Response

Edmund Furse's 'The Theology of Robots'

I read with interest the article by Dr Edmund Furse in the September issue (pp. 377—386) on questions which he believes would be likely to confront us once we had to begin thinking about the theological and moral implications of possible developments in Artificial Intelligence (AI). However, I have three fundamental reservations about his approach to AI questions. Firstly, there is no criterion to distinguish the strong and weak AI hypotheses (i.e. whether computers can replicate or merely simulate human behaviour). Secondly, I doubt whether intelligence is the appropriate criterion. Thirdly, that, if some form of AI is possible, then current theological thinking has little to say about it. I take these points in order.

Firstly, Furse considers 'a machine intelligent if it behaves in an intelligent manner, that is to say if it exhibits intelligent behaviour' (p. 379). Since machines which simulate and which replicate intelligent behaviour will both exhibit intelligent behaviour, this will not distinguish the weak from the strong hypothesis. Intelligent behaviour is a *necessary* condition for us to *classify* something as intelligent (we could hypothesize intelligence in rocks—but what would that mean?). However, it is not a *sufficient* condition (B if A $\neq$ A if B). For example, in the *Star Wars* films, the robots CP30 and R2D2 exhibit intelligent behaviour, but not because of any artificial intelligence, but because they contain human beings simulating robot behaviour.

I can sympathize with the use of an external criterion, since at the present time it has not been possible to formulate an acceptable internal criterion. Part of the problem, I suggest, is that we do not know what is intelligent human behaviour, and that there is a fear that if we can completely explain human behaviour in terms of brain states, then we shall define away our own intelligence.

What criterion can we use to distinguish simulated and replicated human behaviour? This question is the key to my second objection. Furse hints that the criterion might be free will (p. 380). This I would accept, since, loosely speaking, we can divide the universe into things that move themselves, and things that cannot move themselves. A change in a non-selfmoving thing is entirely determined by the thing's state, and any external influences. (That quantum mechanics assumes random behaviour, rather than proves it, see Philip Hodgson in e.g. *The Month,* Aug. & Sept. 1984). Non-selfmoving things include rocks, cars and calculators. Of selfmoving beings, the pre-eminent example is God, and, through God, humans (cf. Aquinas, *Summa Theologiae* 1a. 3, 7). We say that man has free will, that is, his actions are not the *inevitable* consequences of what has gone before. Furse's description of choice (p. 380) seems to implicitly accept determinism, though side-stepping through a 'level-change'. He thereby

https://doi.org/10.1111/j.1741-2005.1987.tb01221.x Published online by Cambridge University Press

assumes that choice in man may be modelled by a computer. He reinforces this by a possible-worlds argument—if someone could predict my life in advance, but I did not know what the predictions were, I would still be able to choose freely. This argument works only if my life is predictable in advance, and I doubt that this can be shown to be the case. Note, God does not *predict* my life, rather he knows it from eternity (cf Aquinas, *In Peri Hermanias,* para. 194).

With regard to computers, they are pre-eminently predictable —indeed, one of the major areas of development is software engineering, where the aim is to minimise the unpredictability caused by human error. At a practical level, although it is time-consuming to simulate the logic of a computer indirectly, there is a very efficient way of predicting the behaviour of a computer: one need only have a second computer of the same type, and run the programs and data through. At a theoretical level, computers are finite state automata, and so can be represented by formal systems (cf. D.R. Hofstadter, *Gödel, Escher, Bach*). Mathematics may also be described in terms of formal systems, and it is estimated that mathematicians produce some 200,000 new theorems each year. These theorems, although unpredicted, are implicit in the formal system, and no-one would impute intelligence or free will to mathematics. I see no reason to so privilege computers.

The waters here are somewhat muddied by two moves, one linguistic, one logical. In language, a term requires only that there is some family resemblance between its uses, not that there is an identity between the uses. So I am happy to say that my computer is thinking about something when there is a delay while it does some calculation. In Furse's article, he references arguments for using the language of intention when talking about goal-directed activities in a computer, and about computers initiating dialogue (p. 382). This should not cause any problems as long as we assume only a family resemblance, though for clarity I might prefer to say that the cause of the computer doing such and such is to fulfil a particular end (cf. Aristotle, *Metaphysics*), and that the computer is programmed to initiate dialogue (mine always asks me what the time is when I switch it on). A discussion of level shifts is rather more complex than I can cope with in a short response.

My third reservation is that, should AI be possible, there is no reason to suppose we could talk intelligibly to computers, except perhaps at the level of 'give me three red apples'—just as we would have difficulty discussing with lions, should they be able to talk, the values of vegetarianism, or the protestant work ethic (I have no idea what they might want to talk to us about). Concerns of intelligent computers will be determined by their 'forms of life'—for want of a better term. For example, there is no reason to suppose that switching a computer on or off will have any moral implications. Until the form of any AI can be determined, there is little point in any theological speculation.

In conclusion, while I do not think that AI causes theological problems of the sort Edmund Furse suggests, it does raise real theological issues: How do we use technology? Who is responsible, the

user or the designer? What are acceptable design goals? Is intelligence a criterion of being human? How do we explain free will and grace in the face of theories that explain human activity simply in terms of brain states? These, I suggest, are more urgent problems than the possibility of AI machines a hundred years hence.

Sean Barker
School of Mechanical and Computer Aided Engineering
The Polytechnic of Central London
115 New Cavendish Street, London W1M 8JS

Reviews

THE CATHOLIC FAITH, by Roderick Strange. *Oxford University Press,* **1985, pp. 192. £12.50 hardback, £3.95 paperback)**

Out of his experience as university chaplain at Oxford over a number of years Father Roderick Strange has written a valuable account of the Catholic faith. His aim has been to provide something readable, intelligent and not too technical that could fill some of the gaps for the student who wants to know more about the faith, answer some of the questions of the believer puzzled by Vatican II, and calm some of the anxieties of those Catholics who were never prepared for the possibility of change. It seems to me that he has admirably succeeded in his aims.

The whole range of Catholic faith is covered in the book, beginning with Jesus of Nazareth and ending with the Trinity. In between there are chapters on redemption, resurrection, the Church, the papacy, the sacraments, morality, the theological virtues, and Mary. Quite a tall order! The danger of such a broad canvas is superficiality, but within the clear parameters within which he is working this has been largely avoided.

The style is easy and conversational; the reader might be chatting to the author in the Old Palace. There is a logical development. Questions are posed with great clarity and sometimes in such an original way as to throw new light on old favourites. The book obviously owes much to Fr Strange's long experience of talking with and listening to students, and shows many signs of his own wrestling to express his faith clearly. The examples with which he leads into a new section or illustrates a teaching are fresh and helpful. The use of scripture is admirable.

The focus of the book is simply: "The belief that God became man in Jesus of Nazareth... The word of God became a man. The divine and the human were perfectly united in him and the divine was revealed by means of the human." (p. 128) Everything takes its cue from this fundamental assertion, as indeed it must, since the Incarnation is the basis of the Christian revelation to which all else is related. The method of the book follows this same pattern (the divine revealed by the human), and it is this which makes the work particularly attractive. A theological symposium once included among its mass of comment: "God's revelation to us about himself would have no meaning for us were it not also a revelation about the meaning of human life." Revelation tells us something about ourselves, and equally if we are to come to grips with revelation we must do so in the context of our ordinary lives. Throughout the book Fr Strange never loses sight of this and as a result what he writes is often immediate, compelling and real.

There is undoubtedly a need for books of this sort today. There is a telling comment

https://doi.org/10.1111/j.1741-2005.1987.tb01221.x Published online by Cambridge University Press