Formalizing Analogical Transitions within a Multi-Level Dynamic Mechanistic Framework

Talk given at 'Cold Bodies, Warm Machines' Conference, Düsseldorf, 09-11/09/16.

Firstly, what is the Dynamic Mechanistic Framework?[1] Simply put its the way in which we can understand the properties of a system by disassembling it and seeing how its parts work either individually or in combination with other parts. It is a mechanistic framework because it acknowledges that the concept of *mechanism* has been, and still is, a key idea of modern science by which the characteristics of a system are explained according to the causal structure of its parts. Though the *classical* mechanistic framework has been critically overcome by the recognition of the ubiquity of non-linearity in dynamic systems theory, the concept of mechanism as structure-based explanatory framework for the functioning of a system may be retained if it is suitably dynamicized. This mechanistic framework is multi-level and dynamic because it is neither 'ruthlessly reductionist' nor rabidly anti-reductionist – it recognizes that the functional characteristics of a system at one level of analysis may be emergent properties that are not well described by the lower level dynamics that support them, without claiming that these properties are irreducible to the behaviour of its structural components (this is called weak as opposed to strong emergence). The framework itself is dynamic since it progresses by several stages of *decomposition* and *localization* of functional properties in structural components – if a property cannot be directly localized in a structural component then it may be hypothesized as a property of the interaction of two or more components, and if this hypothesis fails to hold a further level of complexity can be adduced, up to the point where a property can be assumed to emerge at the global level of the whole system.

Another important consideration that necessitates a *multi-level* analysis is that not only is order dependent on structure but randomness and unpredictability are relative to the theoretical framework or computational set-up (architecture, processing power, input stream). A system or an input is unpredictable only in relation to the theoretical framework that imposes its form of prediction. To understand something means to impute to it some kind of characteristic behaviour under certain circumstances, and this entails prediction, abstraction, and generalization. If we want to act on the world we cannot renounce prediction or abstraction, and we cannot thereby avoid the possibility of some random perturbation to our system. However, we may construct systems that are more or less robust or resilient to perturbations, more or less capable of the dynamic revision and reconstruction of theoretical frameworks. Crucially, different theoretical frameworks are necessary for each level of analysis, and the dynamics of randomness, unpredictability, and robustness or resilience to perturbation are distinct at each level[2] – for example physics, biology and social systems may be distinguished along precisely these lines, as well as classifying the descriptive capabilities of different architectures in

---

[1] The methodology and its explanation are derived from Bechtel, W. (2010) *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. MIT Press.

[2] See: Calude, C. & Longo, G. (2015) *Classical, Quantum and Biological Randomness as Relative Unpredictability*. Natural Computing, Springer.; Longo, G. & Montévil, M. (2012) Randomness Increases Order in Biological Evolution. *Frontiers in Physiology*, n. 3, 39.; Bravi, B. & Longo, G. (2015) The Unconventionality of Nature: Biology, from Noise to Functional Randomness. Invited Lecture, *Unconventional Computation and Natural Computation Conference* (UCNC), Auckland (NZ), 31/8 - 4/9/2015, proceedings to appear in Springer LNCS, Eds. Calude, C.S. & Dinneen M.J. pp. 3-34.

the computational hierarchy from finite automata through to universal Turing machines.[3] Moreover, the multi-level dynamic mechanistic framework provides an engineering perspective by which the functional property of intelligence itself is decomposed and the resulting explanatory hypotheses enable traction not only on the problem of engineering artificially intelligent agents, but also on re-engineering the human.[4]

Why formalization? Formalization may be understood as the explicitation of implicit theoretical frameworks or models, and making these models explicit allows for them to be analysed and transformed. Implicit rule-governed practices will always be in excess of any explicit formalization of them (this was amply demonstrated by the later Wittgenstein's investigation of language games),[5] however formalisation may act as a point of leverage giving externalized traction on the ideological sedimentation of implicitly structured practices (a point well made by Joshua Epstein with regard to the efficacy and politics of computational modeling).[6] In this sense formalization may be understood as a kind of decomposition of the implicit structure of a theoretical framework, allowing us to grasp its functional characteristics and change it. Moreover, formalization of a model makes possible its implementation in another substrate, such as an artificially intelligent agent.

Why do we talk about analogy here? Firstly, we can consider the functionalist analogy, which understood the brain as a computer, as the initial stage in a dynamic process of decomposition and localization of the functional properties of cognition (a process that moves from the first-wave functionalism of Putnam etc based on symbol manipulation to the dynamic model of neofunctionalism based on connectionism put forward by the likes of Bechtel). Already with Turing it was recognized that if the brain in a very general sense is something like a computer it can be specifically differentiated from the kind of computational processes occurring on a discrete state machine at the lower levels of the computational hierarchy since while the latter is functionally context-free and iterates without variation, the former is context-sensitive or further, as Turing had it 'prone to exponential drift', which we may understand in today's parlance as 'sensitive to initial conditions' or non-linear.[7] This breakdown in the analogy does not definitively destroy its usefulness – in fact an analogy that directly maps is hardly fruitful - the usefulness of analogies depends on whether any testable consequences can be deduced from them.[8] In this case the analogy has drawn out the difference between computational processes occurring in discrete state machines and those occurring in biological brains. In fact, modeling itself may be understood at the most general level as an analogical process based on abstraction that ignores certain details while highlighting how other features overlap.

Secondly, analogical transitions between different contexts are not just a natural aspect of human intelligence but a key cognitive operation of any intelligent agent, along with

[3] Crutchfield, J.P. (1994) The Calculi of Emergence: Computation, Dynamics, and Induction. *Physica D, Proceedings of the Oji International Seminar Complex Systems - from Complex Dynamics to Artificial Reality.* Elsevier North-Holland, Inc. New York, NY, USA.
[4] Negarestani, R. (2014) *The Labor of the Inhuman*, in Eds. Mackay, R. & Avanessian, A. #Accelerate: The Accelerationist Reader. Urbanomic.
[5] Wittgenstein, L. (2009) *Philosophical Investigations*. Wiley-Blackwell. Cf. Brandom, R. (1994) *Making it Explicit: Reasoning, Representing and Discursive Commitment*. Harvard University Press. pp.13-30
[6] Epstein, J. M. (2007) Generative Social Science: Studies in Agent-Based Computational Modeling. Princeton Univeristy Press.
[7] Longo, G. (2008) Laplace, Turing and the "imitation game" impossible geometry: randomness, determinism and programs in Turing's test, in eds. Epstein, R., Roberts, G., & Beber, G. *Parsing the Turing Test*. pp. 377-413, Springer.
[8] Arzi-Gonczarowski, Z. (1999) Perceive this as that – Analogies, artificial perception, and category theory. *Annals of Mathematics and Artificial Intelligence*, 26, 215–252.

induction, recursion, and the management of behaviour through the apprehension of contextual frames, discursive markers, and conceptual contents.

As Arzi-Gonczarowski notes, Kant understood that analogy isn't based on an imperfect similarity of two things, but a *perfect similarity of relations between two quite dissimilar things*[9] – and moreover understood these kind of supramodal transits in terms of categories. It is extracting these perfect similarities that underlies the flexibility and generativity of higher intelligence (whether carbon-based, silicon-based or a hybrid). Gentner, for instance, models the use of analogy in various subprocesses of learning and reasoning, such as memory retrieval, mapping and structural alignment between different perceptions or environmental contexts, as well as evaluation, abstraction, re-representation and adaptation.[10]

Analogy is key to the development of AI for the same reason that it underlies human reasoning. Furthermore, in interacting with human agents, artificially intelligent agents will need to analogize in order to respond to human behaviour. At the very general level we can distinguish two broad forms of analogical transition – one based on recognition of already given similarities through a diagnostic process of interpretative analysis, and another that is capable of creatively discovering or inventing new analogical transitions through a generative process of synthesis.

In order to understand the centrality of analogical transitions to the general problem of intelligence it is necessary to acknowledge the way in which cognition is embodied, embedded, enacted and extended. It is impossible to separate perception, conception or action from the environment that supports it and to which it is directed. The fact that interaction with an environment plays an essential role in intelligence was recognized in Turing's early writings, and developed more recently by others, such as Clark and Hutchins. The latter proposes the concept of a cognitive *supraindividual* including both the environment and the intelligent agent acting within it, which can be fruitfully aligned with the insights of second-order cybernetics.[11]

This generatively extensible supraindividual context-sensitivity can be applied not only to the intelligent agent but also to its perceptions, conceptions, and actions, which can take on an infinite number of different senses, since there is no limit on the number of contexts in which their meaning is transformed. However, this relativity, plasticity, and genericity should not collapse into absolute relativity or sweeping skepticism regarding epistemic limitations since some critical invariable aspect of meaning is held by individual perceptions, conceptions or actions and this must be shared across contexts in order for analogical transitions to make sense (in fact, we should rather think of higher intelligence as meta-contextual for this reason).

Grasping this invariance may present problems in natural language due to the tendency toward an infinite regress of definitions, and definitions of definitions, each of which will have context-sensitive connotations as well as invariants (this is the problem of the excess of practical reason to formalization we talked about earlier). However, contemporary mathematics has fortunately enabled a different approach.

[9] Caygill, H. (1995) *A Kant Dictionary,* Blackwell Publishers, Great Britain. p.66 quoted in Arzi-Gonczarowski, Z. (1999) Perceive this as that – Analogies, artificial perception, and category theory. *Annals of Mathematics and Artificial Intelligence*, 26, 215–252.
[10] Gentner, D. (1998) Analogy, in: *A Companion to Cognitive Science.* Blackwell, chapter II(1), pp. 107–113. quoted in Arzi-Gonczarowski, Z. (1999) Perceive this as that – Analogies, artificial perception, and category theory. *Annals of Mathematics and Artificial Intelligence*, 26, 215–252.
[11] Hutchins, E. (1995) *Cognition in the Wild.* MIT Press, Cambridge, MA.

Category theoretic morphisms give formal tools to rigorously describe structure preserving paths (that is, invariants) between *different* perceptions, conceptions or actions within a *single* interpretative context, or dually, for a *single* perception, conception or action across *different* interpretative contexts.

When an analogy is modeled by a morphism, then the monotonicity of that morphism captures the invariant core or 'uniformity' of the analogy, while the non-monotonicity provides a detailed description about the points where the analogy breaks down.[12] Such break-downs or slippages in the analogical transition do not invalidate it but rather require the acknowledgment of non-monotonic variability, and the construction of a non-rigid abstract schema that generalizes over these differences, avoiding determination only where necessary.

The generation of perceptually, conceptually of actionally incisive cognitive images of the environment, as well as the generation of highly structured analogies, are both based on awareness of lawlike patterns or structure preserving transformations between perceptual, conceptual or actional constituents. The ability to construct an intelligent cognitive image of the environment is then linked to the capacity to critically analyse and creatively synthesise insightful analogies since both require the exploitation of lawlike patterns or structure preserving transformations.

Since neural nets can trawl vast archives and compile massive data sets, they are able to discern and catalogue the relationship between a far greater number of patterns than any single human is capable of. Moreover, because they operate according to formalized criteria that are both inspectable and reprogrammable, they may explicitly reveal patterns that are obscured for humans by their ideological sedimentation in implicitly structured practices. This is not to say that neural nets or algorithms are immune to ideological bias – on the contrary, prejudice is often embedded in computational systems, often this is opaque to analysis due to the complexity of the system or to proprietary rights over its management, and moreover this may exacerbate the expression or enaction of bias in computer assisted human behaviour. However, formal tools can give us traction on this problem, not only allowing for the emancipatory amelioration of the computational system, but also facilitating the acknowledgement and transformation of predispositions in human behaviour. For example, a neural net designed by a team at google was programmed to trawl Google News texts looking for patterns of conceptual associations or word embeddings that could be represented by a simple vector algebra in a multi-dimensional vector space.[13] These word embeddings, for example 'Paris is to France as Tokyo is to Japan', capture structure preserving transformations or analogical transitions in the formal language of vector space mathematics. However, since the neural net discovered these analogical transitions by consuming a vast diet of professional journalism, it inadvertently picked up the latent gender bias that pervaded the input data as a whole. Even though gender bias was not noticeable within each single input article, at least to the journalists and editors responsible, the aggregate output displayed a marked prejudice, so that the neural net would make analogical transitions such as 'man is to computer programmer as woman is to homemaker'. Because this bias was captured and represented by formal mathematical tools it was not only legible in the warped geometry of the neural net's vector space but also transformable by applying a distortion

---

[12] This is demonstrated in in Arzi-Gonczarowski, Z. (1999) Perceive this as that – Analogies, artificial perception, and category theory. *Annals of Mathematics and Artificial Intelligence*, 26, 215–252.

[13] Bolukbasi, T., Chang, K-W., Zou, J., Saligrama, V. & Kalai, A. (2016) Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. arXiv:1607.06520.

of inverse proportions, a process the google team called 'hard de-biasing'. However, judging the shape of the warp, or the appropriateness of the analogical transitions that populated it, could not itself be formalized and automated in the present system since it would require some kind of artificial *general* intelligence capable of making normative judgments regarding the appropriateness of associations made on the basis of gender difference. As a result this task was outsourced to democratic vote by a small group of Amazon's mechanical Turks. The irony that there is a marked gender asymmetry in the makeup of the mechanical turk workforce should not be lost here (hard debiasing relies on structurally embedded employment asymmetries). Nevertheless, it should not detract from the remarkable power of the neural net and its formalized mathematical tools to apprehend and transform the latent ideological bias sedimented in journalistic practice.