Meta-level Inference and Consciousness

META-LEVEL INFERENCE AND CONSCIOUSNESS

Alan Bundy

D.A.I. RESEARCH PAPER NO. 187

To be published in *Proc. of the Second Anglo-French Philosophy Colloquium.*

1983

META-LEVEL INFERENCE AND CONSCIOUSNESS

Alan Bundy
25 February 1983

Of all the mental phenomena, exhibited by the human mind, none seems to offer more of a challenge to AI modelling than **consciousness**. So much so that it is rarely discussed (I would say unconsciously avoided) in AI. This paper addresses the question "Can AI shed light on the phenomena of consciousness?". The question can be fruitfully divided into two subquestions.

(a) Can AI modelling help us to understand the role of consciousness in cognition?

(b) Could we build a computer program which experienced the sensation of consciousness?

My personal answers to these questions are:

(a) Almost certainly yes, in fact, progress has already been made, and this paper is a contribution to that debate.

(b) This is the classic problem of 'other minds' in a new guise. We might answer answer (a) to most peoples satisfaction, and this might enable us to build a machine which behaves as if it were consciousness. Whether we will then be happy to attribute consciousness to it, and whether we will better understand the subjective experience of consciousness are open questions. We will not address this question further in this paper.

One piece of AI evidence which we will bring to the study of consciousness is some work on the guiding of mathematical proofs using a technique called **meta-level inference**, which has been developed by the author and his co-workers. The AI problem, for which meta-level inference is an attempted solution, is the **combinatorial explosion**. This arises in the modelling of mathematical reasoning when several rules can be applied to a goal to produce subgoals. Several rules then apply to each of these subgoals, and the number f subgoals rises exponentially (or worse) with the depth of the proof. Even modern electronic computers cannot cope with the storage overload that this rapid growth produces, with the consequence that only simple theorems can be proved automatically by exhaustive search.

In order to prove more interesting theorems, it is necessary to guide the search for a proof, so that the only rules applied to goals are those with a high probability of being in the proof. In meta-level inference this guidance

information is formalised as an axiomatic meta-theory. This meta-theory describes the representation of the original (or object-level) theory. Inference in the meta-theory induces an implicit inference process in the object-level theory. This technique has been used by the author and his co-workers in the domains of equation solving [Bundy and Welham 81], program verfification [Bundy and Sterling 81] and physics [Bundy, Byrd and Mellish 82]. We will investigate the relation between meta-level inference and consciousness.

In folk psychology, the terms: conscious, self-aware and introspection are all intertwined. One consequence of the discussion below will be to suggest that these concepts should be disentangled. In order to infer the contradiction which will force this disentanglement, let us identify consciousness and self-awareness, and regard introspection as the mechanism of both.

But if introspection is the mechanism of consciousness then, far from being a hard problem, modelling consciousness should be one of the easiest. The obstacle to progress in most areas of AI is that the processes we are trying to model are carried out at an unconscious level and are therefore unavailable to study by introspection. This explains why we have made such good relative progress in, say, chess compared with, say, vision - the processes involved in playing chess are much more accessible. But consciousness is, **by definition** available to introspection. It should be easy for us to say what consciousness consists of. Why has this not been done already?

Several people [Sloman 78, Minsky 68] in the AI literature have suggested that one role of consciousness is in solving the combinatorial explosion. This arises, not only in mathematically theorem proving, but whenever a task involves non-trivial inference. The argument is that control of search demands self-awareness of the goals and rules of the problem space and how these fit together, i.e. meta-level knowledge. Introspection is a meta-level inference process where the conscious agent infers which means to apply to achieve particular ends.

At first sight this seems to fit very well with my research, described above, where control of search is effected by meta-level inference. But this should have made my research considerably easier than it in fact was. The meta-level inference steps were not discovered simply by introspecting while solving equations, etc. In fact, considerable study, hypothesising, rationalization, and experimentation went into the design of the programs. It was a trial and error process.

So if introspection does not consist of meta-level inference, what does it consist of? We can all conduct the experiment of introspecting while thinking, to see what seems to be going on. The results of my own experiments are:

- It is much harder than I at first thought, due to the interference between thinking and thinking about thinking.

- Much, but not all, of the process is conducted in English. I often imagine myself arguing with some fictitious person or composing an essay.

- Sometimes I catch myself thinking of a concept, but not having the English word for it. So not all my conscious processes are in English. Some consist of visual concepts (imagery).

- Only rarely do my conscious processes consist of making control (search) decisions, i.e. only when such decisions are particularly tough. Normally, they consist of 'output' rather than 'traces', 'proofs' rather than 'searches'.

- Using the ideas of my research projects, I would say that introspective processes are at the **object** rather than **meta** level, except when a control decision is particularly tough, when the whole show shifts up a level, the meta-level becomes conscious but the meta-meta-level which decides the issue is still hidden.

An analogy for the relationship between conscious and unconscious processing, which is often offered from Computer Science, is that between compiled and interpreted code. Like unconscious processes, compiled code runs faster, but is more difficult to trace than interpreted code. Programmers often use interpreted code, when developing a program, because it is easier to trace, debug and modify. They then compile it when it is running satisfactorially. Similarly, when we are learning a new skill (riding a bike, walking etc.) we are conscious of every move we make, we can easily modify what we are doing, try out new ideas etc., but we are very slow. When we are satisfied with our performance, the procedure becomes unconscious (automatic, a habit) and much faster and effortless. However, if our procedure contains bugs they are now extremely difficult to remove (breaking bad habits). The procedure must be brought back to consciousness, relearnt and constantly practiced. We seem to have difficulty remembering to use the new conscious procedure, instead of the old unconscious one and so must pay constant attention if we are not to fall back into old habits.

Note that Computer Science provides us with a whole range of different levels of compilation, from 'hardwiring' at one extreme to Sussman's 'careful mode' at the other. We may well want to make use of this diversity to enrich the conscious/unconscious dichotomy to account for subconscious, hardwired and other levels of mental process.

The implication of this analogy for control decisions is that for the most part they are made by procedures long learnt and compiled away. There must be some general purpose ones for dealing with new situations and they must often be wrong. The reasons why we do the things we do will usually be obscure and

questionable. Learning to do things differently will be fairly traumatic, but possible, so there is a mechanism for making some of the unconscious decision-making processes to the fore, even if we need help with the more deep seated ones.

Note that the degree of compiledness of the meta-level can be different from that of the object-level. Thus we can pay attention to the correctness rules of algebra without paying attention to the approriateness of their application, and vice versa.

What conclusion can we draw from the above? The analogy of interpreted versus compiled computation seems a fruitful one to apply to the distinction between conscious and unconscious processing. The analogy of meta-level inference seems a fruitful one to apply to self-awareness, but not to introspection. It seems we must tease apart the concepts of self-awareness and introspection. The analogy of interpreted/compiled code suggests a refinement of the dichotomy between conscious and unconscious processing to include several levels. Overall the analogies from AI/CS suggest a reworking and refinement of the simple terminology that we inherit from folk psychology.

REFERENCES

[Bundy and Sterling 81]
Bundy, A. and Sterling L.S.
Meta-level Inference in Algebra.
Research Paper 164, Dept. of Artificial Intelligence, Edinburgh, September, 1981.
Presented at the workshop on logic programming for intelligent systems, Los Angeles, 1981.

[Bundy and Welham 81]
Bundy, A. and Welham, B.
Using meta-level inference for selective application of multiple rewrite rules in algebraic manipulation.
Artificial Intelligence 16(2):pp189-212, 1981.

[Bundy, Byrd and Mellish 82]
Bundy, A., Byrd, L. and Mellish, C.
Special Purpose, but Domain Independent Inference Mechanisms.
In Proceedings of ECAI-82, pages 67-74. European Conference on Artificial Intelligence, 1982.
Also available from Edinburgh as DAI Research Paper No. 179.

[Minsky 68] Minsky, M.L.
Matter, Mind and Models.
In Minsky, M. (editor), Semantic information processing, . MIT Press, 1968.

[Sloman 78] Sloman, A.
The Computer Revolution in Phiosophy: Philosophy, Science and Models of Mind.
Harvester Press, 978.