

Varieties of cognitive science

Cognitive science is currently the scene of a number of exciting debates. The so-called 'classical' approach, which has dominated cognitive science since the 1950s, is increasingly being challenged on various fronts. Evolutionary psychologists and researchers in artificial life accuse classical cognitive scientists of ignoring the fact that natural cognition is not designed to solve abstract problems and prove theorems but to solve particular adaptive problems. Those working with a 'situated' view of the mind are challenging the classical commitment to internalism. Finally, proponents of dynamical approaches claim that the discrete models favoured by the classical approach are too coarse-grained and impute too much internal structure to the mind.

Because of the challenges they pose to the classical approach, the evolutionary, situated and dynamical approaches may all be referred to as 'non-classical'. One of the main questions I address in this thesis is whether or not these non-classical approaches are compatible with classical cognitive science. I argue that they are, in fact, compatible with the classical approach, with the important proviso that compatibility comes in different kinds. In the final chapter I outline a vision of a comprehensive 'integrated non-classical cognitive science' that combines the three non-classical approaches into a single conceptual bundle.

It is hard to assess these sweeping claims about cognitive science in general without reference to a particular field of research. The emotions constitute one such field, and, moreover, one that is eminently suited to assessing the compatibility of the classical and non-classical approaches. Emotions were ignored by most classical cognitive scientists, and some of the main proponents of the classical approach even went so far as to claim that they were strictly beyond the purview of cognitive science altogether. Later, some models of emotion were developed within the classical framework, but these models provided no way of distinguishing emotion from cognition. I argue that the non-classical approaches remedy this problem, and together provide a new way of thinking about the emotions which I dub 'the interruption theory'. Since the interruption theory borrows insights from all three of the non-classical forms of cognitive science, it serves as a good example of the integrated non-classical approach that I recommend.

What is cognitive science?

Cognitive science is a massive field, grouping together many formerly distinct disciplines, such as artificial intelligence, linguistics and the neurosciences, with branches of philosophy, psychology, and anthropology. Indeed, it may be more appropriate to speak of 'the cognitive sciences' rather than of 'cognitive science' in the singular, since the diversity of theoretical approaches and methodologies that now refer to themselves as 'cognitive' may preclude any view of them as a single discipline. This is not my view. I think that the cognitive sciences have enough in common to warrant speaking of a single entity, 'cognitive science' in the singular, that has a 'classical' form and various 'non-classical' variants.

The two key features that are shared by all forms of cognitive science are:

- (i) the computational theory of mind (CTM): the idea that the mind is a computer; and
- (ii) a design-based approach: the methodological maxim that a good way to understand any natural mind is by designing artificial ones.

I think it is the second clause – the commitment to a design-based methodology – that most clearly distinguishes cognitive science from previous approaches to the study of the mind. The computational theory of mind is certainly important, but the idea of computation is such a loose notion that to define cognitive science on this basis alone would be to risk vacuity. It is not enough to say that the mind is a computer; one must then set out to think how such a computer might be built. When we succeed in building a thinking machine, we will know a lot more about thought. Similarly, when we are able to build machines that can feel happy or sad, we will know a lot more about emotion. This is what I mean by the cognitive approach to understanding emotions.

Varieties of cognitive science

The two core features of cognitive science are shared by all the approaches that I discuss in this thesis. It is only this that allows such different kinds of approach to the study of the mind to be regarded as forms of *cognitive science* in particular rather than merely forms of *psychology* (which I take to be a much more general term). These two core features provide the unity underlying the different approaches.

Within cognitive science, thus defined, I distinguish different approaches based on the stance one takes on particular issues. The evolutionary approach is defined by its emphasis on functional questions; what is the mind, and its various components, *for*? In other words, why did the mind evolve? The situated approach is defined by its rejection of internalism.

And the dynamical approach is defined by its preference for continuous models over discrete-state machines. The classical approach is defined, by default, by its obliviousness to evolutionary-functional questions, its commitment to internalism, and its preference for discrete-state machines.

During the first three decades of cognitive science, from about 1950 to about 1980, the characteristics that I take to define the *classical* approach went largely unchallenged. During that period, few cognitive scientists paid much attention to evolutionary questions, and most were committed to internalism and to modelling the mind with discrete-state machines. Nobody referred to this set of features as embodying a particular *form* of cognitive science. Many assumed that these features were just as essential to cognitive science as the commitment to CTM and to a design-based methodology. It was only when certain sections of the cognitive science community began to question these assumptions that it became clear that cognitive science was not, in fact, essentially committed to ignoring evolutionary questions, to internalism, and to discrete-state machines. Only then were these three features seen as defining a particular *form* of cognitive science, rather than defining cognitive science *per se*. The particular form of cognitive science that was marked by these three features came to be called the 'classical' approach in retrospect, when various dissident groups within the cognitive science community wished to question one of the features without thereby excluding themselves from cognitive science itself.

Pluralism

The identification of particular *forms* of cognitive science had the merit of making clear that the bracketing of evolutionary questions, the commitment to internalism, and the preference for discrete-state machines were all logically independent from the basic idea of CTM and from the choice of a design-based methodology. However, it also had a downside; it fractured cognitive science into a set of warring schools, each of which had a tendency to exaggerate its disagreements with the others. A vision of the underlying unity of cognitive science was lost, and cognitive scientists used the new labels to pigeon-hole one another. To the proponents of the non-classical approaches, it became an easy rhetorical ploy to refer to everyone before 1980 as classical. A typology of approaches became a way of classifying scientists.

Scientists are people, however, and people are much more complex than schools of thought. The latter can be defined in conceptual terms, as I have done with the various forms of cognitive science, but people are not so consistent. It is rare to find a scientist who pursues one kind of approach with single-minded dedication throughout his whole life. Thankfully, most people are more flexible than that. There probably never

was a pure classical cognitive scientist, in the sense of one who never said anything about evolution, nor ever doubted internalism, nor ever wondered about the possibility of modelling the mind in continuous terms.

Nevertheless, flexibility comes in degrees. Even though there may never have been a pure classical cognitive scientist in the sense just described, there have been, and still are, cognitive scientists who are more closely identified with one approach rather than another. One of the aims of this thesis is to persuade cognitive scientists to be more flexible. The integrated non-classical approach I recommend is all about such flexibility.

Alan Turing: the pioneer of integrated cognitive science

In the minds of many cognitive scientists, the name of Alan Turing is associated exclusively with the classical form of cognitive science. Turing's work does not brim with references to evolution; his emphasis on internal memory seems to make him a strong internalist; and the 'universal machine' to which he gave his name was a discrete-state machine. All these features figure strongly in his great paper of 1950, 'Computing machinery and intelligence' (Turing, 1950), which can therefore be taken as inaugurating the discipline of cognitive science in general and its classical form in particular.

Turing's legacy, however, turns out to be much broader than this. When one looks more closely at the 1950 paper, for example, one finds there not just the seeds of the classical approach, but the germs of all the so-called non-classical approaches too. In the section on 'learning machines', for example, Turing proposes a method of designing machines based on an analogy with natural selection, thus anticipating the techniques of artificial life by almost forty years (Turing, 1950: 52). The distinction between memory and processing may be far closer to the situated approach to cognition, with its emphasis on exploiting external resources to ease the burden of computation, than is usually realised (Wells, 1998: 275). Turing's remarks, in section about the importance of giving a computer a body in order for it to have the same experiences as a normal child anticipate current research in robotics (Turing, 1950: 53). And although Turing put his money on digital computers having sufficient resources to pass his test for machine intelligence, he did not argue that non-digital computers did *not* have such resources. Indeed, he clearly states that the human nervous system is 'certainly not a discrete state machine' (Turing, 1950: 47), and argues that digital machines could pass his test only because they are capable of mimicking the behaviour of non-digital systems sufficiently closely. It turns out that Turing anticipated many of the supposedly novel challenges to the classical approach that are generating so much debate in cognitive science today.

Turing's legacy is, then, far richer and more eclectic than is generally believed. It may be more accurate to regard him, not as the founding father of the classical approach, but as the first pioneer of the truly integrated cognitive science that I propose in my final chapter. Turing did not merely leave us with a fascinating thought-experiment about how to test machines for intelligence and a detailed set of proposals about one way (the classical way) to build such intelligent machines; he also left us with a number of provocative complementary suggestions, suggestions that have recently been developed independently by various cognitive scientists who object to the classical approach in one way or another.

The question of how Turing's rich legacy came to be reduced, in the minds of most commentators, to a mere fraction of what it really is, would make an interesting case study in the history of science. Whatever the reasons for this impoverished interpretation, one notes its subtle influence in even the most rigorous scholarship. When Douglas Hofstadter and Daniel Dennett, for example, published Turing's great paper in an anthology of philosophical works about the self, they chose to excise many of the passages I have just referred to as evidence of the richness of Turing's legacy (Hofstadter and Dennett, 1981).

J. A. Scott Kelso notes the same pattern of systematic misrepresentation in an interesting little aside in chapter one of his book, *Dynamic Patterns*. He tells a story about a famous scientist who was always arguing that the brain is not a Turing machine. When Kelso pointed out to the scientist that there was another Turing, the response was adamant: 'No, no, there's only one Turing. You know, the Turing of the Turing machine!' After Kelso wrote some equations on the board describing chemical patterns, the scientist paused and stared at him. 'Ah, I see what you mean,' he said. The equations that Kelso had written on the board were the very ones Turing had used to describe 'the chemical basis of morphogenesis' in another paper that has since become a classic in developmental biology (Turing, 1952; Kelso, 1995).

The anonymous scientist in this revealing anecdote had clearly heard of Turing's other work, for he recognised the reference when Kelso wrote up the diffusion equations that Turing had published in 1952. Yet, until he was reminded of this, the scientist insisted that there was 'only one Turing' – the father of the programmable digital computer. The 'other Turing' – the Turing who showed how patterns in nature can emerge without any programmer at all simply by means of a set of dynamic equations – had been eclipsed by the monolithic image of Turing the classical cognitive scientist.

Even Kelso is somewhat restrictive, however, when he captions his box, 'The two sides of Turing'. In addition to the two Turings that Kelso mentions, there are all the other facets that I have just mentioned, such as the Turing who anticipated recent ideas about the value of giving computers humanlike bodies, and the Turing who foresaw the field of artificial life. And it is not necessary to go to Turing's other papers to find these men. They are all there in the great paper of 1950.

Turing even seems to have anticipated recent ideas about the mind emerging from complexity theory. I do not examine complexity theory in this thesis, so I will conclude this introduction by a brief discussion of its relevance to cognitive science. Complexity theory takes ideas from dynamical systems theory and applies them to systems composed of a wide range of components or of a large number of similar components.¹ One of the key ideas in complexity theory is the idea of supercriticality. Complex systems usually exhibit one or more phase-transitions in which the addition of just a few extra components can produce an abrupt shift in the system's behaviour.

Turing's paper pre-dates the development of complexity theory by at least three decades, yet it contains a thought-provoking idea about the role of supercriticality in mental development. After discussing the phase transition in atomic piles that occurs when they reach critical mass, Turing asks if there is a corresponding phenomenon for minds. He answers in the affirmative:

The majority of (human minds) seem to be 'subcritical', that is, to correspond in this analogy to piles of subcritical size. An idea presented to such a mind will on average give rise to less than one idea in reply. A smallish proportion are supercritical. An idea presented to such a mind may give rise to a whole 'theory' consisting of secondary, tertiary and more remote ideas. Animals' minds seem to be very definitely subcritical. Adhering to this analogy we ask, 'Can a machine be made to be supercritical?'

(Turing, 1950: 51)

Another central idea in complexity theory is that complex adaptive systems tend to hover around the critical points, rather than living deep in

¹ There is no widely accepted definition of complexity theory. Some authors seem to treat it as a synonym of dynamical systems theory, or nonlinear dynamics, while others state explicitly that 'chaos is not complexity' (Bak, 1996). The central plank of the theory, in my view, is the idea that special mathematical tools are required for understanding systems in which the high number of components precludes any attempt to derive systemic properties directly from component properties. Care must also be taken to distinguish this theory from the branch of computational theory that is concerned with the complexity of certain mathematical functions, for this too goes by the name of complexity theory (Andy Wells, personal communication).

the subcritical or the supercritical regions of their phase-spaces. In Stuart Kauffman's evocative terminology, complex adaptive systems tend to be poised 'at the edge of chaos', where behaviour is neither entirely ordered nor completely chaotic (Kauffman, 1995: 86-92). Kelso is somewhat more prosaic and prefers to speak of 'the intermittency mechanism', though his meaning is essentially the same (Kelso, 1995: 99). The following passage, again taken from Turing's 1950 paper, seems to anticipate just this idea:

Intelligent behaviour presumably consists in a departure from the completely disciplined behaviour involved in computation, but a rather slight one, which does not give rise to random behaviour, or to pointless repetitive loops.

(Turing, 1950: 55)

With hindsight and close-reading, then, Turing's prescience can be seen to extend even into the most recent thinking in cognitive science. Yet this does not match the common idea of Turing today as the founding father of the classical approach.

The impoverishment of Turing's legacy reflects a general narrowing of perspective that has marred cognitive science for much of its history. In this history, something has occurred akin to what Steven Jay Gould has called 'the hardening of the modern synthesis' in evolutionary biology (Gould, 1983). Varela, Thompson and Rosch argue that the cybernetics movement, which directly preceded the emergence of cognitive science proper, was characterised by a much more pluralistic approach to cognition – one that, for example, still regarded the digital nature of computation as an open question rather than an accepted dogma (Varela, Thompson et al., 1991: 37-39). If Varela *et. al.* are correct in their historical account, my arguments in chapter six about the need for a similar pluralism in cognitive science can be seen as a plea for cognitive science to return to its eclectic roots in what we might call the 'pre-classical' era of cognitive science in the 1940s, when the design-based approach was initiated by disciplines with other names such as 'control theory', 'information theory' and 'cybernetics'. The various species of cognitive science I discuss in this thesis – both classical and non-classical – can then be seen, not as antagonists, but as pieces in a complex jigsaw, all of which are necessary if we are to get the whole picture about mental phenomena. In the second part of chapter six, I go on to argue that this eclecticism is especially vital when it comes to getting the whole picture about emotion – although Turing, it must be said, was silent about this important part of our mental life.