Emergent Representation/Green Cognition

Chris Thornton
Cognitive and Computing Sciences
University of Sussex
Brighton
BN1 9QH
UK

Email: Christopher.Thornton@firenet.uk.com
Tel: (44)1273 678856

May 21, 2003

1 Introduction

Researchers in various fields have begun to exploit the dynamical systems approach to cognition and adaptive behaviour. The approach is still maturing but it is widely perceived as showing great promise in terms of adding a new dimension to the general enterprise of Cognitive Science. The issue of *representation*, which has traditionally been a central concern of cognitivists, is less strongly emphasised in this new approach. In some cases it may even be dismissed as being irrelevant or counter-productive. Brooks, for example, has spoken of the 'representational bottleneck' which is created when cognitive function is mediated though a fully detailed world model, pointing out that in many cases 'explicit representations and models of the world simply get in the way' [1, p. 140].

This new development (see [2] for a closely related argument) puts the onus on the representationalist to state the case for representation more clearly. There is a need to clarify the meaning of the term 'representation' and to show what work it does, if any, in the engineering or explanation of cognitive function. That this cannot be done simply through reference to an existing theory of representation is rather curious given the profligate usage of 'representation talk' by many researchers. But such a theory has not yet emerged despite many years of work and various gestures in the appropriate direction.

The present paper presents a model of representation which, I believe, fulfills both of the requirements listed above. It specifies exactly what 'representation' must mean when the term is used in the cognitive context and it shows precisely the work the concept does for the cognitive engineer and the theoretician.

2 Green Bottlenecks

Brooks' image of the 'representational bottleneck' is a potent one. It conjures up a vision in which relatively primitive sensory processes are forcibly mediated through a fully detailed model of external reality. This mediation must, one feels, impose an extra burden and thus lead to a reduction in overall efficiency. It is only natural to conclude that cognitive processes can function in a more direct and efficient way.[1]

This attitude to representation seems to me rather shortsighted. I would argue that although the imposition of a fully detailed world model on a cognitive process may well be counterproductive, the emergence of detailed world modeling is in fact a necessary consequence for a resource-limited cognitive process operating in a structured environment. The argument, as I now develop it, begins with the meaning of the word 'represent'.

My dictionary defines 'represent' as a transitive verb meaning to 'symbolize, act as embodiment of, stand for, correspond to, be specimen of'. A cognitive agent that makes use of any sort of internal embodiment of any property of an external environment can thus be said to have an internal representation of that property. To some, the word 'representation' may conjure up the idea of baroque, computational structures — frames, scripts and the like. However, it is important to note that there is nothing in the definition of the term which commits it to any particular medium. The representational medium used in a particular case might thus be a static entity, a structure, a dynamic process of some sort or a structure of processes.

What reasons do we have for believing that cognitive agents make use of internal representations of external phenomena? Real(istic) agents implement a wide range of behaviours which are contingent on a wide variety of environmental properties. (Consider, for example, a bee moving slightly to the left to avoid an obstacle, or a bird sheltering under a tree during a rainstorm, or a man laughing in response to a joke.) Where we have an agent which produces a behaviour that is contingent on (i.e., conditionally evoked by) some property of its environment, we know that the agent necessarily implements the contingency for that behaviour (i.e., its 'conditional'). And, by definition, the implemented contingency is then literally the property's embodiment within the agent. Thus all environmental properties which feature within the agent's behavioural contingencies have an internal representation within that agent.

If we are dealing with an agent facing no resource constraints whatsoever we cannot say very much about the way in which these embodiments will link together. However, consider what happens when we allow for the fact that most realistic agents face profound resource constraints and thus can be expected *not* to duplicate internal mechanisms unnecessarily.

Imagine, in particular, that we have two properties P1 and P2 both of which feature in behavioural contingency C1. In addition we have a property P3 which

[1] The impact of the argument can be increased by recounting the well-known anecdote of the 'classical' robot which spends hours blithely computing its next chess move while the room fills with smoke.

is emergent from P1 and P2 and which features in contingency C2.[2] A resource-limited agent can be expected not to duplicate the implementation of P1 and P2 in the implementation of P3 but rather to mirror the emergence relation that exists in reality — letting P3's implementation access the implementations of P1 and P2.

The more resource-limited the agent is, the more we should expect it to exploit such redundancies by 'removing' avoidable duplications. A profoundly resource limited agent, then, can be expected to remove a large proportion of such duplications. In doing so, it effectively recovers a large proportion of the emergence relations that exist between properties of its environment. This process produces internal structures which correspond to external structures and underpin the agent's fabric of behaviours. Thus, profound resource limitations can be the driving force for the construction of what are, in effect, detailed internal models of behaviourally salient 'manifolds' of the external environment.

To illustrate the point, imagine that C2 is the contingency for a 'pushing' behaviour of a mobile agent. Environmental properties P1, P2, and P3 all belong to an object which is assumed to be positioned somewhere close to the agent in the external environment. P1 is the object's 'size' while P2 is its 'flatness'. P1 and P2 feature within the contingencies for different components of the agent's navigation behaviour. P1 is in the contingency for a 'landmarking' reaction (e.g., if the object is big then use it as a landmark) while P2 is in the contingency for a 'walk-over' reaction (if the object is flat then walk over it). The pushing behaviour itself is contingent on the close object being both not flat and small. Since the property of 'not-flatness' is already encapsulated in P2 while the property of smallness is already encapsulated in P1, the agent will save resources by re-using the internal representation of P1 and P2 in the implementation of C1. In doing so it effectively recovers the emergence relation between P1, P2 and the 'pushability' property which is part of the contingency for pushing.

The general implication is that representation is naturally viewed as a *resource-saving* strategy for agents which produce complex (i.e., multiple-contingency) behaviours in structured (i.e., many-leveled) environments. I call representations of this type 'green representations' because of the resource-saving effects they produce. The position which maintains that agents will use green representations is thus 'green representationism' (cf. [3,4,5].)

Note that green representations cannot, in any sense, 'get in the way' as Brooks might have it, since their defining property is the way in which they *clear* the way of redundant processing steps. Note also that green representationism does not make any assumptions about representational media and is thus not in any sense committed to the frames, scripts etc. of classical AI. Nor is it committed to the requirement for 'fully detailed' models of the environment. Nor is it in any sense incompatible with the dynamicist's approach to cognition. It *does* support the expectation that complex agents operating in complex envi-

[2] 'Emergent' is used here in the sense that momentum is an emergent property of mass and velocity.

ronments will necessarily recover behaviourally salient properties and structures of the relevant environment and thus *is* compatible with the belief that agents will engage in some approximation of 'modeling'. However, it is crucial to note that the 'world' which is expected to be modeled (in a particular case) is simply the sum total of behaviourally salient properties and phenomena rather than any imposed picture of reality.

3 Concluding comments

The invitation notice for the workshop posed the question 'is adaptive behaviour in agents a matter of getting the representations right, or is it a matter of getting the dynamics right? ... [Or] should we be expecting to use both concepts of dynamical systems and representation in our investigations of adaptive behaviour and cognition?'. Green representationism provides an emphatic answer. Representations and dynamics are not, in any sense, mutually exclusive. So, yes, we should be expecting to use both sorts of concept in our investigations.

Brooks was, I believe, absolutely right to draw attention to the representational bottleneck. As is well known, the bottleneck is a simple but highly effective device for preventing spillage. The representational variety, it seems, is no exception.

References

[1] Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence, 47* (pp. 139-159).

[2] van Gelder, T. (1992). What might cognition be if not computation?. Research Report 75, Bloomington, IN47405: Cognitive Science, Indiana University (Indiana).

[3] Thornton, C. (1993). Representation as function-hiding. CSRP 282, Cognitive and Computing Sciences, University of Sussex.

[4] Thornton, C. (1994). Tracking the neuro-symbolic continuum: learning by explicitation. Working notes for Symbolic-Connectionist Integration Workshop, ECAI'94, Amsterdam.

[5] Clark, A. and Thornton, C. (1993). Trading spaces: computation, representation and the limits of learning. Cognitive Science Research Paper 291, Brighton BN1 9QH: University of Sussex (Price:1.50).