

# JOHN LUCAS

John Randolph Lucas

18 June 1929 – 5 April 2020

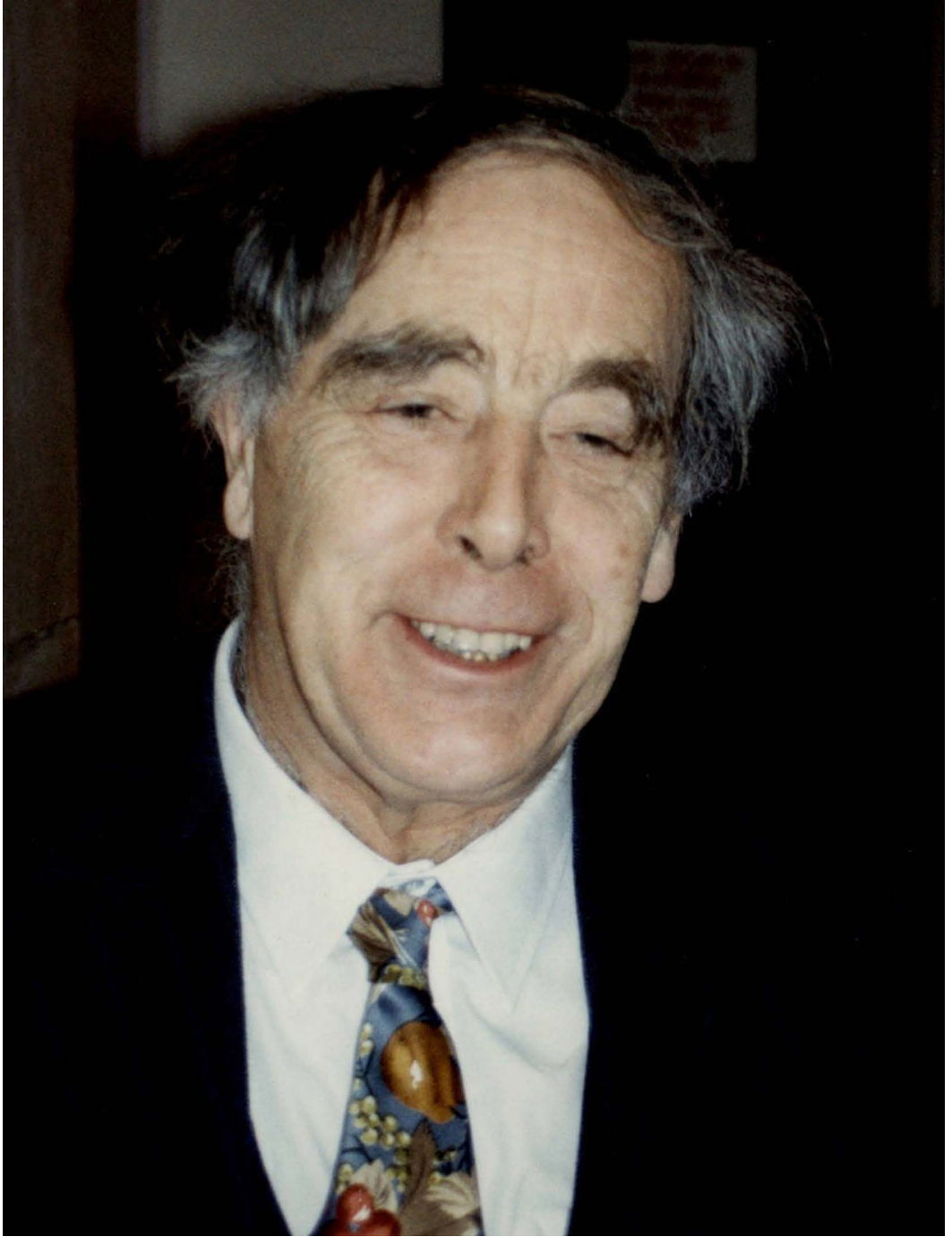
elected Fellow of the British Academy 1988

by

RICHARD SWINBURNE

*Fellow of the Academy*

John Lucas wrote many books covering many different areas of philosophy. His contributions were often challenging and controversial, opposing contemporary trends in philosophy (both the logical positivist and ‘ordinary language’ philosophies current when he began to philosophise, and some of the metaphysical theories subsequently current), and opposing contemporary Western secular morality, but usually tentatively and with sympathy for rival views. He wrote much on space and time, on moral responsibility, and on political philosophy; and is well known for his firm advocacy of human free will, based on his argument that Gödel’s Theorem has the consequence that human behaviour is not subject to deterministic laws. He also wrote many incisive and amusing essays on academic politics.



J.R. Lucas

## Life

John Randolph Lucas was born in London on 18 June 1929, into a family of Church of England clergy. His father was a vicar until he became an Archdeacon, and his maternal grandfather was a Bishop; his younger brother, Paul, subsequently became a vicar. His parents, following the practice of many educated but not especially well-off English parents seeking what they regarded as the best education for their children, sent him from an early age to a 'preparatory school' as a boarder, and then from the age of 13 to a 'public school' (that is, in English as understood outside UK, a private school). Lucas's public school was Winchester College, one of the seven or eight most distinguished public schools, and the most academically orientated of all of them. It was very expensive to send a child to a good public school, and Lucas obtained an entrance scholarship which made this possible. He did not enjoy boarding school life, but was happier at Winchester than at his preparatory school. In the last two or three years at school, students specialised in one fairly narrow area of study; and he specialised in science rather than in languages or history. But he gradually became interested in the foundations of physics and mathematics, and so in philosophy. He obtained a scholarship to Oxford University, and began his life as an undergraduate at Balliol College, Oxford in 1947. He studied mathematics for the examination which all students had to take after their first few terms, but was then allowed to change his course of study to read 'Greats'. Greats involves studying the history of ancient Greece and Rome, and philosophy, including a substantial element of ancient Greek philosophy (principally Plato and Aristotle). As texts were studied in Greek and Latin, and although he had learnt a significant amount of Greek at his preparatory school, he had to spend an extra two terms at Oxford learning Greek before beginning the main course. The central element of undergraduate teaching at Oxford was then (as it is now) the weekly (or sometimes twice a week) tutorial, individually or in pairs with a tutor who is a senior academic. Oxford philosophy in the late 1940s and early 1950s, was much influenced by logical positivism. But for Lucas (as he puts it in a paper about his early life) 'the tenets of contemporary philosophy seemed just too silly to be worth bothering about.' His principal philosophy tutor was Richard ('Dick') Hare, who was highly sympathetic to logical positivism as an account of the meaning of 'factual' statements, and gave an account of the meaning of moral assertions as 'non-factual' but rather 'prescriptive'. But Lucas was looking to philosophy for eternal truths, and so his tutorials with his tutors were battles. He was fortunate to have Bernard Williams as a fellow student at the same stage as himself and claims that the two of them 'used to take pleasure in concerted campaigns to confute our tutors' cherished arguments and force on them a change of mind.' After getting his BA with first-class honours in 1951, he went on to do graduate work at Oxford on a senior scholarship at Merton College, where he then became a Junior Research Fellow. He spent the years 1956 to

1960 in various places other than Oxford – two years as a fellow of Corpus Christi College, Cambridge, a year as a research fellow at University of Leeds, and a year as a Proctor Fellow at Princeton. He returned in 1960 to Merton College, Oxford, where he became a permanent Fellow as well as a permanent Oxford faculty member for the rest of his academic life. John Lucas was very much, and very pleased to be thought as, an ‘Oxford man’.

He was also very much a ‘family man’. In 1961 soon after becoming a permanent Fellow of Merton College, he married Morar Portal. They had four children – Edward, who became a well-known journalist; Helen, who became a GP; Richard, who became an entrepreneur; and Deborah. Deborah was born with Down’s syndrome, and John and Morar cared for her at home with great love. From 1976 they lived in term time in Merton Street, within a few yards of the gate of Merton College. In vacation time, John, Morar, and their family retreated to their home in the West Country (at one period a house near Exeter, at a later period one in Somerset), where John wrote his philosophy. And when John retired in 1996 from his post at Merton College, John, Morar, and Deborah lived permanently in Somerset.

In Oxford John Lucas was very fully involved not merely in his teaching, which he did with diligent enthusiasm and with great care for the welfare of his pupils, but in the political life of the college and university, and more widely in defence of the environment and consumer rights, and in church affairs. He had an ideal of what a university should be like, and he wrote very many witty and provocative articles in successive issues of the *Oxford Magazine* advocating that ideal and many different practical proposals for implementing it, including recommendations for how his fellow academics should respond to various proposals coming from the university authorities. He thought that the university should be run by all its teachers and researchers, and not by its vice-chancellor and a few administrators, let alone by businessmen far removed from university life. Thus for example he favoured the system whereby the vice-chancellor was a tenured Oxford academic seconded from his or her normal duties for two years, and opposed the change subsequently implemented to a system whereby the vice-chancellor who need not be already an Oxford academic was appointed for a much longer period. In the same spirit he argued strongly that the different views advocated in discussions of committees of the University Council should be made public. The extent of his influence is apparent from this incident, which he records in a *Magazine* article some years later: ‘There was once a leak from the Hebdomadal Council. The Assessor told her husband, who told my wife, who told me that Monday afternoon had been spent discussing what Lucas would say if various courses of action were adopted, leading to the conclusion that it would be best to do nothing.’ He was a strong advocate of well integrated joint BA degree courses between philosophy and some other subject to which it was evidently relevant, and was a keen supporter of the creation of the joint degrees in Mathematics and Philosophy and in Physics and Philosophy.

On first becoming a permanent fellow of Merton College, he was active in helping to oppose the Oxford city plan to build a ring road directly through Christ Church Meadow, which is a large green area in the centre of Oxford, lying between Merton College and the river Thames. This plan, put forward in a car-loving and so road-building era, was of a kind that would never be seriously considered by any planning authority today, in view of the enormous damage it would do to the environment; and even in 1961 the opposition to this plan proved successful. His support for the rights of consumers against large manufacturers was illustrated by his campaign against the terms of the warranty provided by BMC insurers to purchasers of new cars. He noticed that this warranty deprived the purchasers of their normal legal rights to free repair or replacement of products which failed to work satisfactorily. So he bought a single share in BMC, attended its annual general meeting, and made a speech denouncing the company for selling such insurance: and as a result, all car insurers' warranties no longer deprive purchasers of any legal rights. For a time he was a member of the Church of England's Commission on Christian doctrine, and also of the Church's Commission on Marriage and Divorce. In this role he had to defend his very moderate high church religious views both against the then fashionable liberal views of theologians who disliked precise statements of church doctrine, and also against the views of more fundamentalist evangelical theologians. When Robert Runcie, an old friend of his, was Archbishop of Canterbury, he sometimes discussed with Lucas what he ought to say in some speech which he was due to give in the House of Lords or in the Church of England Synod.

### Approach to philosophy

John Lucas wrote many books covering many different areas of philosophy; and on the whole supported a fairly common-sense view about them in a conversational style with examples showing sensitivity to different viewpoints, in contrast to many contemporary philosophers who defend hard, precise, philosophical doctrines by hard, precise, philosophical arguments. Anyone whose philosophical education began in Oxford in the 1950s, when 'ordinary language philosophy' began to share dominance with logical positivism, might be expected to begin their philosophical discussion of some topic with an analysis of how its crucial words are used in ordinary language – for example, to begin a discussion of perception with an analysis of how 'it looks red' is normally used, and to begin a discussion of the nature of knowledge with an analysis of how 'know' is normally used. That is certainly a useful way of ensuring that the discussion of a philosophical topic is not a discussion of a topic invented by the philosopher, but rather a discussion of a topic on which non-philosophers are seeking illumination. But it is now generally recognised in the analytic tradition (that is, the tradition which evolved over

the last hundred years mainly in Anglophone countries and now far beyond these), that a major task of philosophy is to see what, if any, is the metaphysical foundation for the distinctions made in ordinary language. Lucas often began his discussion of some philosophical topic with a description of how words and sentences are used in ordinary language and the normal beliefs which we express by means of it, before putting forward any recognisably philosophical thesis. Lucas put forward in his British Academy Philosophical Lecture of 1986 on 'Philosophy and Philosophy of' a general justification of this approach, similar to that of Neurath. There are, he plausibly claimed, too few incorrigible truths, from which to reach a world view by deductive reasoning. So, he argued, philosophy must start from whatever beliefs we find ourselves with, and criticise some of them on the basis of others of them, using non-deductive arguments to reach a moderately justified world view. This general approach to different philosophical problems is evident in his *Reason and reality* (PDF file put on the net in 2006, and published as a printed book in 2009). This was a large book containing his final published views on many of the issues of metaphysics which he had discussed previously in separate books and on many which he had not discussed previously. In this book he began his discussion of realism (the issue of what kinds of thing are real) with the remark (p. 220) that 'although the analysis of ordinary language cannot, as was once claimed, give us all the answers in philosophy, it is nevertheless a useful tool; it can save us from bad mistakes.' And so, for example, he endorsed J.L. Austin's claim that 'the meaning of the word "real" is shown in what it is being contrasted with'. But he criticised Austin's choice of examples to illustrate such contrasts, such as the contrasts between 'real coffee' and 'ersatz coffee', and the contrast between 'real silk' and 'artificial silk', which Austin used in order to suggest that contrast between the real and the non-real was merely a contrast between different kinds of mundane objects and properties. Lucas argued that the disagreements among philosophers about which kinds of entities are real are real philosophical disagreements, and that 'what constitutes reality is revealed by what is denied by various versions of anti-realism – philosophical doctrines denying the reality of some sorts of entity commonly believed to be real'. He distinguished different kinds of anti-realism, denying the existence of different kinds of entity – phenomenism denies the existence of material objects, behaviourism denies the existence of mental events distinct from their manifestation in behaviour, moral subjectivism denies that there are objective moral truths, and so on. He went on to give different careful accounts of the different marks of reality which would justify us in claiming that some kind of entity is real – for example, that it exists independently of the observer, that statements about it are knowable, that it has causal influence, and so on. And on most of these issues. Lucas took the common-sense view that the entities whose status was disputed are indeed real.

## Space and time

In sharp contrast to the ‘ordinary language’ philosophy of the 1950s, analytic philosophy has now sprouted some extreme metaphysical doctrines, often purportedly dependent on science, of which Lucas has been powerfully critical. One of these doctrines concerns the nature of time. Many philosophers have espoused the ‘block universe’ view that all moments of time in a sense always exist, and the distinction between past and future is merely a distinction in the point of view from which we view the world, and not a distinction in the world. John Lucas opposed this view, arguing in several books that there is a deep distinction between what has happened and what will happen. His first book about the nature of time was *A Treatise on Time and Space* (1973), a large book which covered many deep metaphysical issues. Lucas followed Kant in approaching the issues of time and space by means of arguments purporting to show that we have to think of ourselves as existing in a time and space of a particular kind, if we are to make sense of our experiences. He claimed, surely correctly, that our conscious experiences have a temporal structure – they are experiences of one state of affairs being followed by another state. He argued that we must think of the before-and-after structure also as governing our interactions with other people. One argument which he gave for this was that we could not have a conversation with another person unless each of us construed the speech of the other as taking account of what the other person had said previously. More generally, we could not make a difference to the world unless we thought of the world being one way to start with and then as a result of our actions subsequently a different way. So there could be neither conversation nor agency unless the same one-dimensional temporal order governed all events; hence the impossibility of backward causation (a cause causing an earlier effect). He argued that ‘same cause, same effect’ needs to operate for us to be agents making a difference to the world; ‘different cause, same effect’ would make that impossible, and so – he argued – some form of the second law of thermodynamics would need to operate. He argued that a physical space would make it possible for ordinary objects subject to change to be qualitatively identical (in their intrinsic qualities) but numerically distinct. He analysed P.F. Strawson’s claim that there could be different qualitatively identical objects in a purely auditory world, and argued that the features that objects would need to have in an auditory world to make that possible would make the purely auditory world have a spatial character. He went on to argue that if change was to be possible, there must be impenetrable objects existing in a space of more than one dimension. A Euclidean space is simpler than any other space, and so we should assume our space to be Euclidean in the absence of contrary evidence. And he then proceeded to argue for the necessity, at least in a simple and comprehensible universe inhabited by conscious beings, of various other features of space and time – for

example Space being three-dimensional, and relations between spatial and temporal intervals being governed by the Lorentz transformations.

This book set the discussions of its topics in the perspective of discussions about them in classical Greek philosophy, Christian theology, the 17th-century scientific revolution, and 20th-century physics. It included, as well as the very general arguments which I mentioned in the previous paragraph, some rigorous mathematics. At a mundane level, the book is impressive in being a large very elegantly printed book containing lengthy extracts in Russian and Greek (almost always translated), equations in red and blue type and complicated lattice diagrams, printed in the pre-digital era when such complicated printing would have been very expensive – how did Lucas persuade the publisher (Methuen) to produce it? At a deeper level, the scope of this book and the way in which it connects many different issues is enormously impressive. Reviewers noted its many insights into issues about which they had not thought. But one very general issue which they raised was that it was not clear how far Lucas was purporting to show that time and space have certain necessary features, how far he was purporting to show that we must think of them as having certain necessary features, and how far it was necessary for it to have certain features in order for us thinkers to have thoughts at all.

Space and time were the major topics of three more of Lucas's books. *Space, Time, and Causality* (1984), was intended as an introduction to philosophy of physics for physics students, and to show them the relevance to their studies of philosophical problems. It does so in a clear way, and could have served as an introduction to philosophy of physics for anyone with a relatively elementary knowledge of physics. It discussed almost entirely issues raised by classical physics, expressing the hope that similar books would deal with philosophical issues raised by Relativity and Quantum theories. *SpaceTime and Electromagnetism* (1990), jointly authored with the physicist Peter Hodgson, showed how the Special Theory of Relativity is a consequence, not merely of certain experimental results, but of imposing on Newtonian mechanics certain conditions of simplicity and symmetry, and of the individuation of particles; and it went on to clarify the philosophical issues raised by Special Theory, especially the issue of how 'the relativity of simultaneity' should be understood.

*The Future* (1989) describes the complexities and vagueness of much ordinary language use of tenses, and the prospects for a tensed logic which knocks our tensed talk into logical shape – that is, analyses that talk in ways which conform to simple stateable logical rules. Lucas criticised the rule on which most logicians insist, that if a statement is true at one time, it is true at all times. He had two very different criticisms of that rule. His first criticism was that the rule fails to distinguish two different kinds of ordinary language statements about future events, which he called 'predictions' and 'conjectures'. He understood by 'predictions' statements which are true at the time at which they are made if and only if both there is good evidence for them at that time and what they assert



subsequently happens, and by ‘conjectures’ statements which are true if and only if what they claim will happen does happen. It would follow that the former have a truth value which may change as the evidence for them changes. So the meaning of ‘it will rain on June 30th’ uttered on June 29th would vary dependent on whether it was meant to be a conjecture or a prediction. If it was a conjecture, then it would have been true when uttered on June 29th if and only if it rained on June 30th; and if it was a prediction, it would have been true on June 29th if and only if both it rained on June 30th, and all the evidence available on June 29th strongly supported the hypothesis that it will rain on June 30th. Lucas’s analysis of our ordinary use is however subject to the criticism that it seems to confuse what is meant by a claim about the future with what would justify a speaker in asserting it. Lucas had made a similar distinction in *The concept of Probability* (1970), when he used the fact that we sometimes say ‘it is true that p’ in circumstances when we have what we regard as conclusive reason for believing p, and ‘it is probable that p’ in circumstances when we do not have what we regard as conclusive reason for believing p, to claim that this contrast in the circumstances in which we use the two expressions reflects a difference in their meaning. He argued that (when so used), these two statements are contraries, and that in their ascription to propositions truth and probability are (p. 12) ‘in the same line of business’. His critics however denied that truth and probability are in the same line of business, on the grounds that statements may have different degrees of probability at different times on different evidence and also have truth values (normally supposed to be either true or false); and their probability may change without their truth value changing. So this first criticism provides no good grounds for abandoning the ‘if true, always true’ rule.

A significant number of philosophers would however agree with Lucas’s second criticism of the ‘if true, always true’ rule, also advocated in *The Future*. He defended the metaphysical thesis that fully to describe how the world really is, we need tensed language as well as tenseless language. Tensed sentences such as ‘it is raining now’, ‘it will rain tomorrow’, or ‘it rained yesterday’, he argued, can be analysed in terms of the operation of an operator – such as ‘it is true today’ or ‘it will be true tomorrow that’ or ‘it was true yesterday that’, on a tenseless sentence such as ‘it rains’. Lucas distinguished a sentence type (any utterance on any occasion of words arranged in the same way) from a sentence token (a particular utterance of a sentence type). He then claimed that while tenseless type sentences, if true at one time, are true at all times, token tensed sentences of the same sentence type, if true at one time, may be false at another time. For example, ‘it rains on June 30th 1989’, if true at one time is true whenever it is uttered; but ‘it will rain tomorrow’ uttered on June 29th 1989 may be true on June 28th 1989, but false on June 27th 1989. As observers from a timeless perspective, we need only tenseless sentences to describe the world. But, as agents in the world, Lucas claimed, we need also tensed sentences to describe it. For in order to know how to act and react, we need to

distinguish what has happened already from what is likely to happen in future. That the danger is past is ground for rejoicing, but that there is likely to be danger in future is grounds for fear. In terms of the way in which this issue is more usually discussed, in order to state everything true about the history of the universe, we need to list the events in its history individuated both by McTaggart's A-series of times in terms of how long ago they occurred or how long in the future they will occur, as well as by the B-series of times in terms of the dates at which they occurred (for example, June 29th 1989, or June 30th 1989). Lucas was not happy with this way of making the distinction, one reason for which was that in his view saying on June 29th 1989 that it will rain tomorrow is not predicating a property (of occurring on the next day) of an event, but operating on a proposition, 'it rains on June 29th 1989', which has a timeless truth value, to make a different tensed proposition; and so he followed others in developing a 'tree semantics' for the truth values of tensed propositions. But he shared with A-series advocates the strong belief that there is an inherent directedness of time, from fixed past to open future, evident to all experience, for example, that the Battle of Hastings occurred several centuries before the Reformation, the former being fixed at a time when it was not yet fixed whether the latter would occur.

### Free will and responsibility

Lucas shared the normal view of most of us that humans are morally responsible for many of our intentional actions; and that we can only be responsible if we are acting freely. By far the majority view of analytic philosophers in the 1960s and 1970s was that having 'free will' is simply a matter of acting freely; and that acting freely is logically compatible with being caused deterministically to act as we do. This view, called 'compatibilism' was normally spelled out as the positive doctrine that having free will is simply a matter of not being subject to 'constraint'. Being subject to 'constraint' meant being physically forced unavoidably to do some action, or caused unavoidably to do it by a recognised psychological compulsion such as kleptomania, or threatened with dire consequences if you do not do the action; but merely being caused by your brain state or your psychological condition of a kind not considered an illness to do some action did not, on this then fashionable view, impede your free will. Lucas's book *The Freedom of the Will* (1970) began with a very firm defence of the contrary view, 'incompatibilism', that it is a necessary condition if a human is to have free will, that their actions are not fully determined by any prior causes at all. He acknowledged (p. 15) that to say that an action is 'free' may mean any of the many different things, including 'not being under constraint' in some way. We may indeed often talk of someone who does some action when they are not threatened with dire consequences if they do not do it, as acting 'freely', but there remain issues of whether they are doing it 'freely' in some other sense,

and in particular whether they are doing it ‘freely’ if they are caused to do it by their brain state. And Lucas went on to argue, very firmly, that ‘absence of constraint’ is not ‘the relevant sense of freedom’ when we are concerned ‘with responsibility’. He argued that there was no justification whatever for claiming that you are unfree if caused to do an action by your kleptomaniac condition, but free if caused deterministically to do an action by some psychological condition which was not considered a mental illness. More generally, he argued (p. 28) that determinism would deprive humans ‘of any real say in the course of events’ because they would be ‘totally dependent on other factors outside their control’. The debate between compatibilists and incompatibilists remains a philosophically vigorous one, but opinions are far more evenly divided than they were in 1970, and Lucas’s arguments helped the incompatibilist cause. Although some of his books on other topics show considerable sympathy for rival views and a certain tentativeness about their conclusions, the arguments of this book in favour of incompatibilism are sharp and the conclusions which it reaches are clear, definite, and persuasive.

After arguing in favour of incompatibilism, Lucas devoted the rest of this book, as he had devoted several earlier articles and would devote many subsequent chapters or articles, to arguing that human beings are not (normally) predetermined to perform their intentional actions, and – more positively – do have free will in the sense that makes them morally responsible for their intentional actions. This, he claimed, can be shown by two mathematical and physical discoveries of the 20th century – Gödel’s Theorem and quantum theory. He devoted a small part of *The Freedom of the Will* and of *Reason and Reality*, to arguing that quantum theory shows this. In claiming that quantum theory provides good evidence for human actions not being fully determined by their brain states, and so for humans having limited free will, Lucas was one of a fairly small minority of philosophers and scientists who have made the same claim. But Lucas is best known for his claim that Gödel’s first incompleteness theorem shows that humans have free will. He sought to prove this in a two-stage argument. The first stage was to show that physical determinism implies the existence of a formal system that models the mind. The second stage was to show by appeal to Gödel’s first incompleteness theorem that there can be no such formal system.

The first stage of his argument proceeded from the claim that

there is only a definite finite number of beliefs which, according to the physical determinist, a particular human being can be said to hold. If this is so, the reasoning of any particular human being can be viewed as a logistic calculus ... The beliefs held at the outset are the ‘initial formulae’ (‘primitive propositions’, ‘postulates’ or ‘axioms’); and the types of inference drawn by that particular person (whether or not we regard them as valid, sound, or cogent) will be the ‘rules of inference’. Thus each human being’s reasoning if he can really, as the physical determinists allege, be completely described in physical terms, may be viewed as a proof-sequence in some logistic calculus. (*The Freedom of the Will*, p. 132)

(This idea is reminiscent of J.H. Woodger, *The Axiomatic Method in Biology*, 1937.) By a logistic calculus Lucas meant what would now be more usually called a deductive formal system. It doesn't follow straightforwardly that if some human's reasoning can be described by a logistic calculus, that what the human being will do with the resulting beliefs can also be so described; the calculus would need to show when the human will and when the human will not act on her beliefs. Also, Lucas did not provide a detailed explication of how to read off from a physical determinist's claim to have fully explained a human mind, a logistic calculus, that is a formal system with a precise syntax, which he needs for the second stage of the argument. But, supposing these matters to be resolved, physical determinism implies mechanism.

The second stage of Lucas' argument for freedom of the will was to refute mechanism by his 'Gödelian argument', first published in his paper 'Minds, machines and Gödel' in 1961: 'Gödel's Theorem seems to me to prove that mechanism is false, that is, that minds cannot be explained as machines' (p. 112). By 'Gödel's theorem', Lucas means Gödel's first incompleteness theorem. This proved that in any formal system adequate to prove basic arithmetic of addition and multiplication of the natural numbers, there will always be a well-formed sentence *G* such that if the system is consistent, then *G* is not provable in the system, and *G* is true if and only if the system is consistent.

Lucas argued that if a would-be mechanist puts forward some particular computer programme as 'a complete and adequate model of the mind', then—since a computer programme is tantamount to a formal system—there will be a Gödel sentence for its corresponding formal system which, if the system is consistent, will not be provable in that system; and whose truth is implied by the consistency of the system. (If the system were inconsistent, it could not be a correct model of any actual process.) But in order to claim to know that the system represents all the truths that can be known by some human, the mechanist must claim that the system can prove its own Gödel sentence. Yet Gödel's theorem proves that if the system is consistent, that cannot be done. So the mechanist must accept that that their system has failed to model all that their mind can know. This argument proves weak anti-mechanism, the thesis that there cannot be a computer programme of which we can know that it generates all that some human can know; and in my view this is an important conclusion. But it does not establish strong anti-mechanism, the doctrine that there can be no formal system that generates all that some human can know. Showing that no programmes or formal systems put forward by would-be mechanists could model the mind does not show that no programme or formal system (among the infinitely many unknown to us) could model the mind.

Lucas had another argument for the truth of Gödel sentences, based, not on the claim of a would-be mechanist, but on our understanding of the proof of Gödel's first incompleteness theorem: 'any rational being could follow Gödel's argument, and convince himself that the Gödelian formula, although unprovable-in-the-given-system, was none-

theless – in fact, for that very reason – true’ (‘Minds, machines, and Gödel’, p. 115). This argument would, if correct, have established strong anti-mechanism. However in making this argument, Lucas overlooked the fact that the conclusion of Gödel’s first incompleteness theorem, that the Gödel sentence is not provable in the given formal system, depends on the hypothesis that the system is consistent, which is not always known. Furthermore, if this argument were correct, it would imply too much. For the truth of its Gödel sentence implies the consistency of a formal system. So if this further argument were correct, it would establish the consistency of every formal system for which Gödel’s first incompleteness theorem is provable. But there are formal systems for which Gödel’s first incompleteness theorem is provable whose consistency is unknown – for example Quine’s system *New Foundations*.

There are reasons to doubt that any version of Lucas’s Gödelian argument could succeed in proving strong anti-mechanism. One is that there is a theorem by William Reinhardt (1985) that there can be no proof of strong anti-mechanism in the system  $EA_T$  (Epistemic Arithmetic with Tarskian Truth Theory), and Lucas’s arguments can be formalised in  $EA_T$ . Another is that Gödel investigated the implications of his incompleteness theorems for the relationship between minds and machines and was clear that on our present understanding of knowledge and truth, strong anti-mechanism could not be proved outright. Gödel made this point in a paper written in 1951 and only published posthumously, key points of which were published by Hao Wang in 1974 in his book *From Mathematics to Philosophy*, where he reported what are ‘in Gödel’s opinion ... the two most interesting rigorously proved results about minds and machines’. The second result is Gödel’s disjunction: ‘*either* the human mind surpasses all machines (to be more precise: it can decide more number theoretical questions than any machine) *or* else there exist number theoretical questions undecidable for the human mind.’ The first disjunct is strong anti-mechanism, which Lucas claimed to have proved outright from Gödel incompleteness, while Gödel saw that all that could be proved is the disjunction, *either* strong anti-mechanism *or* that there are truths unknowable by human minds. It is puzzling that Lucas did not engage with Gödel’s study of the implications of his incompleteness theorems for the relationship between minds and machines.

Alan Turing in his 1950 paper, ‘Computing Machinery and Intelligence’, had briefly formulated and rejected a version of an argument similar to that of Lucas. Lucas’s argument got significant support from Roger Penrose in his books, *The Emperor’s New Mind* in 1989 and *Shadows of the Mind* in 1994, in which he argued that Gödel incompleteness shows ‘the un-tenability of the viewpoint . . . that our thinking is basically the same as the action of some very complicated computer’. Penrose’s argument in the first book, elaborated in the second, is essentially the same as Lucas’s Gödelian argument, which has come to be labelled the Lucas-Penrose argument. In *Shadows of the Mind* Penrose propounded a new argument, much more complicated and subtle than the Lucas-Penrose

argument, which requires type-free theories of truth and knowledge in which to formalise it, and so is not ruled out by Reinhardt's theorem (see above). Penrose came to his initial idea that anti-mechanism follows from Gödel's incompleteness while he was a graduate student in Cambridge, eight or nine years before Lucas published his seminal paper 'Minds, Machines and Gödel'; and so Penrose's espousal of the Lucas-Penrose argument is not evidence of Lucas's influence. Nevertheless, Penrose's standing as one of the greatest mathematicians and theoretical physicists of his generation testifies to the powerful attraction of Lucas's Gödelian argument.

Lucas's views on the Gödel incompleteness theorems gave him what he considered to be an important scientific foundation for his strong conviction of the (to him) evident sharp difference between humans and inanimate objects. He took further the views which he expressed in an early paper on 'The soul' in his contributions to the two series of Gifford lectures at Edinburgh in 1971-3, given in the form of lectures by Lucas, Anthony Kenny, H.C. Longuet-Higgins, and C.H. Waddington, and discussions between them. These lectures were published in two small books, *The nature of mind* (1972), and *The development of mind* (1973) by Edinburgh University Press. In his contributions Lucas emphasised that humans have conscious experiences and act for reasons, while inanimate objects do not have experiences and do not act for reasons. In consequence he was strongly opposed to behaviourism, the doctrine that all statements about human thoughts and feelings can be reduced to statements about their actual and hypothetical (that is what they would do in certain circumstances) behaviour. But, as with many of the other topics which he discussed, this common-sense approach did not lead to any systematic theory, which in this case would have been a precise dualist theory

Lucas's claim that humans do have free will, and so are morally responsible for some of their actions, led to his discussion in his book *Responsibility* (1993) of the degree of our responsibility for different actions. A consequentialist must hold that we are equally responsible for all the (foreseeable) consequences of our actions and of our failures to act. But Lucas argued that on the contrary, while we are responsible for our actions, we are not responsible for our failures to act – unless there are special reasons why we ought to have acted. He discussed (pp. 45-51) Bernard Williams's example of 'a traveller in South America who comes on a village where a hit squad is about to shoot twenty of the villagers, and their captain says that if the traveller will himself pull the trigger on one of them, the others will go free'. If consequentialism were recognisably true, there would be no dilemma: the best consequence would evidently be obtained by the traveller shooting one of the villagers, given – Lucas assumed – that the traveller reasonably believes that the hitmen will shoot all twenty villagers unless the traveller kills one of them. Lucas argued (p. 51) that 'if the traveller refuses, and the guerrillas carry out their threat, [the traveller] is not automatically and necessarily answerable for what happens. He has not killed anyone. The deaths are due entirely to the guerrillas' actions, not to his inaction.

The chain of causal responsibility is broken by their autonomous action. They do not have to kill the hostages. It is entirely up to them whether they do it or not. The responsibility is therefore theirs.’ Lucas acknowledged (p. 48), however, that ‘consequences are always relevant, though sometimes not very relevant, and some bad states of affairs are ones we are always under great obligations to avert’. In the rest of the book, Lucas went on to defend various positions on various controversial moral issues. ‘In so far as we take pride in what our predecessors have done, and enter into their achievements ... we identify also with the bad things they have done, and make their misdeeds our misdeeds for which we must answer’ (p. 77.) In this Lucas shares the morality of many contemporary Western intellectuals. But on other moral issues he is out of line with that morality. Punishment, he argued, should be backward-looking and so retributive (on the basis of past misdeeds), and not forward-looking and allocated on utilitarian grounds of prevention, deterrence, and reform. Likewise, contrary to Rawls, he argues that desert arising from past services is ‘an appropriate basis for the just allocation of benefits’, though not the only such basis. Sexual intercourse should ideally take place only between married couples. It is a fault of meritocracy that it ‘plays down the importance of marriage, the family, and the home’ (p. 255).

### Political philosophy

Lucas’s first book, *The Principles of Politics* (1966) was a large book working out in detail the kinds of political organisation and constraint needed for a society of humans with a nature like ours. To constitute a society humans have to interact with each other and to share some values. But actual humans are imperfect – that is, only partly unselfish, only partly rational, and only partly well-informed. He claimed, plausibly enough, that in their political writings, a few philosophers exaggerated the imperfections of humans, as did Hobbes who regarded humans as entirely selfish; but that more philosophers have assumed that humans are more unselfish, more rational, or more well-informed than they actually are, as did Kant and R.M. Hare. The book went on to work out the consequences of Lucas’s account of human nature for the best form of political organisation and the best kinds of limitation of freedom necessary for humans who have the natures they actually have. Society will need laws limiting freedom of its citizens in certain respects and limited punishments of those who break the laws, a recognised process for creating and repealing laws, judges to interpret laws and to determine who has broken which laws and what their punishment shall be, and Lucas analyses the best ways of satisfying such requirements. *On Justice* (1980) was concerned with what makes procedures for reaching decisions just ones, such as the rules that no person should be judge of their own cause, and that all persons should be entitled to equal consideration by the

law. Justice is at least in part backward-looking; it punishes or rewards persons for what they have done or what they are like, and not what they would be like if they were punished or rewarded. Hence punishment must be in part retributive. It sends a message to the wrongdoer that society is hostile to him or her for what they have done. But (p. 147) ‘justice does not require that there be a rigid tariff’ of punishments for particular breaches of a particular law, but it must take account also of the nature and circumstances of the victim and the criminal; there is a place for mercy in the allocation of punishment. Distributive justice, the just way of distributing different goods to different members of society, should take into account many things other than the needs of those members, varying with what the goods are and who the members are – for example, ‘I ought to consider my family first in allocating my time and attention, and worldly goods’ (p. 166). As in most other areas of philosophy Lucas was hostile to simple general formulae. He devoted a chapter to Rawls’s theory that the rational way to organise society is the way all of us would choose if we chose how society should be organised from the standpoint of a previous life, ignorant of which future person we would be in that new society. Lucas argued cogently that this was a confused suggestion, because what sort of a society I would choose to live in must depend on what sort of person I would be in that society; ‘I may be, for all I know, a compulsive murderer, in which case I should be far better off if there are no punishments’ (p. 193). In ignorance of what sort of a person I would be in that society, there can be no definite answer to which sort of society I would choose to belong. Lucas was similarly critical of the view attributed to Nozick and Hayek that ‘the keeping of covenants constitutes the whole of justice, and that the market economy is ... of necessity the fairest’ form of organisation (p. 214); he pointed out that the only agreement that the weakest can get from the strongest may be unjust. In *Democracy and Participation* (1976) he analysed the advantages and disadvantages of different kinds of democracy and different levels of participation in the decisions of government and their practical implementation; and the strength of different arguments for why citizens have an obligation to obey the law, and the limits to that obligation. *Ethical Economics* (1996) was co-authored by Lucas and M.R. Griffiths, a management consultant. It analysed the moral responsibilities of businessmen to other businessmen with whom they were negotiating, to their employees and to their customers, which form the framework within which they were right to price and market their goods so as to make money.

All of this may be thought to be a deep and detailed exposition of the common sense of many moderately conservative British people of moderate education and moderate means, concerned for their own well-being and that of their family, but sensitive to the needs of others. But it is also one that will make that common sense appealing to others who do not usually share that outlook.



## Religion

As an analytic philosopher and a Christian believer, Lucas made an early contribution to 'analytic philosophy of religion', the application to traditional religious claims of the methods and results of the kind of philosophy practised in the analytic tradition. Analytic philosophy of religion began with the publication of two influential collections of essays – *New Essays in Philosophical Theology* (1955), edited by Antony Flew and Alasdair MacIntyre, and *Faith and Logic* (1957), edited by Basil Mitchell, a close friend of Lucas. All the contributors to *Faith and Logic* were Oxford philosophers or theologians, and John Lucas was one of them. His contribution to it was his paper on 'The soul'. He argued that in talking about a person's 'soul' we are talking about their personal qualities and experiences and emotions, their 'inmost self which may long for God' (p. 135). Talk about the soul was useful in emphasising the very real difference between persons and inanimate objects, which he re-emphasised in his Gifford lectures; and it was at this early stage that he vigorously opposed behaviourism, then very influential as a result of Gilbert Ryle's *The concept of Mind*. Lucas claimed plausibly that Ryle's argument depended for its justification on the thesis of logical positivism that the meaning of statement is to be equated with the method by which it can be verified; and he gave powerful arguments for rejecting that thesis. *Freedom and Grace* (1976) is a collection of short papers and occasional sermons delivered on miscellaneous occasions on various issues centred around the tension arising between that human free will in which Lucas so strongly believed and the human condition of self-centredness. They bring to life, in simple non-philosophical language with a sensitivity to the human condition, many associated doctrines of grace, providence, sin, atonement, redemption, and forgiveness. A theme which runs through them is that 'the fundamental reason why I ought to do what I ought to do, is because I love God, and doing what I ought to do is an expression of my love of God' (p. 92). At about the same time Lucas gave a lecture in Durham Cathedral on *Butler's philosophy of religion vindicated* (1978), subsequently published as a separate pamphlet. This defended Butler's non-deductive and to some extent pragmatic defence of Christian theism, directed both against those who claim to have sound deductive arguments for the existence of God, and those who claim to have sound deductive arguments for the non-existence of God. In this lecture, Lucas anticipated the approach of much contemporary analytic philosophy of religion to arguments for the existence of God, although in a more tentative way. Analytic philosophy of religion is much the poorer for the fact that Lucas wrote no systematic treatise on any topic in the philosophy of religion. Nevertheless we have reason for gratitude for those individual chapters or few paragraphs in books of his primarily concerned with other topics, in which he comments on the relation of some view to Christian theology. In particular, in several places in his work on space and time and other writings, he emphatically opposed the long tradition

of much Christian theology that God is timeless and changeless. He claims that a God who is a personal agent, as he is always depicted in biblical books, could not know infallibly the future actions of free human agents, and would be able to change his mind, and would be no less perfect for being like this. In this respect also Lucas anticipated subsequent developments in analytic philosophy of religion; many recent philosophers of religion have advocated this temporal account of God, now called ‘open theism’. John Lucas’s sometimes slightly tentative religious faith was a central element of his outlook on life.

### *Acknowledgements*

I am very grateful to Morar Lucas, for answering many questions on John Lucas’s life, and to Tim Horder, a current editor of the *Oxford Magazine*, for providing me with copies of many past articles by John Lucas in that magazine. Very many thanks to John Tassioulas for providing me with an assessment of Lucas’s contributions to political philosophy, to supplement my own very limited knowledge of that area of philosophy. I am enormously grateful to Daniel Isaacson for providing me with a long report expounding and analysing John Lucas’s many publications of his Gödelian argument, and comparing them with Penrose’s arguments against strong AI, and with Gödel’s arguments from his incompleteness theorems on the relationship between minds and machines, on a short summary of which my whole discussion of the matter is almost entirely dependent. I am however fully responsible for the selective use which I have made of that summary, and any consequent errors. Isaacson in turn wishes to record his gratitude to Stewart Shapiro for his generous willingness to respond to formulations of these issues.

### *Note on the author*

Richard Swinburne is Emeritus Nolloth Professor of the Philosophy of the Christian religion, University of Oxford. He was elected a Fellow of the British Academy in 1992.

This work is licensed under a Creative Commons Attribution-NoDerivatives 4.0 International License.

*Biographical Memoirs of Fellows of the British Academy* (ISSN 2753–6777) are published by The British Academy, 10–11 Carlton House Terrace, London, SW1Y 5AH  
[www.thebritishacademy.ac.uk](http://www.thebritishacademy.ac.uk)