FORMAL PHILOSOPHY 2021

Dilectiss Liu[1]

WHAT MACHINE PHILOSOPHY ISN'T

Abstract. This short paper clears up three misunderstandings about machine philosophy. First, machine philosophy does not demand computational or formal philosophy. Instead, it only calls for a *grounding* of philosophical theorising in statistical learning, not that philosophy must proceed by means of statistical learning. Second, machine philosophy does not entail the collapse of metaphysics. In fact, it doesn't affect metaphysics more than any other philosophical field. Third, while machine philosophy potentially entails the illegitimacy of widely discussed philosophical problems, this consideration isn't a worry for machine philosophy.
Keywords: machine; philosophy; metaphilosophy; methodology; intuitions; theorising.

Для цитирования: *Liu D.* What Machine Philosophy Isn't // Логико-философские штудии. 2021. Т. 19, № 2. С. 117–122. DOI: 10.52119/LPHS.2021.82.98.009.

1. What is Machine Philosophy

Machine philosophy is a methodological foundation for doing philosophy. It proposes that philosophical theorising be epistemically grounded on statistical principles, and thereby continuous with the physical sciences. For this end, machine philosophy entails a stance on philosophical data, and the activity of theorising over this data. Agreeing with standard practice, machine philosophy treats intuitions as the primary source of philosophical evidence. However, instead of taking them as either *a priori* evidence about physical reality, or as outright pseudo-evidence, machine philosophy treats intuitions as fallible evidence, which reflect *objective* facts about our *socio-linguistic realities.* As for theorising, machine philosophy rejects standard boolean attitudes, which it sees as the key culprit for the methodological anarchy in philosophy. In particular, boolean argumentations such as using *reductio ad absurdum* have led to descriptivists to adopt an overly cautious take against statistical bias in theorising, which has in turn led to overfitting.[2] The standard response to this is a boolean one—to treat philoso-

[1] *Dilectiss Liu*, Research Associate, University of Tübingen.
dilectiss.liu@uni-tuebingen.de

[2] Strawson (1963) raised the original point against revision in philosophy. Post-Gettier epistemology is a manifestation of this attitude.

phy as normative.[3] Machine philosophy rejects this false dichotomy, by replacing the boolean attitude with a fundamentally statistical one. In this capacity, philosophical theories are *descriptive* of our socio-linguistic realities in virtue of being *statistically adequate models of our intuitions*. In this capacity, philosophical theorising should *revise* over our intuitions, with the goal of producing *true* theories.

2. Relating to Computational or Formal Philosophy

First, machine philosophy doesn't entail computational philosophy. Computational philosophy is the methodology of applying computational methods to philosophical problems. Examples include using agent based modelling (ABMs) to study moral norms or social epistemological problems.[4] These have indeed worked quite well. However, machine philosophy simply calls for philosophical theorising to be governed by statistical norms, and provides a foundation for doing so. As a result, philosophical problems *could lend themselves* to computational solutions. However, machine philosophy does not entail that philosophers *should* use computational methods. After all, scientific theorising as we know it had been practised even before computational methods became feasible.

Second, machine philosophy does not entail formal philosophy. This is analogous to machine philosophy's relation to computational philosophy. Machine philosophy simply claims that philosophical methodology should shift from a priori, boolean reasoning to statistical methods like in the sciences. This does not itself entail that philosophers should use only formal methods. Again, scientists doesn't exclusively engage in statistical modelling, even if their *reasoning* are ultimately grounded on statistical norms. Similarly, mathematicians don't generally do first-order logic, even if the foundations of mathematics grounds mathematical propositions in set-theoretic terms. However, machine philosophy does entail that philosophers *should* strive to be as statistically fine-grained as feasible. This means that, whenever a more refined statistical method

[3]This stance can be traced back to *explication* in Carnap (1950), and the defence of it against Strawson in Carnap (1963). For a contemporary summary of the revisionist attitude, see Cappelen (2018). Also in Cappelen, it has been noted that there isn't a clear line between so-called descriptivist (Strawsonians) and revisionists. Jackson (1998) is a prime example of this fuzziness. Machine philosophy is in a sense, a clarification of the Jacksonian sentiment and the Carnapian project.

[4]Goldman (2010) had laid the groundwork for using ABMs to study social-epistemological problems by setting out a system for social-epistemological settings with the parameters: choices; agent; evidence; norms. Mayo-Wilson (2014) uses ABMs of research networks to study the reliability of testimonial norms in scientific communities. Zollman (2010) uses network structures in ABMs to study the social epistemological problem of communication. These studies have been done in NetLogo, a high level language specifically designed for building ABMs. See also Weisberg, Muldoon (2009); Mayo-Wilson, Zollman, Danks (2011). In ethics, Skyrms (2003, 2010) laid the groundwork by building a game-theoretic framework to account for cooperation. We have examples such as Alexander (2007) and Muldoon, Hartmann (2014), who used ABMs to study how moral norms can evolve out of basic game-theoretic norms.

is available, the philosopher shouldn't settle on less refined methods. If it's possible for a particular philosophical question to be answered by a machine learning model, there is no reason for us to stop at the archaic procedure of linguistic or conceptual analysis. This does not meant that we cannot *start* with simple analysis, just that we should *also* use the more advanced methods *in addition*, in order to refine our theorising. Moreover, there is absolutely no reason to assume that we need a lot of data to do machine philosophy well—the intuitions that we have considered in the literature are perfectly fine.[5] This is neither a pleasantry nor an annoyance, it's merely what scientists have been doing for quite a while.

3. On the Issue of Metaphysics

Metaphysicians often think of their theories as being about the mind-independent world, and intuitions are *a priori* evidence of that. Machine philosophy considers intuitions to be empirical facts about our socio-linguistic reality. So it seems that metaphysics has no place in machine philosophy. However, this would be a misunderstanding on several levels.

First, it would be an equivocation on 'about' as an epistemic relation of grounding with 'about' as a linguistic relation of reference. Intuitions are epistemically grounded in our socio-linguistic realities. However, they refer to the external world. Intuitions are ultimately about *our categorisations* of entities *in the world*. Observations are likewise not as 'objective' as one might like to believe.[6] Pointing out that intuitions are community-dependent (as observations are) is a call for modesty, not a rejection of the ability for philosophers to study the mind-independent world.

Second, whether our intuitions refer to our socio-linguistic realities or the physical reality has no bearing on the *truth* of metaphysical propositions. If metaphysicians

[5]Moreover, we have a ML technique 'Less Than One'-Shot Learning from Sucholutsky, Schonlau (2020). This technique is much more data efficient than previously thought possible for statistical learning. It allows statistical learning to effectively generalise on what we might think of as anecdotal data. This is good news for philosophers, as we'd rather sit in our armchair than do corpus studies.

[6]Proietti et. al (2019) has experimentally realised Wigner's thought experiment from Wigner (1961) where two observers experience irreconcilable realities at the quantum level. Granted, observer-dependent evidence need not entail observer-dependent quantum theories, since either locality or statistical independence could be violated by quantum mechanics, as demonstrated in Hensen et. al (2015). However, the point stands that we cannot in general assume observer-independency for observational evidence. Moreover, the point of the experiment wasn't that observations are observer-dependent, this is trivial (via being theory-laden). The point was that observer-dependency could be a feature of the measured reality rather than human quirks (including our beliefs etc.). This reinforces the QBist reading of quantum mechanics (and Proietti et al. support this view) where the very target of a quantum theory *just is* the reality *as experienced by an observer*. However, it does not entail anti-realism. In fact, the experiment demonstrates the opposite: the reality that exists as we know it may be necessarily observer dependent as a matter of physical reality. Since intuitions are trivially observer-dependent in virtue of having no obvious external reference, we need not delve further into this point.

insist that their subject matter is the mind-independent world, they're free to do so—so long as they do not treat their metaphysical intuitions as infallible boolean evidence, and so long as they try their best to theorise in accord with statistical principles.

Third, what matters for philosophical theorising (not just metaphysics) isn't about the 'mind-dependencies' of our subject matters, but their objectivity. Intuitions, like observations, are necessarily observer-dependent and theory-laden. There is a trivial sense in which all of our data are 'mind-dependent'. However, that has no bearing on whether the evidence is objective. The relevant sense of objectivity for theories is the existence of objective criteria for measuring the truth of a theory. For intuitions, this criteria is a combination of expert agreement and statistical significance, much like for observations. What matters is that our theories aren't idiosyncratic, not whether the truth of a theory is domain-bound.

4. On the Issue of Legitimate Philosophical Problems

One consequence of machine philosophy is that certain traditional philosophical issues that are grounded on boolean considerations might become illegitimate. For example, Cartesian considerations of epistemic certainty becomes inoculated once we no longer think of justification as boolean. However, there seems to be plenty of philosophers who take many traditional philosophical issues very seriously. One might argue that this latter fact constitutes a counterexample to machine philosophy. This objection has four fatal flaws.

First, there's the division within philosophy. Already, many philosophical problems have been abandoned by philosophers,[7] and it isn't a worry for these people that some traditional philosophical problems aren't solvable or well-formed. Just because a lot of people cared about phlogiston or aether doesn't make them genuine scientific topics.

Second, machine philosophy isn't supposed to describe what philosophers actually do. So the fact that a lot of people actually tackle certain problems isn't the relevant data. Machine philosophy tries to describe what *proper* philosophy looks like.

Third, machine philosophy isn't as radical as it might first appear. So long a problem is solvable by scientific methods, it's a real philosophical problem. I suspect that many of the active problems are solvable, just that we need to abandon boolean ways of tackling them. Of course, I don't know exactly how many problems that philosophers work on are rendered nonsensical by machine philosophy and I'm aware that my proposal may look quite radical. Nonetheless, I see it as a natural and necessary step in the progress of philosophical methodology. If some folks have a tough time accepting that, it's not my concern to convince them, just as is my indifference to the strong voices of flat-earthers.

[7]E.g. Bayesian, machine, or function-first epistemology abandoning the Gettier issues. See Bovens, Hartmann (2003), Wheeler (2016), Hannon (2019).

Finally, even if the structure of this *modus tollens* argument is legitimate, machine philosophy begins with the goal of producing *true* propositions, and thereby epistemic rigour in the service of truth, rather than being held hostage by some ungrounded theoretical concerns like the Strawsonian attitude, into which both the conceptual engineers and analysts have inadvertently bought.

References

Alexander 2007 — *Alexander J.* The Structural Evolution of Morality. Cambridge University Press, 2017.

Bowens, Hartmann 2003 — *Bowens L., Hartmann S.* Bayesian Epistemology. Oxford: Oxford University Press, 2003.

Cappelen 2018 — *Cappelen H.* Fixing Language: An Essay on Conceptual Engineering. Oxford: Oxford University Press, 2018.

Carnap 1950 — *Carnap R.* On Explication // Logical Foundations of Probability. University of Chicago Press, 1950.

Carnap 1963 — *Carnap R.* P. F. Strawson on Lingustic Naturalism // The philosophy of Rudolf Carnap / ed. by P. A. Schilpp. La Salle: Open Court, 1963. P. 933–940.

Goldman 2010 — *Goldman A.* Systems-oriented social epistemology // Oxford Studies of Epistemology. 2010. Vol. 3. P. 189–214.

Hannon 2019 — *Hannon M.* What's the Point of Knowledge? Oxford University Press: 2019.

Hensen et al. 2015 — *Hensen B., Bernien H., Dréau A et al.* Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres // Nature. 2015. Vol. 526. P. 682–686.

Jackson 1998 — *Jackson F.* From Metaphysics to Ethics: A Defence of Conceptual Analysis. Oxford University Press, 1998.

Mayo-Wilson 2014 — *Mayo-Wilson C.* Reliability of Testimonial Norms in Scientific Communities // Synthese. 2014. Vol. 191. P. 55–78.

Mayo-Wilson, Zollman, Danks 2011 — *Mayo-Wilson C., Zollman K., Danks D.* The Independence Thesis: When Individual and Social Epistemology Diverge // Philosophy of Science. 2011. Vol. 78, no. 4. P. 653–677.

Muldoon, Lisciandra, Hartmann 2014 — *Muldoon R., Lisciandra C., Hartmann S.* Why are there descriptive norms? Because we looked for them // Synthese. 2014. Vol. 191. P. 4409–4429.

Proietti et al. 2019 — *Proietti M., Pickston A., Graffitti F., Barrow P., Kundys D., Branciard C., ..., Fedrizzi A.* Experimental test of local observer independence // Science Advances. 2019. Vol. 5, no. 9. DOI: 10.1126/sciadv.aaw9832.

Skyrms 2003 — *Skyrms B.* The Stag Hunt and the Evolution of Social Structure. Cambridge: Cambridge University Press, 2003.

Skyrms 2010 — *Skyrms B.* Signals: Evolution, learning, and information. Oxford University Press, 2010.

Strawson 1963 — *Strawson P.* Carnap's Views on Conceptual Systems versus Natural Languages in Analytic Philosophy // The philosophy of Rudolf Carnap / ed. by P. A. Schilpp. La Salle: Open Court, 1963. P. 503–518.

Sucholutsky, Schonlau 2020 — *Sucholutsky I., Schonlau M.* 'Less than one'-shot learning: Learning N classes from $M < N$ samples. 2020. arXiv: 2009.08449.

Weisberg, Muldoon 2009 — *Weisberg M., Muldoon R.* Epistemic Landscapes and the Division of Cognitive Labor // Philosophy of Science. 2009. Vol. 76, no. 2. P. 225–252.

Wigner 1961 — *Wigner E.* Remarks on the Mind-body Question // The Scientist Speculates / ed. by I. J. Good. London: Heinemann, 1961. P. 284–302.

Wheeler 2016 — *Wheeler G.* Machine Epistemology and Big Data // The Routledge Companion to Philosophy of Social Science / ed. by L. McIntyre and A. Rosenberg. Routledge, 2016. P. 341–349.

Zollman 2010 — *Zollman K.* The Communication Structure of Epistemic Communities // Social Epistemology: Essential Readings / ed. by A. Goldman and D. Whitcomb. Oxford University Press, 2010. P. 338–350.