguably, imperfect though they are, computational models of hurricanes are the best tools we have for coping with hurricanes. Our coping abilities turn on bounded, often flawed approximations of the chaotic world around us. If those models are enough for coping, then clearly calling for adequate and synoptic modelling sets the bar too high (as a necessary condition for an AI system to count as coping). On the other hand, if these models aren't enough for coping, then, presumably, we can't cope. Either way, the argument from coping fails: either we cannot cope, or the computational methods we use to cope (which fall short of adequate and synoptic modelling) suffice for coping.

On emulation: here, I have a few worries. First, it isn't obvious that our neuro-cognitive systems are fully complex. For example, it is debatable how much chaos there is in the healthy brain, as opposed to criticality or near-criticality (see O'Byrne and Jerbi 2022).

Second, there is an equivocation lurking in the notion of 'emulation' at issue. Is the aim of emulation to create a perfect replica of a specific token system, e.g., to build a concrete model of a specific hurricane, accurate enough to predict where and when that particular hurricane will make landfall? Or is the aim simply to generate a new sample from the same distribution, a new token of the relevant type? If we are after our own digital immortality, then

maybe we must pursue the former project. In contrast, if all we want to do is build an AGI, we only need pursue the latter project.

But I worry that from the claim that the human neuro-cognitive system is fully complex, taken together with the claim that it is impossible to adequately and synoptically model fully complex systems, we only get the impossibility of *token-level* emulation, leaving open the possibility of type-level emulation.

Maybe we cannot build a model of an actual specific hurricane currently out at sea that will allow us to predict to the minute or square mile when and where it makes landfall. But we can build models of hurricanes that embody the profile of hurricanes in general (see Weisberg 2013 for a discussion of distinctions between kinds of predictive models). So too here: the AGI we build might not be a perfect copy of you or me, but this does not preclude that it is adequate as a type-level emulation—especially so since the type in question is *system that has human-level intelligence* and not *system that is as complex as humans*.

Of course, the token-level question is important, too; it seems relevant to questions of uploading and digital immortality. But there is no simple refutation of the possibility of uploading or digital immortality here since numerical identity over time does not require qualitative identity over time.

For the type-level question, of course, it remains to be shown that such an emulation is possible. Maybe adequate and synoptic modelling of fully complex systems would still be required, even if the thing we create is not constrained to perfectly resemble an existing intelligent being.

This is one way that the coping argument fits into the dialectic of the book: if human-level coping involves harnessing our full complexity in order to ride out storms, and we can do this as well as we would if we had adequate and synoptic models of those storms, then emulating our type *in silico* presumably entails adequate and synoptic modelling of fully complex systems somewhere or other. But if, as I suggest above, we don't cope as well as that, then it does not follow that emulating us entails adequate and synoptic modelling of fully complex systems somewhere or other.

Landgrebe and Smith might advocate a slightly weaker fallback claim, which is that the full (or nearly-full) complexity of our neuro-cognitive systems surely must have something to do with all of our intellectual successes, and so nothing will succeed in emulating us (at the type level) if it cannot at the very least *instantiate* the seven key features that make a system fully complex. This isn't obvious: again, the type to be emulated is not *system that is as complex*

as humans but rather *system that has human-level intelligence*. Still, it merits consideration.

But there is no clear reason to doubt that we can build digital systems that instantiate these features. Think of Conway's Game of Life or other complex systems generated *in silico* via cellular automaton rules. If we assess criteria like drivenness and force-overlay within the simulation, arguably, these are fully complex. So, too, for deep learning systems, especially if we focus on their dynamics during training (as opposed to inference). During training, parameters evolve (and the sampled distribution changes). Stochastic gradient descent is non-ergodic: systems get stuck in local minima all the time. It can also be chaotic or near-chaotic: there are trajectories that pass along the borders of basins of attraction for local minima. Even during inference, some complex features can be seen. For example, the whole point of attentional mechanisms is to allow models to handle context during inference (see Søgaard 2022), and functional differentiation between different neural network layers (attention vs. feed-forward, pooling vs. convolution, etc.) exemplifies element-dependent interactions. Finally, let's not forget about neural organoids, which are programmable assemblies of biological neural cells: these certainly fit the bill if nothing *in silico* does.

Thus, we have a few reasons to doubt that Landgrebe and Smith fully succeed. Even so, their arguments are important and merit further consideration. If our uploads are guaranteed to differ from us, this problematizes the claim that we can survive into them or that they preserve us, even if the matter is far from settled. And I certainly agree with Landgrebe and Smith about the limits of *current* AI systems and that the question of how and to what degree we adaptively harness our underlying complexity is a key open question, one which we must answer to fully understand the difference between biological minds and AI systems. That said, I am still afraid.*

Jonathan Simon
0000-0001-6081-4127
Université de Montréal
jonathan.simon@umontreal.ca

References

LANDGREBE, Jobst and SMITH, Barry. 2021. "An Argument for the Impossibility of Machine Intelligence." ArXiv preprint, doi:10.48550/arXiv.2111.07765v1.

—. 2022. *Why Machines Will Never Rule the World. Artificial Intelligence Without Fear.* London: Routledge, doi:10.1002/9780470693476.

O'BYRNE, Jordan and JERBI, Karim. 2022. "How Critical is Brain Criticality?" *Trends in Neurosciences* 45(11): 820–837, doi:10.1016/j.tins.2022.08.007.

SØGAARD, Anders. 2022. "Understanding Models Understanding Language." *Synthese* 200(6), doi:10.1007/s11229-022-03931-4.

WEISBERG, Michael. 2013. *Simulation and Similarity. Using Models to Understand the World.* New York: Oxford University Press, doi:10.1093/acprof:oso/9780199933662.001.0001.