Frames of mind

John Fox

In memory of Allen Newell

Biological systems such as cells, ecologies and minds/brains, are typically complex and therefore difficult to understand. Their complexity may arise from their static properties (e.g. the huge number, heterogeneity and connectivity of components that are typical of most biological systems) or dynamic features (e.g. non-linearity and non-repeatability of behaviour that makes the identification of simple "laws" notoriously hard). Under these circumstances biological systems often cannot be understood within a single conceptual framework, and theorists typically use multiple explanatory metaphors.

Take the case of "how cells work". The nature of a living cell can be understood using many distinct metaphors, including the following: as a cellular "architecture", built up from many specialised sub-structures and organelles; as a "signalling network" in which lots of little chemical messages whiz from point to point in the cell causing things to happen; as a "bag of chemicals" whose concentrations vary dynamically over time according to quantitative chemical laws, and as a computational system in which DNA sequences are "programs" for building proteins or a set of "rules" that control the cycle of cell growth and division. Different metaphors can be useful for understanding and for teaching, and each can contribute distinctive insights. But each metaphor provides only a partial account, not a unified theory. Disciplines mature when a single unifying framework (metaphor or formalism or narrative) replaces all the partial ones, or at least it is clear how to map from one metaphor to another. (Cell biology is arguably on the way to having a unified theory, though ecology is probably not.)

Cognitive scientists trying to understand "how the mind works" are in a similar situation. Faced with complexity they are generally forced to simplify their research paradigms so that they can make progress. For example they may

1. Simplify the data to be explained (e.g. experimentalists frequently focus on one class of "interesting" experimental phenomena)
2. Simplify the functions to be studied (e.g. reasoning, memory, perception, decision-making, language …)
3. Simplify the explanatory concepts (principles, formalisms, …) they use to obtain a practical trade-off between model complexity and explanatory power.

Furthermore cognitive scientists have probably yet to discover the "right" concepts and languages with which to describe cognitive systems, and formalisms that have the required clarity and expressive power. For anyone who is attempting to understand minds and/or brains the currently available theoretical tools tend to be either expressive but vague (e.g. natural language) or formal and precise but lacking conceptual expressiveness (e.g. classical mathematical functions).

Unfortunately the wish to simplify can be the enemy of scientific realism. Allen Newell, who pioneered the use of computational models in cognitive science and whose ambition was to develop a "unified theory of cognition", criticised the inclination of experimental psychologists to use over-simple experimental paradigms in his famous paper "You can't play 20 questions with nature and win".

In this chapter I wish to take up Newell's aspirations but not his strategy. Newell greatly inspired and influenced me with his conviction that it was possible to understand human as well as artificial intelligence using a computational framework, but in trying to take on the complexities of the human mind he also had to focus on a hugely simplified notion of cognition (as the "firing" of condition-action rules). Computational ideas are extraordinarily varied and versatile for conceptualising, modelling and explaining mental processes, but if the conceptual metaphors are too simple then so will be the theory. If we are to get theoretical purchase on complex natural systems like cells and minds then a single explanatory framework is unlikely to work.

I wish to advocate a "multi-paradigm" approach to this problem, in which we accept that for the foreseeable future no one metaphor will have a monopoly on insight into how the mind works and until a true unified theory of cognition emerges we will need quite a few of them. The goal of this chapter will be to understand the different contributions of a number of metaphors for thinking about mental processes and how we gain additional leverage through several complementary metaphors, in the way that modern biology conceptualises the cell.

Drawing on various traditional perspectives (including philosophy, neuroscience, cognitive psychology, evolutionary psychology, AI, computing and mathematical logic) I shall first look at a few general metaphors of mind and some of their current instantiations. In order to explicate these metaphors I shall review (inevitably rather superficially) a number of frameworks that have influenced the cognitive sciences (as distinct from the rather different ideas about mind in literature, philosophy, folk psychology etc which are interesting in other ways). The main frameworks are likely to include the following.

- Minds as information processing architectures, from classical cognitive psychology (Neisser 1967 on); universal information processors (SOAR, Newell et al), to assembling information processing systems out of computational cognitive modules (Cooper)

- Minds as processes, in which declarative "content" is distinguished from procedural control (e.g. production rules and conflict resolution; deduction and theorem proving; agents and their reactive and deliberative control modes)

- Minds as cognitive agents (e.g. Belief-Desire-Intention agents: Cohen and Levesque; Rao and Georgeff and the Domino agent (Fox);

- Minds as "layered" architectures, including "3 layer agents" (e.g. NASA automata); knowledge based problem-solvers (e.g. Beveridge), human neuro-cognitive organisation (Shallice).

- Minds as meta-cognitive systems including "kinds of minds" (Dennett); complexity of minds (Das, Fox).

- I also hope to make some comments on the currently fashionable view, from evolutionary psychology, of minds as adaptive, evolving capacities that gave primates a competitive advantage.

The main theoretical effort of the chapter will be to try to relate the different frameworks together, to push a knitting needle through the different narratives, as it were, to see how they provide complementary perspectives on a single system. This will not be a unified theory of cognition in Allen Newell's sense, but perhaps a step in that direction.

John Fox, Cancer Resarch UK Version 3, July 2003