Review of *Human and Machine Thinking* by Philip N. Johnson-Laird. Lawrence Erlbaum. 1993. £19.95. 189 pp Indexes.

Reviewed for THES by Dr Susan J. Blackmore, Senior Lecturer in Psychology, University of the West of England, Bristol.

"We have never fully understood a process" declares Johnson-Laird, "until we can build a machine to carry it out" (xv). This is what this book is all about, only the processes are rather tricky ones and the building of the machine is proving harder than anyone first thought. The processes are precisely those that seem to make us human - perception, thinking, language and consciousness. The machine is a computer.

This connection between understanding and mimicking underlies the whole field of artificial intelligence, in which Johnson-Laird has been influential for many years. Some researchers want to understand the human mind while others want to build the machines. Either way, the fundamental question of AI is can machines think?

The classic way to tackle it has long been Alan Turing's famous "Turing test", which a computer program can pass by being mistaken for a human. But Johnson-Laird says it is time to retire the Turing test. The question is not whether computers can think - obviously they can and sometimes they do it very crudely and stupidly. The critical question is rather "Can a machine think in the same ways that a human being does?" (xvi). It is not good enough to make a crude simulation of thinking. What is needed is the real thing - a computer that actually thinks like we do. An analogy might be made with a weather forecasting computer. It might do very well at predicting and describing storms and gales but you would not find a cold wet wind howling inside the box. We do not need to mimic human thinking but embody it.

To do this, Johnson-Laird argues, we need a theory of human thinking. Between the extremes of human calculation and free association of ideas, the "clocks and clouds of the mind", lie the more important processes of deduction, induction and creation.

Deduction is something that computers seem to do much better than humans, at least if you think of it as a purely logical process. Johnson-Laird, however, is not concerned with making the most logical deduction machine but with how humans actually deduce. He makes a strong case that we do it not by formal logic or syntax but by constructing mental models. This is revealed in the many kinds of errors we make in everyday life. For example, we search for conclusions compatible with our mental models and when we find one we stop looking - with potentially catastrophic results if there are other conclusions we have overlooked. Some kinds of deduction take humans longer than others. A purely formal theory of thinking cannot explain this but the mental model theory specifies how the models are built up, how many are needed for each kind of deduction, and therefore how long the process will take. Advantages of this way of reasoning are that it is simple to check models, common sense knowledge can be used in making the models and they can be updated as new information comes along.

Of course most human thinking is not deduction because we are more interested in truth and usefulness than validity. Since we often don't have enough information we have to use induction,

going beyond the facts available. Induction adds semantic information but still uses similar concepts. The interesting difference comes in creativity in which, according to Johnson-Laird, the underlying concepts themselves are changed. The information may be overlapping with what came before and the concepts incommensurable with previous ones. In creation we restructure the concepts used to understand things in novel and non-deterministic ways.

His analysis of creation, from musical innovation to scientific creativity, provides interesting insights. The modular structure of our mental skills shows why we are so much better at criticising our creations than generating them. Creation cannot be entirely Lamarckian nor entirely Darwinian but has to be a mixture. We cannot learn creativity in general but only in specific domains. But I wondered why creativity cannot be deterministic and whether all the beautiful creations of chaotic processes are excluded.

For all three kinds of thinking the key argument is that we use mental models and not formal rules. But what are they? Apparently mental models use tokens to represent each entity and relations between tokens to represent relations between entities. I am not much the wiser. This book does not explore how the world comes to be divided into entities or what these tokens might be. Johnson-Laird claims that at the moment machines only *simulate* semantics whereas a machine using mental models would be computing semantically, not just syntactically; that is with meaning and not just structure. So how would machines get to this point?

To understand *meaning*, he claims, machines would have to have as rich connections with the world as we do. Here I agree. Yet our best hope of understanding these connections may lie in the development of neural nets that can make their own concepts and divide the world up according to their experience rather than by rules a programmer might give them. But Johnson-Laird barely mentions neural nets or the revolution they have brought about in cognitive science. Until we understand how machines can build their own concepts I do not think we can come back to the really interesting questions Johnson-Laird hints at – would a computer instantiating his mental models feel the way we do? Would it be conscious? Could it have a moral, aesthetic or spiritual life? He believes they would but I am not sure he has yet provided a theory of thinking that explains how or why.