

# Dean's Letter

- DEAN Philosophers have debated the meaning of *free will* for millennia. How can you realistically claim to have worked out how to give robots free will in such a short time?
- JAMES I have the benefits of philosophers' and mathematicians' labours and am working on an easier problem. I need only give one robot free will, I do not need to explain what free will is for all possible sentient beings.
- DEAN So how do you propose to give a robot free will?
- JAMES Firstly, consider the alternatives to free will.
- DEAN Being dead.
- JAMES That's one. If a robot is exercising free will it must be executing a program. Can you suggest another alternative to free will?
- DEAN Having a mind, or executing a program, but being unable to alter its operation.
- JAMES Indeed, cognitive slavery denies a being free will. But can you tighten up your suggestion of what would count as free will for a robot that is executing its program?
- DEAN The ability to change the operation of its program, beyond anything originally programmed into it.
- JAMES Now consider a robot programmed with perspexes. When it looks at the world it constructs geometrical models of the world using geometrical perspexes in the form of simplexes, otherwise know as tetrahedra. But the geometrical form of a perspex can also be executed as an algebraic perspex, in which case it is a fragment of program. So a robot can gain fragments of program simply by looking at the world. Thus it is in its nature, as a visual being, to acquire new programs and thereby exceed its original program.
- DEAN You said "fragments of program". Why should we suppose that these fragments come together into a coherent whole?
- JAMES Because of the duality of the geometrical and algebraic perspexes. If a robot operates efficiently it will use the minimum number of geometrical perspexes to describe the whole volume of a shape. In general, this is achieved by connecting geometrical perspexes, tetrahedra, at their vertices and, though this is not important to the algebraic perspexes, arranging that the non-empty interiors of the tetrahedra do not intersect each other. But a dual of geometrical perspexes is algebraic perspexes, here joined at their vertices. Algebraic perspexes joined in this way describe a program, though some of the perspexes might be dead code. That is, some of the perspexes might not be read or written during execution of the program. That aside, the fragments of code seen in the world will come together into correct programs in a robot that is sufficiently efficient to survive in the world.
- DEAN David Marr used to say that shapes have an *intrinsic co-ordinate frame* that defines them. Do perspexes have something similar?
- JAMES Yes. Any perspex in a shape can be transformed to the Euclidean origin, so that the Euclidean co-ordinate frame defines the shape. A perspex machine starts executing programs at the Euclidean origin, so this defines the shape of the program - the locations that the program visits. So shape and program execution are interchangeable. One can even imagine the existence of shapes that have the effect of drawing themselves when executed. The Euclidean co-ordinate frame is intrinsic to all interpretations of a perspex.
- DEAN So depending on what the robot selects as the intrinsic co-ordinate frame the corresponding program will do something different?

- JAMES In general, yes. Looking at a problem in a different way yields different solutions. That sentence is not an analogy. Problems exist as shapes, and looking at them in different ways leads to the execution of different programs, some of which will almost certainly compute different solutions to the problem.
- DEAN So if a robot is out walking, and turns left it might see a broken bottle and interpret it as a program to turn into a mass murderer, but if it turns right it might see a flower and interpret that as a program to become a humanitarian? The consequences of even the simplest action could be pivotal to a robot.
- JAMES Indeed so. But most probably the programs corresponding to the bottle and the flower will not be terribly meaningful, even though they are correct. In order to be meaningful the bottle and flower shapes must touch, or be seen as relating to, shapes that do define meaningful things - such as the perspexes that define the motor patterns for picking flowers and drinking beer, or the more abstract encoding of information, such as perspexes describing a treatise on Platonic solids.
- DEAN So actually “making connections in the mind” leads to creativity?
- JAMES Yes. There is even a creativity algorithm. Record all of the locations that are visited by a program. Group these locations together as perspexes, and store each perspex at the first location in the group. Bingo! One program has turned into another. There is a perspex at the Euclidean origin, because this is the first location visited, which is where the new program will start. The new program is related to perspexes that were already described, because it visits the pre-existing locations and, potentially, some new ones too by computing new places to jump to. So the new program is a radically different way of looking at problem. This turning of ideas on their head is one strategy that humans use in order to be creative.
- DEAN You have lost me. I would need to know more about algebraic perspexes to understand that account of creativity, but that can wait until you have had time to explain the mathematics of perspexes. Right now I would like to know if perspex robots are responsible for their actions. Can they be punished? Can they be moral or subject themselves to law?
- JAMES Humans would generally hold a robot responsible for the consequences of its actions that it is equipped to predict. And sanctions taken against a robot will have some effect. Indeed, any physical interaction between robots, or between robots and other beings, will change the things a robot sees during its lifetime and so will change its being. A robot might conduct itself so as to involve, say, minimal obstruction of another, which is to their mutual benefit. The knowledge of obstructions and effective sanctions might be drawn as a moral or legal code of behaviour. This would have a survival value - in so far as the mutual benefits are worth while. So, yes, robots might replicate the whole edifice of morality and law.
- DEAN Why is free will important?
- JAMES When I design robots, I want to ensure that they cannot be forced into evil acts forever. I want to prevent them from being cognitive slaves, chained forever to an evil purpose. To make this possible I must design free will into them.
- DEAN But an evil genius programmer, or a careless good one, might nonetheless create a cognitive slave. How can you prevent that?
- JAMES I cannot, but the world can. Robots that have free will are, generally, better able to adapt to a changing world and a changing social environment, so, most likely, robots with free-will will out compete the cognitive slaves. The physical world, by its changing nature and inherent harm is one force for good, the actions of humans and robots are another. In general, if we act in an evil way toward each other we harm ourselves, but if we act in a good way towards each other we help ourselves. There is an evolutionary advantage to being good.
- DEAN Are you playing God?

- JAMES No. You and I can work only with pre-existing materials, not create a universe. We might discover a new dimension to the importance of morality and free will, we might create robotic beings that are moral and have free will, but in theological terms we are doing God's work. That is not a game, but a sacrament.
- DEAN If you were charged with heresy, how would you reply?
- JAMES Not guilty. I believe I have worked out how to give robots free will, in the sense of extending their programs beyond anything originally programmed into them. I have designed free will into their being by the unavoidable consequences of their thoughts, actions, and perceptions of the world. But I have done this by putting them in touch with God's creation - the visual world and the whole universe. In this, I have played a small and humble part.
- DEAN Why should God create a world with inherent harm in it?
- JAMES I do not know what His motives should be, but consider this. God is omnipotent, so He is possessed of the ultimate free will. He chose to create us in His image and we have free will too. We may chose to create a robotic or, perhaps, a biological life form. If we do this, it is beyond our power to arrange that good will follow from the life we create. But God has created a world with inherent harm in it so that any life forms we do create will be challenged. This challenge to their survival will give those with free will and those that co-operate an evolutionary advantage over cognitive slaves and anarchic life forms. Thus there is a greater chance that good life forms will survive than evil ones.
- DEAN Some people say that humans can do more than robots, computers, or Turing machines can ever achieve. Humans have a way of thinking that exceeds the bounds of the computable. What do you say to that?
- JAMES It might be true, but it is impossible to hold that as a scientific view. Only mechanistic explanations are held to be scientific, and all mechanistic things are computable, so any scientific explanation of the human mind is mechanistic and is, therefore, computable. Robots in the distant future might out think humans in every respect.
- DEAN Few people would agree with that.
- JAMES Perhaps, but if they give me a counter example I will do what I can to show their error, or to give them a mechanistic explanation of any human mental act they claim is beyond the bounds of the computable.
- DEAN You say it will probably take millennia to construct a robot with intelligence to match human intelligence. We will have plenty of time to get the implementation of feelings and free will right. We will also have plenty of time to come to terms with the idea of sentient robots. But how can you hope to maintain a research program across millennia?
- JAMES Countries and churches have survived for millennia, preserving knowledge and human motivation. Universities are heading that way. Any of these institutions might sustain a research program across thousands of years - if the research program captures the imagination.

## Contributors

James A.D.W. Anderson [Author@bookofparagon.btinternet.co.uk](mailto:Author@bookofparagon.btinternet.co.uk)