LONG
SHORTTER
MEMORY
Arity & Abstraction

What does it mean to hijack the latent space of a computation? Not just any computation, but a cognitive act about which we can say – a decision was made, an inference took place here. First there is the admission of a latent space itself, an interiority – to assert an algorithm could develop, in the language of Kant, an ‘inner sense’. Then there are the implications – that computation could maintain its own myriad languages of thought, as Fodor once proposed, that its acts are not performances for us, may not be performances at all, that they may ultimately be unintelligible to human reason.

By latent space we denote the tensorial data passed between inner layers of an Artificial Neural Network (ANN). Tensors are high-dimensional representations of input data, structures which are reshaped as they flow through the net. Reshaping here refers to dimensional plasticity, to moving fluidly between representations of an input space, learning higher level embeddings of data – in short, engaging in acts of multi-level abstraction.

The crisis of explication that characterises contemporary AI can be viewed either as a semantic limitation proper to the ANN model – its inability to attain the requisite level of concept formation – or instead a more fundamental limitation on the order of linguistic correspondence, a problem arising from the act of mapping human concepts to this latent space. A non-correspondence of vocabularies is merely the observation that no necessary bijection exists between language sets. Inuits may create $n$ words for snow, ‘umami’ may only be translatable through analogy, and so on. But the latter perspective implies an altogether stronger claim, namely the incommensurability of cognitive acts mediated by diverse languages, a principle first proposed by Feyerabend in historical form, and subsequently critiqued by Putnam.
A computational theory of mind can quickly converge on its own claim of incommensurability, with implications for the epistemic status of inferences made by AI. In this account, reason is modeled as a set of linguistic statements, a ‘canon’ of every sensical aethetic statement in a language. The assumption here is that reason can be modeled as a context-free grammar $R$, given by the tuple:

$$ R = (V, T, P, S) $$

Where $V$ is a set of variables representing the vocabulary (a set of strings), $T$ is a finite set of terminals, namely the symbols that form sentences, $P$ is a set of production rules that form the recursive definition of a language, and $S$ is a start symbol from which induction can progress.

$R$ is conceived as a grammar of inferential production rules, learned via interaction with an environment, along with a verification process, via a procedure we shall leave undefined for now. In this formalism, sets of statements are themselves sensical propositions, they can be combined and retain not only soundness but meaning. $R$ is constrained by its environment at the level of rule production, but not in terms of its generative capacity. To propose an infinite domain of such statements is to impose no theoretical limit on the conceptual labour characteristic of reasoning.

Even within the invariant grammar of $R$, a power set of statements is irreducible to the canon, demonstrable through Cantor’s Theorem – it is always possible to diagonalize out of $R$. It turns out the canon is no authority after all. The epistemic repercussions are considerable – if AI represents a bootstrapping of human rationality in the project of unpacking reasons, human reason cannot presume to circumscribe this domain. The aim of such a model is not to instigate a relativistic relation to AI, but rather to attempt to grasp the tractability of rational interpretation in the face of runaway statistical inference, to get a sense of the epistemic limits of said labour.

The strictly formal conception of reason offered by such a theory is not without its problems. Brandom reminds us that an account of reason cannot dismiss
either the pragmatic insistence on material rules of inference – the practice of
concept use – or the rational emphasis on the expressive role of logical
vocabulary. Likewise, the Sellarsian view of concept acquisition as “mastery of
the use of a word” implies a linguistic act integral to material practice. Pragmatic
objections as to the role of social speech acts in concept formation can be
integrated into the model in the form of agent interactions, but material
correctness remains an elusive concept in this formalism. A verificationist theory
of meaning – the claim that statements cannot be rational premises without
material validity – is irreconcilable with the unpacking of reasons sketched out by
a purely set theoretic account. Semantics provides a challenging hurdle to this
functional closure of reason, as exemplified by Putnam in developing the position
known as semantic externalism, the notion that meaning is not in the mind.

Such computationalist accounts of mind can be accused of over-emphasizing the
inferential role of formal logical vocabulary in reasoning – of conflating logic with
the act of justifying, of giving and asking for reasons. In Brandom’s expressivist
account of logic, the role of logical vocabulary is not as metalanguage, but rather
a formalization of ordinary vocabulary to the level of propositions – an epistemic
mediator of truth statements, akin to an expressive toolkit which makes explicit
our commitments. Logic neither undergirds nor circumscribes reason in
Brandom’s scheme – it is neither foundational nor strictly transcendent, but rather
autonomous – one can reason without recourse to logic, simply grasping the how
of causes without the formal expression of such. Reasoning can be, in a sense,
informal, without descending into purely irrational belief, and this forms part of
Brandom’s subtle account of reason, in which normative statements provide a
bridge to logical vocabulary, a position he calls normative pragmatics.

Nevertheless, contemporary AI confronts us with modes of statistical inference
that lack explication but exhibit validity within domains marked out by human
epistemic practices. An open attitude to inhuman reasoning might instead view
the semantic void at the heart of AI as a result of linguistic incommensurables
arising from a process of abstraction. Abstraction as a practice of mapping
between domains introduces a challenge to a theory of language in the realm of
representation, namely of providing an account of how cognition transitions in and out of the properly linguistic conceptual domain.

One might consider such a conceptual domain emerging from the learned tensorial representations native to ANNs. A vector space may not seem a plausible basis for a language at first sight, but what exactly is required for a language? No more than a vocabulary and a syntax. ANNs’ ability to learn internal embeddings of various spoken languages is the basis for recent developments in natural language processing. Assuming a *distributinal semantics*, relationships between words are captured in a lower dimensional vector space, allowing the model to exhibit an awareness of abstract concept classes, to construct analogies, and so on. Such abstractions are motivated by the *curse of dimensionality* – namely the difficulty of working with sparse vector representations – the latent semantic indexing is simply a side effect of this learned representation.

Embeddings are not limited solely to language – any input domain can be collapsed into a common vector space, as evinced by models which combine image and word data into a single embedding, allowing for breakthroughs in image captioning. Autoencoders, another family of ANNs, also leverage an ability to reduce the dimensionality of an input domain, essentially achieving a distributed form of compression akin to an embedding. These acts alone are not evidence of language formation as such, but rather a capacity to abstract both the syntax and semantics of existing (visual, textual) vocabularies into a novel encoding. It is in transitioning out of embeddings back into the input domain that we see how they can constitute internal languages – take for example, translation ANNs that are able to translate between language pairs they were never explicitly trained on, indicating an internal *interlingua* with a strong claim to the status of a language.

The era of *feature engineering* – spoon-feeding a vocabulary to a model which could be used to reason about its inferences, to account for its decisions by shaping its representations of the input data at the outset, is quickly receding.
This conceptual dictation of man over machine turned out to be counterproductive – sidelining human concepts improved the models’ ability to generalize. Even within supervised deep learning today, there is no shared means of reasoning about decisions – concepts are applied as labels attached to desired classificatory outcomes, which themselves cannot be the basis for inference.

If embeddings create the ground for internal language models, then agent interactions form the basis for the development of external languages proper to ANNs. Machine learning techniques based on interaction – namely adversarial modes of learning – have created a phase transition in the generative capacity of AI models. A predictive processing (PP) model of mind ascribes a central role to such generative models in its account of perception. In PP, perception is characterised as the output of a predictive model, and sense data no more than an error to be back-propagated. The refinement of such models, which is the continual process of perception itself, provides the means to learn a causal matrix that underpins our knowledge of the world. Such an account of perception resonates with contemporary deep learning, in which generative nets often train themselves against discriminatory nets. The development of learning techniques which focus on agent interaction promises a fertile ground for agent languages, arising from acts of communication in service of just such a generative optimization process as that posited by PP.

Another key architectural development for dealing with language in deep learning models is Long Short Term Memory (LSTM). LSTMs constitute a family of ANNs composed of recurring units housing \textit{tanh} and \textit{sigmoid} activation functions marshaled by \textit{forget gates}. The nets learn which information to retain, and which to discard, when performing tasks on sequences of data. This is an architecture for forgetting, imperfect recall as a means of managing attention at different scales. The labour of grasping patterns from a stream of symbols hinges on discrimination, compression, detecting redundancy, and judging relevance. Strategic forgetfulness, it turns out, is integral to the performance of such feats of learning. These feats, like the entire regime of deep learning, still exist firmly within the domain of inductive logic, framed entirely as the generalization of patterns in input data.
Brandom’s *inferentialist* account of reason would locate contemporary AI at merely the first rung of inference – to climb the ladder of abstraction that affords the kind of self-reflection proper to those acts we consider rational is the challenge put to AI by such accounts. If machine learning is to develop beyond mere acts of labelling – classificatory feats of inductive inference – then it must develop language(s), be they internal *mentalese* or external modes of communication. As Brandom notes, to be counted as a concept user one must move beyond simple differential responsiveness to stimuli, as exhibited by a thermostat or a pigeon trained to respond to different colours – to acquire concepts means instead to deploy them within a web of inferences, to offer them as premises or conclusions in acts of reasoning. This marks the rationalist distinction between sentience and sapience.

In order for concepts to enter into rational roles, they must provide justification for each other on multiple levels of abstraction. A hierarchy of synthetic and analytic concept formation must be at play. It is this multi-level labour of abstraction which latent spaces and embeddings make theoretically possible. But even so, AI remains bound to an inductive regime of logic that precludes the kind of normative claims which play a central role in reasoning. The work of Judea Pearl is instructive here in suggesting a path forward for AI, not only as a critique of deep learning, but as an exploration of inferential webs in the form of bayesian networks. Pearl develops computational modes of causal inference which are conspicuously absent from mainstream AI.

One way of describing ANNs' capacity for abstraction might be through the concept of *arity*. ANN architectures can be formalized as a system of nested functional composition, a complex network of activation functions learned over time. The *arity* of a function denotes its promiscuity, its degree of interdependence within a network of relations. Neurons in this analogy are simple functors which take *n* functions and compose them with an activation curve, outputting a single function in their place – they exhibit *n-arity*, where *n* is bounded by the number of neurons in a layer, in the limit case of a fully connected architecture. The *arity* of a connection in such a network is the dimensionality of that relation.
Arity can act as a short-hand for both connectivity and dimensionality in a functional model, akin to degree in graph models, an indicator of valence within a broader system. Learning is in effect a rewiring of the neural architecture that renders plasticity as a variadic reconfiguring of the net. Learning turns neurons into variadic functions, as some functional relations are enhanced and others discarded by tuning their weights in an effort to minimize error – variadic plasticity within a latent space becomes the key to generalization.

The work presented in the pages that follow is a visual and textual exploration of abstraction in ANNs and an inquiry into the generative capacity of LSTM architectures. The methodology presented is akin to administering an injection, a violation of sorts, one which questions the integrity of a body. Submitted to this procedure, a black box of cognition reveals a membrane with dispositions, a tangle of tensorial fibre, matter imbued with memory.

An LSTM is injected with random data adopting a specified distribution (pareto, gaussian, etc), effectively hijacking its latent space. This is fed through the untrained network, which outputs values as intensities of light. The output is an exposure of its internal structure – a nesting of activation curves, smooth gradients, bulbous growths, and fibrous strands. The exposed image passes through a second network in the form of an autoencoder which scales the output to arbitrary sizes, using a lossy representation stored in a distributed manner amongst its neurons. Dimensionality reduction here is a means of compression, collapsing representations in an input space into their characteristic features.

Alongside these exposures a Recurrent Neural Net (RNN) learns to read a corpus of contemporary theory on computation, finitude, and mathematics. The RNN has read the work of Luciana Parisi, Gregory Chaitin, Quentin Meillasoux, and others.
Language is but a stream of characters to the RNN. It has no a priori notion of the concept ‘word’, let alone grammar. It too determines which associations to forget and which to retain, a limited memory allowing it to compose synthetic philosophical claims. This is a regime entirely devoid of semantics but rich in syntax and sensicality. It exhibits a form of creativity redolent of Desjardin’s theory of knowledge, namely emergence through generative recombination. Its generative capacity – learnt through many hours of training – hints at the constructibility of grammars, at a generalized axiomatics of inhuman linguistic formalisms.


025029

A passion for a contingency in the world.
02507

A situation that is also a part of the situation.
02508

The procedure of fidelity, an ontological schema.
02595

The axiomatics of a most general universe.
02571

Faithful to the multiple as multiple-being.
A non-language must be a presentation.
Ignorance is the axiom of choice.
02567

The inadmissibility of the pure event.
02569

A truth is only a periphery.
The critique of the individual as a minority of segments, which interpret one another forever.
02532

Predicates retreat to expressions and intensions.
02578

It is a question of distinguishing between bodies, the book is also a milieu of seven.
02583

The randomness of the capitalist axiomatic substratum.
A real object is a statement and the possibility that it fits the pattern of a theory.
A secret is always a correspondence between bodies.
The program is evolved by a process of interaction.
The new situation retains a stratum of its formalization.
025109

A conception of reason as a continual contagion of computation.
02592

Taking a computational account of processes as a historical combinatorial algebra.
A prototype of infinitesimal programs that transform the notion of a group.
02588

The fundamentally digital architecture of the brain projects a local singularity.
02534

The computation of cognition cannot be stored in discrete or finite quantities.
02528

Outside the risks are enough to define the nomads.
02543

The effects of the state are supernumerary to the situation.
02545

The landowner finds it impossible to establish the constitution of the animal.
CLAUSES
Perception is never a mere contact of the mind with the object present; it is impregnated with memory-images which complete it as they interpret it. The memory-image, in its turn, partakes of the 'pure memory', which it begins to materialize, and of the perception in which it tends to embody itself: regarded from the latter point of view, it might be defined as a nascent perception. Lastly, pure memory, though independent in theory, manifests itself as a rule only in the coloured and living image which reveals it.

Whenever we are trying to recover a recollection, to call up some period of our history, we become conscious of an act sui genesis by which we detach ourselves from the present in order to replace ourselves, first in the past in general, then in a certain region of the past - a work of adjustment, something like the focussing of a camera. But our recollection still remains virtual; we simply prepare ourselves to receive it by adopting the appropriate attitude. Little by little it comes into view like a condensing cloud; from the virtual state it passes into the actual; and as its outlines become more distinct and its surface takes on colour, it tends to imitate perception.
We place ourselves at the point of view of a mind unaware of the disputes between philosophers. Such a mind would naturally believe that matter exists just as it is perceived; and, since it is perceived as an image, the mind would make of it, in itself, an image.

We have examined what takes place in the organized body, we have seen movements transmitted or inhibited, metamorphosed into accomplished actions or broken up into nascent actions. These movements appear to us to concern action, and action alone; they remain absolutely foreign to the process of representation. We then considered action itself, and the indetermination which surrounds it and is implied in the structure of the nervous system, - an indetermination to which this system seems to point much more than to representation.

We propose to show, in the course of this study, that both are right; and that there is no essential difference between the light and the movements, provided we restore to movement the unity, indivisibility, and qualitative heterogeneity denied to it by abstract mechanics; provided also that we see in sensible qualities contractions effected by our memory. Science and consciousness would then coincide in the instantaneous. For the moment all we need say, without examining too closely into the meaning of the words, is that the point P sends to the retina vibrations of light. What happens then?
Reject then the share of memory, consider perception in its unmixed state, and you will be forced to recognize that there is no image without an object. But, from the moment that you thus posit the intracerebral processes besides the external object which causes them, we can clearly see how the image of that object is given with it and in it: how the image should arise from the cerebral movement we shall never understand.

Far from depriving matter of anything perceived, we must on the contrary bring together all sensible qualities, restore their relationship, and re-establish among them the continuity broken by our needs. Our perception of matter is, then, no longer either relative or subjective, at least in principle, and apart, as we shall see presently, from affection and especially from memory; it is merely dissevered by the multiplicity of our needs.

This radical powerlessness of pure memory is just what will enable us to understand how it is preserved in a latent state. Without as yet going to the heart of the matter, we will confine ourselves to the remark that our unwillingness to conceive unconscious psychical states is due, above all, to the fact that we hold consciousness to be the essential property of psychical states so that a psychical state cannot, it seems, cease to be conscious without ceasing to exist.
As, in such cases, the object has disappeared while the brain persists, he holds that the cerebral phenomenon is sufficient for the production of the image. But it must not be forgotten that in all psychical states of this kind memory plays the chief part. Now, we shall try to show later that, when perception, as we understand it, is once admitted, memory must arise, and that this memory has not, any more than perception itself, a cerebral state as its true and complete condition.

A priori, indeed, we may expect the clear distinction of individual objects to be a luxury of perception, just as the clear representation of general ideas is a refinement of the intellect. The full conception of genera is no doubt proper to human thought; it demands an effort of reflexion, by which we expunge from a representation the details of time and place. But the reflexion on these details - a reflexion without which the individuality of objects would escape us - presupposes a faculty of noticing differences, and therefore a memory of images, which is certainly the privilege of man and of the higher animals.
Memory actualized in an image differs, then, profoundly from pure memory. The image is a present state, and its sole share in the past is the memory whence it arose. Memory, on the contrary, powerless as long as it remains without utility, is pure from all admixture of sensation, is without attachment to the present, and is consequently unextended.

Let us admit for a moment that the past survives in the form of a memory stored in the brain; it is then necessary that the brain, in order to preserve the memory, should preserve itself. But the brain, in so far as it is an image extended in space, never occupies more than the present moment: it constitutes, with all the rest of the material universe, an ever renewed section of universal becoming. Either, then, you must suppose that this universe dies and is born again miraculously at each moment of duration, or you must attribute to it that continuity of existence which you deny to consciousness, and make of its past a reality which endures and is prolonged into its present.
This survival of the past per se forces itself upon philosophers, then, under one form or another; and the difficulty that we have in conceiving it comes simply from the fact that we extend to the series of memories, in time, that obligation of containing and being contained which applies only to the collection of bodies instantaneously perceived in space. The fundamental illusion consists in transferring to duration itself, in its continuous flow, the form of the instantaneous sections which we make in it.

Consciousness, then, illumines, at each moment of time, that immediate part of the past which, impending over the future, seeks to realize and to associate with it. Solely preoccupied in thus determining an undetermined future, consciousness may shed a little of its light on those of our states, more remote in the past, which can be usefully combined with our present state, that is to say, with our immediate past: the rest remains in the dark. It is in this illuminated part of our history that we remain seated, in virtue of the fundamental law of life, which is a law of action: hence the difficulty we experience in conceiving memories which are preserved in the shadow.

Excerpted from Henri Bergson, 'Matter and Memory'