In Defence of Functional Analysis

Joe Faith

School of Cognitive and Computing Sciences, University of Sussex, Brighton, UK

`josephf@cogs.susx.ac.uk`

Abstract

Computationalism presupposed a modular-functional analysis of cognitive behaviour, and its failure has encouraged the search for alternative analytical techniques, such as behavioural decomposition and dynamical systems theory. This pa-

of what Hendriks-Jansen

systems that detect pheromones, the presence of the affordance that the pheromone signals is proportional to the activity of a local set of receptor cells. In contrast, when an odour is detected by an olfactory bulb it produces an oscillatory pattern that is distributed across a very wide area [44]. How can neural mechanisms evolve, or learn, to use such general sensory systems to reliably detect the presence of an affordance?

In such cases, the more successful mechanisms will be those that are better able to generalise from one case of a behavioural regularity to another. Consider shift invariance. In this case the regularity required is a similar response to an image that may be presented at different positions on a retina. It would be possible to teach, or evolve, a Type II mechanism to respond suitably to a number of particular image presentations. However, since it is a Type II mechanism then the way in which it achieves these responses may have nothing in common with each other: the behavioural regularity may not be due to a regularity in the underlying mechanism.

This mechanism is likely to be robust against certain changes in input: the addition of noise to the input would only change each input channel by a small amount, and thus we may expect the response of the mechanism to be similar. However other changes, such as moving the image on the retina, result in each input channel changing by a large amount — spots that were dark will now be light, and *vice versa*. There is no reason why a previously successful mechanism would produce similar behaviour given such different novel inputs. By contrast, a Type I solution would, by definition, involve some functional element that responds similarly to the same pattern regardless of its particular position. Such a solution, if it can be found, will generalise more robustly and will thus be favoured by evolution or a learning regime.

For example, one possible (though completely artificial) Type II solution to learning shift invariance would be a look-up table that lists a number of different retinal inputs and their required responses. This solution is completely incapable of 00.37(w-ning)90999.7(b)-1000.67(yw)999.349(and)-10000.3(the)-14999.7(lsuted)-10000.3(im-)]TJ01 that the

positvae cases in commoe.

Type I rghaniaution   nat be a propy of the im-
mechanis; maýt
p   For

active perception, a sensor morphology closely tied to the structure of the environment, and a representational architecture in order to produce adaptive behaviour.

The evolutionary pressure to generalise behaviour thus produces a pressure to localise function. Thus we find that the most robust, stereotyped, behaviours are produced by the most functionally specialised mechanisms. For example, sensory-motor behaviours, such as saccades or fixation movements, that have to be very fast and reliable tend to be produced via very clearly defined topographic cortical maps [39].

It is also worth remembering that not all components are localised, structurally individuated entities like hearts and lungs: in small animals the functions of respiration and circulation can be achieved by diffusion processes. This does not imply that there are no entities — stomata etc — that carry this function, but that they form a distributed, functionally individuated, "component" or subsystem; rather than a localised, structurally indivuated, one. Similarly, componential-functional decomposition in neural networks need not imply the existence of "grandmother" cells, or even of clearly delimited modules. Something can play a well-defined functional role with respect to other components even if it is not topographically localised. For an intuition pump, think of the geographically diffuse functional components of human societies, such as political organisations, classes, companies etc. As was shown above, it is usually only the most stereotyped behaviours that result in localised functional modularisation; and even these do not work in isolation. For example, although only a small number of neurons are directly involved in the gill-withdrawal reflex of *Aplysia*, up to 300 others are simultaneously activated, since siphon stimulation *also* causes many other behaviours: mantle contraction, inking, mucus release, postural changes, respiratory pumping etc. [2]

Functions Without Computations

It must be emphasised that a functional decomposition is relative to a behaviour of the overall system [14]. This is the crucial difference between this approach to functional analysis, and the *modular*-functional analysis of classical computationalism. Fodor [23] argues for the existence of modules that are the prior explanatory atoms of all cognitive behaviour. Modules are general purpose, they play the same role in all behaviour, and so have a fixed function. However, the functional analysis given above *starts* from a particular behaviour, and then asks how it is achieved. Analysing different behaviours may reveal a different functional decomposition with no component playing the same role in each case. It is modularity, rather than functional analysis *per se* that defines a computationalist perspective.

When a single unit is described as representing the presence of a target triangle, this is a description of how

a mechanism achieves a behaviour. Representation is not what an entity *is*, but rather what it *does* in a behavioural context. Representation is not merely a correlation between internal and external state that only exists for an external observer, but a relational, functional property between a mechanism and a particular behaviour that it displays.

3   Functional Analysis and Dynamical Systems Theory

What do we gain by a functional analysis of such simple systems? After all, if a functional analysis is needed in a simple case like the triangle-seeking robot, then there seems no principled reason why it could not be applied in even simpler cases. For example, the neural mechanism of the Braitenberg light-seeking vehicle is actually no more than a pair of crossed wires. However we *could* describe it in functional terms as an input module (comprising the two sensors) that passes a representation of the world (the state of the two wires) to the output module (the motors). Although no principled "bottom line" for functional explanation can be given, this need not invalidate its use. After all, if we consider a vacuum that contains just a few molecules of a gas, then con-

assumption of classical computational psychology, that cognition takes place in a mental "space" built from representations of the world. The only difference is that the rules for the manipulation of these representations are essentially temporal. The weak form of DST therefore stands in the same relation to computationalism as classical connectionism [16]: it shares an explanatory framework, but uses more complicated rules.

The stronger form of DST therefore restricts state variables to be non-mental, i.e. *physical*, properties of the system and its environment [4] [46]. This approach was first used 50 years earlier by Ross Ashby [3]. In the case of agents built from a neural network, the obvious choice of state variables are the activation of single neurons or neural masses [33][24]. However, this restriction has consequences for how DST can be used.

DST is supposed to provide an explanation of how cognitive behaviour is produced, rather than simply describing the dynamics of the internal mechanism. In order to do this a DST model must therefore include not just a whole agent but also its *environment* (though see [34]). Let us assume for the moment that determining the evolution equations for state variables that are internal to the agent is unproblematic. The environment can then be handled in one of three ways. The first is to ignore it by leaving any environmental impact on the system as undetermined parameters: this is obviously no answer for analysing the behaviour of whole agents. The second, as advocated in [46] and [4], is to treat the environment as a dynamical system in its own right, tightly coupled to that of the agent. This, however, is a form of Laplacean reductionism. Consider trying to produce a DST model of a bird trying to land in a tree. The bird must coordinate its body with a swaying branch using its eyes, brain, and muscles. The obvious state variables for the bird will include retinal cell inputs, the activations of the neurons in the visuo-motor system, and muscle nerve outputs. However the retinal inputs to the system will be affected by the most trivial changes in the environment: the wind catching a leaf and causing a shadow to move, for instance. This perturbation of input will, in turn, alter the trajectory in the phase space of the system. If we want to know how the bird lands reliably despite the incidental movement of so many factors in the environment, then it seems as though we will have to model the tree in as much detail as the nervous system of the bird. This is impossible in practise, even if we agree with Laplace that it may be possible in principle. The only case of a full DST analysis of a whole agent-environment system that I am aware of is [33], in which the environment is completely static. In [4] the neural network controller for a hexapod robot is modelled as a dynamical system, but its environment is treated as the body that it controls (the environment external to the body is again assumed to be static), and only a localised 5-neuron subsystem

is *analysed* as such. (This is not meant to detract from the great subtlety of the evolved design. The point is that it is not possible to fully appreciate it from a purely dynamical systems perspective.)

The only alternative for DST is to postulate features of the phase space of the agent-environment system that are immune to the incidental, un-modellable, changes to state variables. This is the way in which Walter Freeman uses DST to model oscillations in the olfactory bulb [25]. Two points should be noted about this model. The first is that it is *not* an attempt to model a wholeform   tri

analil999.959(virony

that the presence of affordances in the environment be

References

[1] P.E. Agre and D. Chapman. *Pengi: An implementation of a theory of activity*, pages 196–201. 1987.

[2] J.S. Altman and J. Kien. Highlighting *aplysia*'s networks. *Trends in Neuroscience*, 13(3), 1990.

[3] W.R. Ashby. *Design for a brain: The origin of adaptive behaviour.* Chapman Hall, 1952.

[4] R.D. Beer. A dynamical systems perspective on autonomous agents. Technical Report 92-11, Case Western Reserve University, Cleveland, Ohio, 1992.

[5] R.D. Beer and J.C. Gallagher. Evolving dynamical neural networks for adaptive behaviour. *Adaptive Behaviour*, 1(1), 1992.

[6] V. Braitenberg. *Vehicles : experiments in synthetic psychology.* MIT Press, 1984.

[7] R. Brooks. A robust

[32] F. Huber and J. Thorson. Cricket auditory communication. *Scientific American*, 253(6):46–54, 1985.

[33] P. Husbands, I. Harvey, and D. Cliff. Circle in the round: State space attractors for evolved sighted robots. *Robotics and Autonomous Systems*, (15), 1995.

[34] N. Jakobi. Half-baked, *ad hoc* and noisy: Minimal simulations for evolutionary robotics. This volume.

[35] J.S. Kauer. Contributions of topography and parallel processing to odor coding in the vertebrate olfactory pathway. *Trends in Neuroscience*, 14(2):79–85, 1991.

[36] P. Maes, editor. *Designing Autonomous Agents: Theory and Practise from Biology to Engineering and Back*. Special Issues of *Robotics and Autonomous Systems*. MIT, 1991.

[37] D. Marr. Artificial intelligence: A personal view. In M. Boden, editor, *The Philosophy of Artificial Intelligence*. Oxford University Press, 1977.

[38] M.J. Matarić. Navigating with a rat brain: A neurologically-inspired model for robot spatial representation. In Meyer and Wilson [40], pages 169–175.

[39] M. Merzenich and J. Kaas. Principles of organization of sensory-perceptual systems in mammals. *Progress in psychobiology and physiological psychology*, (9):1–42, 1980.

[40] J.A. Meyer and S.W. Wilson, editors. *From animals to animats: Proceedings of the first international conference on the simulation of adaptive behaviour (SAB90)*. MIT Press, 1991.

[41] F. Mondada and D. Floreano. Evolution of neural control structures: some experiments on mobile robots. *Robotics and Autonomous Systems*, 16:183–195, 1996.

[42] N.H. Packard. Intrinsic adaption in a simple model for evolution. In C. Langton, editor, *Artificial Life: Proceedings of the workshop on artificial life*, Santa Fe Institute studies in the sciences of complexity. Addison-Wesley, 1988.

[43] C. Scheier and R. Pfeifer. Classification as sensory-motor coordination: A case study on autonomous agents. In F. Moran, A. Moreno, J.J. Merelo, and P. Chacon, editors, *Advances in artificial life: proceedings of the third European Conference on Artificial Life*, number 929 in Lecture notes in artificial intelligence. Springer, 1995.

[44] C.A. Skarda and W.J. Freeman. How brains make chaos in order to make sense of the world. *The Behavioral and Brain Sciences*, 10:161–195, 1987.

[45] T. Smithers. Taking eliminative materialism seriously: A methodology for autonomous systems research. In F.J. Varela and P. Bourgine, editors, *Towards a practise of autonomous systems: proceedings of the first European Conference on Artificial Life*, 1992.

[46] T. Smithers. What the dynamics of adaptive behaviour and cognition might look like in agent-environment interaction systems. In T. Smithers and A. Moreno, editors, *3rd International Workshop on Artificial Life and Artificial Intelligence, The Role of Dynamics and Representation in Adaptive Behaviour and Cognition*, San Sebastian, Spain, 1994.

[47] L. Steels. Towards a theory of emergent functionality. In Meyer and Wilson [40].

[48] L. Steels. The artificial life roots of artificial intelligence. *Artificial Life Journal*, 1, 1994.

[49] C. Thornton. Brave mobots use representation. Cognitive Science Research Paper 401, School of Cognitive and Computing Sciences, University of Sussex, Brighton, 1995.

[50] T. Van Gelder. The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, submitted.

[51] G. Viana Di Prisco and W.J. Freeman. Odour-related bulbar EEG spatial pattern analysis during