Natralization without associationist reduction: a brief rebuttal to Yoshimi

Jesse Lopes[1]

Abstract

Yoshimi has attempted to defuse my argument concerning the identification of network abstraction with empiricist abstraction - thus entailing psychologism - by claiming that the argument does not generalize from the example of simple feed-forward networks. I show that such details of networks are logically irrelevant to the nature of the abstractive process they employ. This is ultimately because deep artificial neural networks (ANNs) and dynamical systems theory applied to the mind (DST) are both associationisms - that is, empiricist theories that derive the principles of thought from the causal history of the organism/system. On this basis, I put forward a new aspect of the old argument by noting that ANNs & DST are the causal bases of the phenomena of passive synthesis, whereas the language of thought hypothesis (LOT) and the symbolic computational theory of mind (CTM) are the causal bases of the phenomena of active synthesis. If the phenomena of active synthesis are not distinct in kind from and are thus reducible to those of passive synthesis, psychologism results. Yoshimi's program, insofar as it denies this fundamental phenomenological distinction, is revealed to be the true anti-pluralist program, by essentially denying the causal efficacy of the mechanistic foundations of active synthesis by referring phenomenology exclusively to associationism for its causal foundation.

Yoshimi has written an excellent defense and exposition of his views. It's correct that I have tried to "assimilate" artificial neural networks (ANNs) and dynamical systems theory (DST) as applied to the mind/brain to empiricism and its abstractive process. I

✉ Jesse Lopes
    redbroom@bu.edu

1   Boston College, Boston, MA, USA

🖄 Springer

should say that I am following a venerable tradition in doing so. These approaches are regularly distinguished from their rationalist alternative in the computational theory of mind (CTM), best exemplified by Fodor's work on the language of thought (LOT) (Cain, 2016 34–35). Their distinguishing feature, with respect to content (and thus to intentionality), is their sub-symbolic or non-symbolic character. This distinguishing feature is neither in dispute (contra Yoshimi (?)) nor neutral (contra Yoshimi (!)) – it is the major point at issue with respect to psychologism and, as I argue, Husserlian causal mechanisms. For if ANNs & DST lack (innate) discrete symbols, they lack content that is not constituted by experiential input. If they lack content that is not constituted by experiential input, then they must abstract it from experiential input, if they are to have content (intentionality) at all. If so, they will necessarily employ the empiricist theory of abstraction, which is the Lockean theory that, since the Cartesian doctrine of innate content is false, content must be abstracted from experience. For either a symbol has its content intrinsically, which is the CTM view, or it has its content extrinsically, which is the empiricist view. By the disjunctive syllogism, and insofar as this is an exhaustive/logical dilemma, ANNs & DST will necessarily get their content (if they can be said to have any) extrinsically. They are thus empiricisms necessarily employing the empiricist theory of abstraction: the abstraction of content based on similarities and non-similarities in sensory experience (in the input).

Yoshimi seems not to like the 'necessity' of the above arguments. As a result, instead of calling my arguments forceful, he has preferred to call them "forcing" arguments. Now, all interesting arguments are "forcing" (i.e. compelling) arguments, insofar as they are true. But I suppose Yoshimi wishes to imply that they are a bit strained, since one of Yoshimi's two goals in his criticism is to say that the arguments are obviously false, since even I regard the approaches as philosophically neutral, as demonstrated by my endorsement of them. Yoshimi is here defending a neutrality thesis which, given the above, is implausible. But it is not a paradox to endorse ANNs and DST, provided they are supplemented by the syntactic symbols of CTM, or what Husserl calls "the natural psychological mechanism of symbolic inference" (1994 42). For in that case, to have content (or intentionality), they will not need to employ the empiricist theory of abstraction, which would otherwise result in psychologism for the system/organism. Such a symbolic mechanism is therefore the causal explanation for the possibility of content identity (and thus logical inference, as both Husserl & Fodor argue), which would otherwise have to be abstracted from similarities in experience. But between similarity and identity in content constitution, there is an infinite regress as well as circularity – according to both Husserl (1901/2001 241–244) and Fodor & Lepore (1992, 2002). Consequently any system constituting content based on its causal history will never achieve content identity (meaning) but can at best approximate it. This is consistent with what Marcus & Davis say concerning Deep Learning, that its similarity amalgams may "approximate meaning" but cannot capture "the real thing" (2019 132). All Yoshimi really says in response to this possibility is that it would "undermine most uses of neural networks or dynamical systems in cognitive science (e.g. any use of the concept of an attractor)." But this is false: see, for example, Fodor's endorsement of attractor landscapes in relation to LOT (2008 159–163). My argument is that DST & LOT are compatible but necessarily distinct in kind, not in degree (and therefore non-reducible), on pain of

psychologism. I develop this point further below by identifying DST & ANNs as the causal bases for the domain of passive synthesis and LOT as the causal basis for the domain of active synthesis.

Yoshimi has another goal in his paper, namely, to minimize the generality of the argument by saying that it is limited to simple feed-forward networks, which I used, following Churchland (2012), as an illustrative example of the type of abstraction that awaits all artificial neural networks. Here Yoshimi shows that he has not appreciated the generality of the argument. The cause of Yoshimi's *ignoratio elenchi* here lies in his failure to realize that it is the invariable detail of their being data-driven machines - and not the variable details of supervised/unsupervised, simple/complex & feed-forward/recurrent - that makes neural networks employ the empiricist theory of abstraction, whenever abstract contents are to be constituted in these machines. Buckner (2018) has independently shown that deep convolutional neural networks (CNNs) employ the empiricist theory of abstraction. As deep CNNs are complex, and their weight adjustments can be unsupervised, it follows, by Yoshimi's criteria, that the details of supervised/unsupervised and simple/complex may be ruled out as logically irrelevant, since Yoshimi does not contest that simple feed-forward networks employ the empiricist theory of abstraction. That leaves feed-forward/recurrent. The only difference between feed-forward and recurrent is that a recurrent neural network (RNN) adds loops between input and output that create context based on its causal history. I confess I fail to see how such recurrent loops could cause a network to avoid statistically separating similarities and differences and generalizing therefrom. In fact, the context that is built by loops allows the network to be more sensitive to the similarities and differences in the input, which for RNNs is usually linguistic (as opposed to imagistic) input. With the advent of deep learning, RNNs developed most prominently into long short-term memory networks (LSTMs), and their abstraction of similarities and differences became clearer, as with deep CNNs above. For example, the word2vec LSTM models exhibit the idea that "similar contexts have similar meanings" (Kelleher, 2019 181). These networks abstract by recurrent processing, as opposed to feed-forward processing, to generate context on the basis of similarities and differences in the input; but, again, that just is the empiricist theory of abstraction. Wherever one looks one will find that networks abstract on the basis of similarity. That is not because of the variable and, as I argue, quite orthogonal details of supervised/unsupervised, simple/complex, & feed-forward/recurrent, but because of the invariable detail that these are data-driven machines, which are therefore dependent for content constitution on their ever-fluctuating causal histories.

Since Yoshimi was not convinced by my previous line of argument, I hope it will not be out of line to propose another "forcing" argument, this time on the basis of the idea that even more generally ANNs & DST are associationisms; and this is to be distinguished in principle from syntactic mechanisms on pain of psychologism. For this, I need a definition: "Associationism is a theory that connects learning to thought based on principles of the organism's causal history" (Mandelbaum, 2015). The law of similarity, as Husserl observes within the epoché, is the basic law of associationism, determining "how a similarity among a variety of similarities becomes privileged to build a bridge, and how each present can ultimately enter into a relation with all pasts" (1926/2001 169). All neural networks (and dynamical systems

approaches to the mind) are associationisms, since they are causal history approaches to the mind, and constitute content, if they do, on the basis of their causal history, governed necessarily by the law of similarity in appearance. As such, they will necessarily employ the empiricist theory of abstraction, if they abstract at all. That's again because when the abstraction of content takes place on the basis of *causal history* - a data-driven process (and this applies to all current machine learning) - that abstraction will proceed on the basis of associated similarities (abstracting from non-similarities or differences). This general argument is *illustrated* by, not essentially *derived* from, Churchland's (2012) example of simple feed-forward networks, which is just a simple example for purposes of clarity of the general nature of network abstraction, for both me and Churchland. Yoshimi simply fails to observe the argument, which is, I emphasize, supported by "the phenomenology of association," revealing "the relation of similarity" as governing all passive synthesis, i.e., data-driven synthesis (1926/2001 163, 168).

I hope therefore it is not too strong to say that Yoshimi is simply wrong: the details of these network associationisms do not matter (much). What matters is the fact that these machines are primarily data-driven associationisms. And what the phenomenology of associationism (passive synthesis) reveals is that they are all essentially governed by the relation of similarity. As a direct result of this law (revealed within the epoché), if they abstract at all, artificial neural networks, unless supplemented by the mechanisms/phenomenology of syntax, will necessarily employ the empiricist theory of abstraction. For what the phenomenology reveals - epistemologically for cognitive science now - is that there is a distinction in kind, not in degree, between the phenomenology of association and the phenomenology of syntax. It is, I claim, syntax and syntax *only* that allows a system/machine/organism to avoid the empiricist theory of abstraction and the psychologism that necessarily results therefrom. For it is only here that logical form, and thus symbols with discrete identical contents, are possible, along with meanings being a function of identical discrete units and the way they're put together (i.e. active synthesis). I shall return to this presently.

One small point, first, about my supposed anti-pluralism. If one reads Part III of "On the psychologism of neurophenomenology" one will find me trying to recommend Yoshimi's approach to low-level phenomenology (Lopes, 2021). This is based on the Husserlian demarcation, just mentioned, between the phenomenology of association and the phenomenology of categorial-syntactic intuition. My only point in this section is that it cannot serve as the *immediate* or *direct* explanatory foundation for higher-level phenomenology - the genuine (content identity) intentionality toward objects, states of affairs, etc. and their functionally related combinatorial syntheses - productivity, systematicity etc.; in other words: active synthesis. My main point, to be entirely clear, is that the similarity-based associationisms that Yoshimi favors *as potentially sufficient* for phenomenology are necessarily insufficient - not that they aren't *necessary*. I concede entirely that they may be necessary – Fodor even regarded DST as necessary for statistical prototype effects, short of grasping, however, Platonic essences (2008 159–163). I think deep ANNs might give some insight into the same function. That does not mean they are sufficient. This point is very Fodorian but also entirely Husserlian. It is, as I say, based on the phenomenology of passive and active synthesis - active synthesis usually being neglected in these types

of discussion. Active synthesis is *indirectly* (not directly) related to passive synthesis on pain of psychologism. For if they were directly related (i.e. continuous), the one would be reducible to the other. That would be exactly what Hume wanted, and logico-syntactic phenomena would be reduced to the causality of associations. But that is psychologism. Someone once said that all experience may begin from sense-experience but it does not therefore arise from sense-experience. The experience that begins from sense-experience but does not therefore arise from sense-experience is categorial-syntactic experience, projected, I argue, from the Husserlian mechanism of syntactically structured, symbolic mental representations – a language of thought. This mechanism is distinct in kind, not in degree, from mechanisms of association, according to Husserl himself (1994 36–46). He uses it to explain how our minds can apprehend both arithmetic and language (46), just as Chomsky uses the syntactic operation of Merge to do the same (2012 15; Cain 2021 130–131).

As a result, we can have our cake and eat it too: we can have a "natural psychological mechanism of symbolic inference" that causally explains the forms of active synthesis without psychologism. The price to pay is that we must distinguish this mechanism from any and all mechanisms of association (ANNs & DST) that are forced, in their ambivalence toward discrete symbols with intrinsic content, to abstract content on the basis of their causal histories, governed necessarily by the law of similarity. Thus we have the following picture:

Phenomenologically (within the methodological, not substantive, epoché):
Associationism ◊ Passive Synthesis (Law of Similarity)
Syntacticism ◊ Active Synthesis (Law of Identity)
Mechanistically (substantively, outside the epoché):
Associationism ◊ Causality (DST & ANNs)
Computationalism ◊ Causality + Syntax (CTM & LOT)

It's that last parameter, which is to be construed as not essentially determined by causality, that prevents active synthesis from collapsing back into passive synthesis. Collapsing back into passive synthesis would mean psychologism for our cognitive life. Accordingly, Husserl argues "the cognitive life, the life of *logos*" must be split into two parts, marking a distinction in kind, or "fundamental stratification": "(1) Passivity and receptivity" (1926/2001 105). This is the level of empiricist associationism, discussed descriptively, and this is also where dynamical systems theory and artificial neural networks can play an explanatory role in our theory of the cognitive life. The second level: "(2) That spontaneous activity of the ego (the activity of *intellectus agens*)" which consists of "syntactical actions" that are not associative, i.e. not determined by purely data-driven processes (ibid.; 1929/1969 106). The former is *directly* related to hyletic data (phenomenal features in state-space) vs. indirectly related (hence "spontaneous") to syntactically structured thoughts and meanings as in the language of thought hypothesis. The spontaneity of this indirectness refers to syntax and its determinations as independent parameters, not passively received from experience, or sensuously inputted into the causal history of the associative network. It is precisely because syntax in the computational theory of mind *is* such an independent formal parameter that it can serve as the rule-bound bulwark, as it were, against

experientially arbitrary associative syntheses, and thus against psychologism, which, for example, reduces transitive inference to a certain relative frequency of associations in experience - a clear epistemological absurdity, as Husserl points out in both *Ideas I* (1913/2014) and the highly mechanistic and computational "On the Logic of Signs (Semiotic)" (1891/1994). If we are to avoid psychologism, therefore, we must hold on to that independent parameter with all our theoretical strength, as Jerry Fodor tried to argue in his own way for the last 50 years. This means always phenomenologically going beyond passive synthesis and, with this descriptive evidence in hand, explanatorily and mechanistically going beyond the sub-symbolic nature of associative neural networks or the non-symbolic nature of dynamical systems theory. Yoshimi's (2009) formerly sharp separation of these two levels is very compatible with my view here. But now Yoshimi throws doubt on this sharpness, not apprehending the psychologistic consequence (2023).

My theory does not mean that associationism's nature and results cannot be investigated phenomenologically. Anything that plays into the cognitive life has some sort of essential way of working and can be investigated phenomenologically; and associationism is no exception. That is, after all, what a large portion of the *Analyses Concerning Passive & Active Synthesis* is about, as Yoshimi emphasizes. The suggestion by Yoshimi that I deny this is a red herring. It distracts from the issue that I'm raising: what kind of mind can apprehend universals/essences/kinds as distinct from instances/particulars vs. only inductive amalgamations of instances/particulars. The phenomenological investigation of the essential nature of associationism, as of anything else that represents a distinct kind of cognitive accomplishment, transcendentally presupposes principles that are not associative in character. Consequently, when Yoshimi attempts to base phenomenology on passive synthesis on the Humean ground that there are laws and types of passive synthesis also revealed phenomenologically, he gets the cart before the horse. As Husserl says, "It is endemic to the nature of the situation that we can only speak of these lower levels [of passivity] if we already have before us something constituted in activity" (1926/2001 275). This of course means that one cannot speak of identity conditions among associative networks without presupposing the theory of content that network graphs were supposed to explain, and which is in fact the fruit of the symbolic mechanism. This is why Yoshimi reduces my talk about (multiply instantiable) contents and thoughts to talk about when theorists think the machine is in the same state, since that is understood. But that is of no use "when the project is to explain what identity/difference of content is, because the notion of identity of labels just *is* the notion of identity of content, and, as far as we know, connectionists/associationists have no theory of conceptual content" (Fodor & Pylyshyn 2015 51). They have no theory of universal conceptual contents and identities because phenomena of meaning cannot structurally be generated by associative networks. Universal conceptual contents and identities are a function of categorial-syntactic intuitions – which I trace causally to the symbolic mechanism - and are not a function of passive synthesis, on pain of psychologism (see the 2nd *Logical Investigation*).

It is the central mystery of transcendental phenomenology how the intuition of universals and the identities of logic and scientific theory could possibly emerge on a Lockean & Humean (associationist) foundation. Indeed, Husserl calls it "impos-

sible" (1926/2001 32). The real question I am raising, the one Yoshimi is ignoring, is: if our theory of the cognitive mind derives *all* of its principles from the rules of passive synthesis - and therefore the causal history of the organism - would we have a psychologism? The answer is clearly **yes**. And what I claim *therefore* is: insofar as Yoshimi immediately founds phenomenology on explanatory frameworks that can only derive the principles of thought from the causal histories of the organism (the definition of associationism), then psychologism results. I merely observe that ANNs and DST, insofar as they are distinguished from symbolic theories, are frameworks that derive their principles from the history of the organism/system. They are thus frameworks that broadly fall under the general idea of associationism, governed by similarity, and therefore employing the empiricist theory of abstraction if they abstract. I don't see how this can be denied (call it a "forcing" argument if you will). As a result, if we are to have not only a theory of passive synthesis, but of active synthesis - respecting the distinction as a distinction in kind (not one of reducible degree) - we must found phenomenology on a theory whose principles are not wholly derived from the history of the organism/system. Otherwise: psychologism. In terms of cognitive science, I am therefore saying (and I hope not too redundantly) that one must found phenomenology immediately (as opposed to mediately) on a syntactic layer of processing, where symbols are process-relevant (contra Smolensky), and therefore governed by rules that are not derived from the statistical history of the system. Now, since the whole idea of Turing/Husserl machines is that the rules governing the syntax of the machine are independent of the associative-causal history of the system (i.e. not based on a mechanism of association); and since this is the only explanatory framework available in the whole of cognitive science that avoids psychologism (which otherwise plagues every associationist approach), it follows that the symbolic "thought-machine" is the only candidate for a causally explanatory basis of phenomenology that avoids psychologism (Lopes, 2022).

That is, I suppose, more than a little surprising to cognitive scientists, phenomenologists, and everyone in-between. But I hope the reader sees how my hands are tied. The above reasoning is hardly willful; yet Yoshimi seems to accuse me of having an unsympathetic attitude to alternative approaches and of being an anti-pluralist. On the contrary, the logical entailments of the above reasoning have absolutely nothing to do with whether I am sympathetic or not sympathetic to alternative approaches to the idea of a 'LOT cum phenomenology' program. For example, if it can be shown that the distinction in kind between mechanisms of symbolic inference and mechanisms of association is false, and therefore that the distinction between the two is graded or continuous in such a way that they do not represent essentially distinct systems, then my reasoning is false. For then the principles of thought governing the one mechanism would be reducible to the other. Or if it is true that associationism does not entail psychologism, and therefore frameworks (deep ANNs, DST) that only appeal to the causal histories of systems/organisms in deriving the principles of thought are not psychologistic, then my reasoning is false. Yoshimi does not (and I think essentially cannot) show either of these entailments. Instead, Yoshimi's strategy is to say that I am contradicting myself by embracing what I am arguing against, or else to say that my argument does not generalize. I have attempted to rebut these claims, in part by presenting a new "forcing" argument, now grounded on the idea of associationism

and the phenomenological distinction in kind between passive and active synthesis and their respective causal bases in (for passive synthesis) mechanisms of association and (for active synthesis) mechanisms of symbolic inference. I hope this argument may prove less "forcing" for Yoshimi and more persuasive.

References

Buckner, C. (2018). Empiricism without magic: Transformational abstraction in deep convolutional neural networks. *Synthese*, *195*, 5339–5372.

Cain, M. J. (2016). *The philosophy of Cognitive Science*. Malden: Polity.

Cain, M. J. (2021). *Innateness and cognition*. New York: Routledge.

Chomsky, N. (2012). *The Science of Language*. New York: Cambridge.

Churchland, P. (2012). *Plato's Camera: How the physical brain captures a Landscape of Abstract Universals*. Cambridge: MIT.

Fodor, J. (1992). *Ernest Lepore. Holism*. Malden: Blackwell.

Fodor, J. (2002). *The Compositionality Papers*. New York: Oxford.

Fodor, J. (2008). *LOT 2*. New York: Oxford.

Fodor, J. (2015). & Zenon Pylyshyn. *Minds without meanings*. Cambridge: MIT.

Husserl, E. (1891). On the Logic of Signs (Semiotic). In *Early Writings in the Philosophy of Logic and Mathematics*. Trans. Willard. Dordrecht: Springer, 1994.

Husserl, E. (1901). *Logical Investigations*. Trans. Findlay. New York: Routledge, 2001.

Husserl, E. (1913). *Ideas I*. Trans. Dahlstrom. Indianapolis: Hackett, 2014.

Husserl, E. (1926). *Analyses Concerning Passive & Active Synthesis*. Trans. Steinbock. Dordrecht: Kluwer, 2001.

Husserl, E. (1929). *Formal & Transcendental Logic*. Trans. Cairns. Hague: Nijhoff, 1969.

Kelleher, J. D. (2019). *Deep learning*. Cambridge: MIT.

Lopes, J. (2021). On the psychologism of neurophenomenology. *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-021-09773-8

Lopes, J. (2022). Phenomenology as proto-computationalism. *Husserl Studies*.

Mandelbaum, E. (2015). Associationist Theories of Thought. *The Stanford Encyclopedia of Philosophy*. Eds. Edward N. Zalta & Uri Nodelman. 2022.

Marcus, G. (2019). & Ernest Davis. *Rebooting AI*. New York: Pantheon.

Yoshimi, J. (2009). Husserl's theory of belief and the heideggerean critique. *Husserl Studies*, *25*(2), 121–140.

Yoshimi, J. (2023). Pluralist phenomenology. *Phenomenology and the Cognitive Sciences*.