EUROPEAN JOURNAL
FOR PHILOSOPHY OF RELIGION

Volume 2 Number 1 Spring 2010

ARTICLES

DISCUSSIONS AND REPLIES

BOOK REVIEWS AND NOTICES

GOD AS THE SIMPLEST EXPLANATION OF THE UNIVERSE

RICHARD SWINBURNE

University of Oxford

Abstract. Inanimate explanation is to be analysed in terms of substances having powers and liabilities to exercise their powers under certain conditions; while personal explanation is to be analysed in terms of persons, their beliefs, powers, and purposes. A crucial criterion for an explanation being probably true is that it is (among explanations leading us to expect the data) the simplest one. Simplicity is a matter of few substances, few kinds of substances, few properties (including powers and liabilities), few kinds of properties, and mathematically simple relations between properties. Explanation of the existence of the universe by the agency of God provides the simplest kind of personal explanation there can be, and one simpler than any inanimate explanation. I defend this view more thoroughly than previously in light of recent challenges.

THE TWO KINDS OF EXPLANATION[1]

I have argued over many years that theism provides a probably true explanation of the existence and most general features of the universe.[2] A major reason for this, I have claimed, is that it is simpler than other

[1] This paper is being published simultaneously with its publication here in Anthony O'Hear (ed.), *Philosophical Essays on Religion*, Cambridge University Press (on behalf of the Royal Institute of Philosophy), with the kind agreement of the Institute.

[2] This paper is dependent on much earlier writing of mine, especially *The Christian God* (Oxford: Clarendon Press, 1994) chs 6 and 7, *Epistemic Justification* (Oxford: Clarendon Press, 2001), ch 4, and 'How the Divine Properties Fit together: Reply to Gwiazda', *Religious Studies* 45 (2009), 495-8. The latter paper was a reply to Jeremy Gwiazda, 'Richard Swinburne's Argument to the Simplicity of God via the Infinite', *Religious Studies* 45 (2009), 487-93. I am grateful to Jeremy Gwiazda whose criticism of earlier views of mine helped me to formulate the view presented in this paper.

explanations.[3] The present paper seeks to amplify and defend this latter claim in the light of some recent challenges.

Explanatory hypotheses are of two kinds – inanimate (or scientific) and personal. In inanimate explanation we explain an event by means of an initial condition (or cause) C and a regularity or law of nature (N), such that these together necessitate or render probable the resulting event (E). To take a trivial example, we explain a given piece of iron expanding (E) by 'the iron was heated' (C) and 'all iron expands when heated' (N). We explain Mars being where it is today in terms of where it and the sun were yesterday and on previous days (C) and Kepler's three laws of motion (N) which together entail it being where it is today (E). In personal explanation we explain by means of a person (S) with certain powers (P), beliefs (B) and purposes (G). By a 'purpose' I mean an intention in what the person is doing; an intentional action is an action of bringing about what the person has the purpose of bringing about. We explain my hand moving (E) by means of me (S), having the power to move my hand (P), having the purpose of getting your attention (G) and believing that causing my hand to move will do that (B). And we explain people's purposes in terms of their desires (D) and beliefs (B). By a 'desire' I understand a kind of liability with which we find ourselves to form certain purposes which render it probable that we will form those purposes. Among our beliefs we have moral beliefs about what is good to do. Moral beliefs motivate us, incline us to do the relevant action, but they may not motivate us as strongly as non-rational desires. So sometimes we have to choose between forming a purpose involving yielding to a desire to do what is bad, or a purpose involving pursuing the good despite contrary desire.

There are however two different ways of construing inanimate explanation, depending on what laws of nature are, that is what are we claiming in claiming that 'all iron expands when heated' or (perhaps more realistically) 'all photons travel at velocity c in vacuo relative to all inertial frames' are laws of nature. We can, I think, these days quickly dismiss the Humean account that they are just assertions about what in fact happens: each bit of iron when heated in the past did expand, and

[3] See for example, my *The Existence of God*, second edition (Oxford: Clarendon Press, 2004), chs 3 and 5.

each bit when heated in future will expand. For there is a physical necessity in the operation of laws of nature not captured by Hume's account. That leaves two serious possibilities. The first is that laws are real things separate from the substances (physical objects), things which determine how those substances behave. The currently discussed version of this is the view that laws are relations between universals (that is properties which may be instantiated in many different substances); 'iron', 'expanding' and 'being heated' are universals tied together (and that has to be construed as tied together in a Platonic heaven) such that when you instantiate 'iron' and 'being heated' you inevitably instantiate 'expanding'. I shall call this view, advocated by Armstrong and others[4], the RBU (relations between universals) account. And so we explain the behaviour of a particular substance by something outside itself which determines also the behaviour of other substances – e.g. other pieces of iron. The alternative view is the substances – powers – and liabilities account (SPL). This was the normal view in ancient and medieval thought, and versions of it have been recently advocated by Harré and Madden and by Brian Ellis.[5] On the SPL account the fundamental laws of nature are causal laws; and these are generalizations, not (as Hume supposed) about what in fact happens, but about the causal powers of substances of a certain kind and their liabilities (either with physical necessity or with a certain physical probability) to exercise them. Thus 'all iron expands when heated' being a law of nature is a matter of every piece of iron having the power to expand and the liability (with physical necessity) to do so when heated. Kepler's laws are generalisations about the powers of planets to move in certain ways and their liabilities to do so under certain circumstances. It is a contingent matter that things fall into a few kinds distinguished

[4] See D.M. Armstrong, *What is a Law of Nature*? (Cambridge: Cambridge University Press, 1983); Michael Tooley, 'The Nature of Laws', *Canadian Journal of Philosophy* 7 (1977), 667-98; and F.I.Dretske, 'Laws of Nature', *Philosophy of Science* 44 (1977), 248-68. Armstrong construes universals in an Aristotelian way (that is, as existing only when instantiated). But that will not explain why their first instantiation had the character it did, e.g why the first piece of iron expanded when heated. That could only be explained if the universals were already tied together, and that would involve their existing before being instantiated, and so in a 'Platonic heaven'.This latter is the view of Tooley, and it is in his way that I have spelled out the RBU account.

[5] R. Harré and E.H. Madden, *Causal Powers* (Oxford: Basil Blackwell, 1975); Brian Ellis, *Scientific Essentialism* (Cambridge: Cambridge University Press, 2001).

by their powers and liabilities as well as by other properties. This ultimately derives from the fact that their constituents are fundamental particles (substances, such as electrons and quarks) which fall into a very few kinds distinguished from each other by their mass, charge, spin etc; these latter being – at least in part – analysable in terms of the powers and liabilities possessed by the particles .

Although I think that the main argument of this paper can be phrased in terms of either account of inanimate explanation, in order not to make the paper too long, I am now going to make the assumption that the SPL account is the correct account. One quick reason for rejecting RBU, is the implausibility of a Platonic heaven containing universals influencing the behaviour of mundane things. Another reason is that it enables us to give a more unified account of explanation[6]. For, given SPL, all explanation, inanimate and personal, now involves substances (either persons or inanimate things) and their powers. I assume that an inanimate thing must be physical (that is, public.) The difference between the two kinds of explanation is now that inanimate things have liabilities (inevitably or with a certain physical probability) to exercise those powers under certain circumstances and so to cause effects, whereas persons intentionally exercise their powers to cause effects in the light of beliefs about what exercising some power will achieve and the purposes which they seek to achieve. Explanation of an event will now consist (in the inanimate case) in the occurrence of circumstances under which some substance (or substances) was liable to exercise certain powers. That a piece of iron expanded when heated is explained by it (S) having the power to expand (P) and the liability to exercise that power when heated (L), and by (C) it being heated. Strictly speaking no law 'all iron expands when heated' is part of the explanation. The 'law' is a mere description of the powers etc., of all pieces of iron, relevant only because it entails the powers of this piece. Explanation of an event by personal explanation will invoke only S, P, B, and G; although all these factors might themselves be explained (in so far as they can be explained) by an inanimate explanation – e.g. B may be explained by the liability (L) of S to have a belief of a certain kind under certain circumstances which did in fact occur; and G may be ex-

[6] For further argument in defence of the SPL account see pp. 179-85 of my 'Relations Between Universals, or Divine Laws?', *Australasian Journal of Philosophy* 84 (2006), 179-89.

plained by S's desire to form such purposes, that is a liability (at least with physical probability) to do so. And conversely the factors involved in an inanimate explanation might be explained by a personal explanation.

THE CRITERIA OF CORRECT EXPLANATION

I suggest that an explanatory hypothesis (or theory) is rendered probably true (or likely to be true) by data (evidence) insofar as (1) the occurrence of the evidence is probable if the hypothesis is true and improbable if the hypothesis is false, (2) the hypothesis 'fits in' with any 'background evidence' (that is, it meshes with other hypotheses outside its scope which are rendered probable by their evidence in virtue of the other criteria), (3) the hypothesis is simple, and (4) the hypothesis has small scope.[7] The scope of a theory is a matter of how much it purports to tell us about the world, in the extent and precision of its claims. (3) and (4) are features internal to a hypothesis, independent of its relation to evidence and so determine its prior probability (probability prior to considering the evidence). While the more the hypothesis claims, the more likely it is to be false (which is what the criterion of scope says), simplicity carries more weight than scope; scientists consider some theory of enormous scope (concerned with the whole universe) probable if it has a relatively simple set of laws. There may be no relevant background evidence, and then criterion (2) drops out. One case of this is when a hypothesis has a very large scope (purports to explain a vast amount) and so there is little if any evidence about fields beyond its scope. Among large scale theories of equal scope, such as theism and rival accounts of why there is a universe of our kind, relative probability depends on criteria (1) and (3) alone; and so in the case of theories leading us to expect the evidence with the same probability (that is, satisfying criterion (1) equally well), on criterion (3) alone. Let me give you two examples – one of each kind of explanation – illustrating how, among theories satisfying the other criteria equally well, the simplest theory (simplest in an intuitive sense yet to be analysed more precisely) is the one most probably true.

[7] For a fuller account of these criteria, but one which does not distinguish the different roles some of them play on the RBU and SPL accounts of laws of nature, see my *Epistemic Justification*, ch 4.

Suppose we find among pages recovered from an ancient library three pages in similar handwriting of an apparently connected philosophical argument. One hypothesis is that the same person wrote all three pages. An alternative hypothesis is that each page was written by a different philosopher; all three philosophers had similar handwriting, and thought independently of the same argument which they wrote down; but only the first page of the first philosopher's text, the second page of the second philosopher's text, and the third page of the third philosopher's text have survived. The two hypotheses are of equal scope – telling us about who wrote these pages; and are such as to lead us to expect the data with equal probability. But, unless there is relevant background evidence, the first hypothesis is obviously more probable in postulating only one person writing the pages rather than three. For my second example consider a theory which renders probable the same astronomical evidence observed so far as does General Relativity. General Relativity does this (on the SPL account) by attributing to stars certain powers and liabilities to exercise them dependant on the structure of the spatio-temporal region in which they are situated; and it predicts further observations tomorrow in virtue of the same powers and liabilities. The rival theory claims that the liabilities of these things to exercise their powers will depend on the structure of their spatio-temporal region in a different way when the expansion of the Universe causes galaxies to have a certain average distance apart (which distance they will attain tomorrow) from the way they depend today. General Theory is more probable than its rival because (put in terms of laws) it consists of only one set of complicated equations and so is simpler than its rival which consists of a conjunction of two sets of complicated equations. So again simplicity is evidence of truth. In so far as General Theory gets support from other theories of physics with which it fits, then unless simplicity is evidence of truth, those other theories would be just as probable as rival theories adjusted in a similar way, and General Theory's rival would fit better with those rivals; and so again General Theory and its rival would be equally probable on the evidence. But they are not, and that is because of the crucial role of the criterion of simplicity.

THE NATURE OF SIMPLICITY

The assimilation of scientific explanation to personal explanation, consequential on adopting the SPL account of laws of nature, makes it possible to give common criteria of simplicity covering both types. So let's look at what the criterion of simplicity, when used in this way as evidence of truth amounts to. It consists of various sub-criteria. First, fewer entities (alias, substances or objects). Astronomers don't postulate an extra planet, unless thereby their data are made more probable. And the historian in my earlier example postulates as few philosophers writing texts as possible. What constitutes one entity as opposed to two? Entities require a certain causal unity to them; they stick together. But whether the parts of a physical thing stick together is matter of degree, and it's not always clear when there are two entities rather than one. (For some purposes you can treat a double-star system as one entity, for other purposes you must treat it as two entities.) But clearly an entity which has no parts is just one entity; and as such a very simple one. Secondly, fewer properties attributed to entities. Don't postulate a new property possessed by (e.g.) a fundamental particle unless it results in a gain of explanatory power. Again, the application of this subcriterion depends on how you count properties. A property defined by similarity to paradigm examples of its application, such as 'green' or 'mass' or 'bright' counts as one property; properties defined as conjunctions or disjunctions of such properties (or as having more complicated probabilistic relations to such properties) count as two or more properties. It follows from this subcriterion that hypotheses are simpler, the more accessible to observation (or experience generally) are the properties which they postulate. This can be illustrated by the well-known philosophical example of two theories to account for the colours of emeralds. 'All emeralds are green' and 'all emeralds are grue' (where 'grue' means 'green before 2050 A.D., or blue thereafter') both render the data about the colours of emeralds now (in 2010 A.D) available equally probable. But the theory 'all emeralds are green' is more probable than 'all emeralds are grue', and this is because 'grue' is defined in terms of an accessible property (green) and another property (the date, whose definition in terms of what is accessible clearly has a certain complexity).[8]

[8] But couldn't there be a being which just recognized things as 'grue' without doing so in virtue of their colour and the date? There could certainly be a being which

Among the properties of objects are their powers and liabilities, purposes and beliefs. In explaining human behaviour we need to attribute to humans as few and as accessible such properties as will suffice to render probable their behaviour. And we need to attribute to them as few and accessible desires and moral beliefs as will explain (at least in part) their purposes; and as few and accessible liabilities to acquire beliefs as will explain their moral and other beliefs. I illustrate this for the case of powers. Someone may walk (as far as we can judge, intentionally) one mile from A to B one day at 3mph, and two miles from C to D another day at 3 mph. But we wouldn't explain his behaviour on the first day simply by his having the power to walk one mile at 3mph on the first day; and explain his behaviour on the second day simply by his having the power to walk two miles at 3mph on that day. Rather we'd explain both pieces of behaviour by his having the power over a period to walk at least two miles in a day at 3 mph. Since both previous powers follow from one equally accessible property, we attribute the latter to the person; and so we believe that in virtue of having the more general power, that person will be able to walk at least two miles at 3mph on other occasions. And so we seek to attribute to humans powers as few and as accessible as will explain their behaviour. Similarily for inanimate things – for example, we attribute to fundamental particles as few forces (that is, powers to affect other substances) as possible.

We explain the exercise of powers by inanimate things by their liabilities to exercise those powers. Liabilities are also simpler, the fewer the (accessible) properties by which we distinguish them. A power of a substance is a power to exercise various amounts of causal influence; and the liability is a liability to exercise a power, and to exercise some particular degree of it under certain contingent circumstances. A body may exert more or less gravitational influence on another body in virtue of its liability to do so being dependent on its mass, the mass of the other body, and their distance apart. And an explanation is simpler, the simpler

classified together (in virtue of their similarity to paradigm examples) the same objects as we would call 'grue' (on the grounds of their satisfying the stated definition). But he would be picking out a different property ('grue*') which – as far as his experience went – was coinstantiated with 'grue'. Yet there could be no guarantee that the two properties individuated in different ways would always coincide. We have no access to the property of being grue* and so cannot use it in our explanations of things.

the mathematical relations between the degrees of the various properties and liabilities, and the simpler the mathematical entities involved in stating the relations. Such a relation or entity A is simpler than another one B, if A is defined and so can be understood without reference to B but not vice versa. For this reason multiplication is a more complicated relation than addition, numerical powers more complicated than multiplication, and vectors more complicated than scalars; and large finite integers are more complicated entities than small ones (you can't understand '5' except as '4+1', but you can understand '4', '+', and '1' without understanding the notion of '5'); and (as their name implies) complex numbers are more complicated than real numbers, real than rational numbers, rational numbers than integers. So an explanation which explains the pressure exerted by a gas on the walls of a container is simpler if it makes this depend on only three other quantities (the volume of the container, the temperature of the gas, and a constant varying with the kind of gas) rather than on four. It is simpler if the relation between the variables involves just multiplication as in the Boyle-Charles law pv=KT, rather than on exponentials, logarithms and square roots, for example $p= \log_e vk^{-1/2}T^2$. If laws of both kinds satisfied the first two criteria equally well, a law of the first kind would be more probably true than a law of the latter kind. For exponentials, logarithms and powers are defined (and so can be understood) in terms of multiplication but not vice versa. So too a law of gravity – $F= mm'/r^2$ is to be preferred to a law $F= mm'/ r^{2.00\ (\text{ten zeros})\ 1}$ if both equally well explain the data (of measurements accurate only to a certain degree).

Likewise with purposes, desires and beliefs. A purpose to visit London and a purpose to learn to sky-dive are separate purposes, because they are not derivable from one equally accessible general purpose. But a purpose to write the first chapter of my book and the purpose to write the second chapter of my book are derivable from an equally accessible purpose to write my book; and that is why the simpler and more general description should be attributed to me, given evidence of my writing the first chapter and then the second chapter, in the absence of counter-evidence. We attribute to persons continuing accessible desires (varying with circumstances in a mathematically simple way, which give rise to purposes at appropriate times – for example, a desire to eat which increases when the person has not eaten for a long time, and decreases

after eating.) Likewise it is simpler to explain many of my beliefs by such general liabilities as the liability (at least with high physical probability) to believe what I am told and to acquire beliefs about the location of objects in my field of vision, than by separate liabilities for each belief. And powers, beliefs, purposes, desires and liabilities are readily accessible properties.

In summary, hypotheses of personal and scientific explanation (on the SPL account of the latter) are simpler if they postulate fewer substances, fewer (accessible) properties (including powers and liabilities), and mathematically simpler relations between them (including mathematically simpler numbers in the statement of these). These features of simplicity are features of the simplicity of the actual components (substances, properties, and the relations between them) of an explanation of some phenomenon independently of whether they are operative in other similar substances.

NATURAL THEOLOGY

Natural Theology of a probabilistic kind claims that the most probable explanation of the existence of the universe and its most general features is that they are caused by God. These most general features include the universal operation of simple laws of nature (that is, in terms of the SPL account, that every physical object behaves in exactly the same way codified in the simple 'laws' of nature), those laws and the initial (or boundary) conditions of the universe being such as to bring about the existence of human bodies, and humans being conscious beings, open to a finite amount of suffering and having some ability to bear it or alleviate it. (I mean by the 'boundary conditions' of the universe those general features of the universe which, in addition to those captured by 'laws' of nature, the universe would need to have, if it did not have a beginning, at all times if human bodies were to evolve – for example enough matter-energy.) These general features, the natural theologian's evidence or data, described in terms of the SPL account are the existence of a vast number of substances all behaving in the same simple way such as to bring about somewhere or other at some time or other subject to some suffering the bodies of conscious humans. Natural theology needs therefore to claim

that the hypothesis of theism (that there is a God) satisfies the criteria of correct explanation set out earlier better than does any rival explanation. As I wrote earlier, for very wide-ranging theories such as theism and any rivals, this will depend only on how well they satisfy criteria (1) and (3).

Criterion (1) is satisfied insofar as the evidence is probable if the hypothesis is true and improbable if the hypothesis is false. I have argued elsewhere at some length[9] that if there is a God, it is quite probable that he would bring about the existence of embodied humans in conditions such as we find on earth (including limited suffering and the possibility of bearing or alleviating it) and so that he would bring about those general features of the universe just described which are necessary conditions for the existence of such humans. The basic reason for this is that God being perfectly good will seek to produce good things; humans are good things of a unique kind having – unlike God – the power to make efficacious choices between (limited) good and evil. It is therefore quite likely that God will produce them. But humans can only make efficacious choices (ones that make a difference) if they live in an embodied state in an orderly universe where they can predict the effects of their actions, and that minimally involves a universe with many substances of few kinds (protons, electrons etc) with simple powers and liabilities. But of course there are innumerable other logically possible hypotheses which satisfy criterion (1) equally well, both hypotheses in terms of many or weaker deities, and scientific hypotheses to the effect that the initial conditions of the universe and its laws of nature are ultimate and have no further explanation (e.g. in terms of God bringing them about) and eventually cause the existence of conscious beings. My concern in this paper however is only to discuss how well different such hypotheses satisfy the other relevant criterion, criterion (3), the criterion of simplicity. I now proceed to inquire, in the light of my analysis of that criterion, how well theism satisfies the criterion of simplicity, and how well any rival hypothesis either of a personal or an inanimate (scientific) kind which satisfied criterion (1) to some significant degree would also satisfy the criterion of simplicity.

[9] See *The Existence of God*, chs 6-13, and the shorter and simpler book *Is There a God?* (Oxford: Oxford University Press, 1996), chs 4-7.

THE SIMPLICITY OF GOD

The simplest kind of explanation of the features which I have described will be in terms of some one first substance (whether some first chunk of matter-energy or a personal creator) which caused, in virtue of its powers and liabilities or its powers, beliefs and purposes, the multitude of substances of a very few kinds with their powers and liabilities. If the universe had a beginning, the 'first substance' would be a substance which caused the emergence and evolution of the universe into this multitude a finite number of years ago. But if the universe has always existed, then this 'first substance' would be one which everlastingly keeps this multiplicity of substances in existence with their powers at all moments in time. Theism (as understood by the Christian and similar religions) postulates a person as the cause of the universe and so provides a personal explanation of its existence. This person is often called 'God'; though in the case of the Christian tradition, we must regard the first substance as 'God the Father' who according to Christianity inevitably brings about from all eternity the other two members of the Trinity.[10] Other theistic religions do not of course have this feature. So can the traditional divine properties be construed in such a way that this unique personal substance, whom in future I will call simply 'God', is a simple substance and the simplest substance which can perform this explanatory role?

God is one person. So theism is inevitably a simpler theory than polytheism. To be a person at all, a substance has to live for a period of time, to have some power (to do intentional actions), some choice (whether free or not) of which actions to do, and some true beliefs. He will have to have some true beliefs about his intentional powers (what he can do); otherwise he will not be able to bring about anything intentionally. I shall assume that there cannot be a timeless person, and so that any

[10] A simple hypothesis is no less simple for entailing complicated consequences. Christianity claims that God the Father inevitably in virtue of his nature brings about the other two members of the Trinity, all of whom together constitute one God. (For an argument in justification of this Christian claim see my *The Christian God*, especially ch 8.) But I suggest that arguments to the existence of that one God must proceed via arguments to the existence of one person on whom everything else depends, and so to the existence of God the Father, whose postulated properties are the same as those attributed to the God of Islam or Judaism.

person who exists exists and so has his properties at moments of time.[11] If there could be such a person, the simplest kind of person would be an everlasting, omnipotent, omniscient, and perfectly free person, a person to whose length of life, power, true beliefs[12], and freedom of choice there are no limits; or rather no limits except those of logic, since any description of what all this amounts to has got to be free from contradiction. A person is everlasting if he exists at all times, omnipotent if he can do all actions, omniscient if he has all true beliefs, perfectly free if he is subject to no non-rational desires which influence how he chooses to act. The concepts of length of time, intentional power, belief, choice, and influence on choice are concepts of properties maximally accessible; we are more familiar with paradigm examples of these properties than of virtually any other properties. The concepts of 'all', or 'unlimited' (that is having no (zero) limits) are concepts far more accessible than concepts of particular large numbers. Hence the properties of everlasting life, limitless power, having all true beliefs, being subject to no causal influences on choice are far more accessible than the properties of living for a particular large number of years, having a particular large finite degree of power, having a particular large finite number of mostly true beliefs, and being subject to fairly few non-rational desires.

A person P is omnipotent at a time *t* iff he has the maximum degree of logically possible power. I suggest that that amounts to this: he is able at *t* intentionally to bring about any state of affairs which it is logically possible for anyone at *t* to bring about (and the description of which does not entail that P did not at *t* bring it about.)[13] However the notion

[11] For argument in defence of this claim see my *The Coherence of Theism*, revised edition (Oxford: Clarendon Press, 1993), 223-9; and *The Christian God*, ch 4 and pp. 137-44.

[12] It is generally agreed that knowledge is true belief not acquired by luck, although there are different views about what 'not acquired by luck' involves. I shall be arguing shortly that all the divine properties which I have been discussing belong to God essentially and so not by luck. So God's true beliefs will amount to knowledge. It will simplify the present discussion if I assume this already established.

[13] Philosophers have found it very difficult to analyse an intuitively simple concept of omnipotence (maximum logically possible power) in such a way as to avoid various paradoxes. For the history of attempts to analyse the concept of omnipotence, see Brian Leftow 'Omnipotence' in T.P. Flint and M.C. Rae (eds), *The Oxford Handbook of Philosophical Theology* (Oxford University Press, 2009). I hope that my analysis avoids all such paradoxes, but if it doesn't the concept is a simple one which makes clear the kind of qualifications which are necessary to avoid paradoxes.

of having an intentional power or an ability to do some intentional action is somewhat unclear. One says that someone has a power or ability when asleep, meaning by that that he could exercise it if he was awake and tried to do so. One might say that someone has the ability to speak French simply because he could learn it if he tried; or even that he has the ability to speak Gaelic for the same reason even if he does not believe that there is such a language as Gaelic. But clearly he has the power in its fullest form if he can exercise it at will, is conscious and knows that he can exercise it at will; only his will is preventing him from exercising it immediately. So omnipotence, being the maximum possible degree of power should be construed as equivalent to the following: a person P is omnipotent iff he is consciously aware of all the states which it is logically possible for anyone to bring about at *t* (and the description of which does not entail that P did not bring it about); and if he chooses at *t* to bring about any such state, it happens. Given the logical impossibility of backward causation, an omnipotent person cannot affect the past[14] or the truth of logically necessary truths, which I assume to include the fundamental moral truths.[15] Hence the choices of an omnipotent person can affect only contingent future states.

A person P is omniscient at *t* (in the most natural sense) iff he knows all propositions true at *t*. I have spelled out being 'perfectly free' as being subject to no non-rational desires influencing his choice. Clearly any agent who makes a choice is influenced by the nature of that choice, what it involves, and so by considerations of reason, the apparent goodness or badness of the action. To believe that some action is good to do necessarily gives you a desire to do it in proportion to its believed worth. The apparent good motivates, and the apparently better motivates more. What is ruled out by 'perfect freedom' are desires to do an action which is apparently bad, or ones which are stronger than its apparent worth would motivate. Hence a person who is both omniscient and perfectly free will be moved to do an action in proportion to its actual goodness. God being perfectly free will set himself to do what he believes best; be-

[14] That is, he cannot affect 'hard facts' about the past, these being ones whose truth conditions are solely in the past.

[15] For my reasons for this assumption see for example pp.151-55 of my 'What Difference does God make to Morality' in R.K.Garcia and N.L.King (eds), *Is Goodness Without God Good Enough?* (Lanham: Rowman and Littlefield, 2009).

ing omniscient (to the extent of knowing all necessary truths and truths about the past) he will have true beliefs about what is the best; and being omnipotent, he will succeed in doing it. So he will always do the best action where there is one. If in some situation there is no unique best action (i.e some incompatible action would be equally good, or there is an infinite sequence of incompatible actions, each less good than the next member of the series), then God cannot do the best. But his perfect freedom will lead him to get as close to that as the nature of the good will allow; and that means that if there are several equal best actions he will do one of them, and in the infinite series situation he will do some good (and no bad) action. So God will be as good as it is logically possible to be, and we may call that 'perfectly good.'[16] So it turns out that God being 'perfectly free' in my sense has the consequence that while he can do evil, inevitably he will never exercise the choice to do so. In another perhaps equally natural sense of 'perfectly free' a perfectly free person would be one who could make any logically possible choice including doing evil. So why my sense of 'perfectly free' rather than the rival sense? Because it is a simpler sense. 'Perfect freedom' in my sense is simply the absence of properties, non-rational desires; 'perfect freedom' in the rival sense would be a complicating feature of God, because it would involve his being influenced by non-rational desires which alone make possible a choice of evil. So I stick with my sense.

Frequently, indeed I would have thought in most situations, there will be no unique best action for God to do. Surely, even if there were no incompatible better action for God to do, it would have been an equal best action for God to make (the universe begin in such a way as to cause) the planet Uranus to rotate in the same direction as the other planets as to rotate in a different direction. And however many planets (in our galaxy or in some distant galaxy) containing living organisms God makes, it would be better if he makes one more. And so on.

But these definitions give rise to a problem, that while there could be an everlastingly omnipotent and perfectly free person, and there could be an everlastingly omniscient person, everlasting omnipotence plus

[16] William Rowe (among others) has argued that unless God always does an action better than any incompatible actions, God cannot be 'perfectly good', and so there cannot be a God of the traditional kind. See his *Can God be Free*? (Oxford: Clarendon Press, 2006). Like many others, I find this view highly implausible.

perfect freedom is incompatible with everlasting omniscience. In the absence of rational considerations relevant to his choice and in the absence of any non-rational desires influencing that choice a God perfectly free and omnipotent would have a free choice of which state of affairs to bring about before the time when the relevant state of affairs came about. But then he would have been free to choose to make any earlier belief about how he would choose false, and so only by cosmic luck could all God's beliefs have been true, and beliefs acquired by luck (see note 12) do not constitute knowledge. So omnipotence plus perfect freedom are incompatible with omniscience understood in the obvious way. On the other hand a more restricted kind of omniscience, knowing all truths about the past and all necessary truths (including the necessary moral truths[17]) is not merely compatible with omnipotence plus perfect freedom but entailed by them. This is because for any past state or any necessary truth, there are future states of affairs which can be defined by its relation to them, from which it follows that if God is to have the choice of bringing about these states, he needs to know all necessary truths and all truths about the past. For example, if God is to know that he can choose now to bring about a third world war or a prime number of planets greater than 7, he has to know that so far there have been only two world wars, and that there is a prime number greater than 7. Since omnipotence entails knowledge of moral truths, omnipotence plus perfect freedom entail perfect goodness; unrestricted omniscience is not necessary for this. Omnipotence is the simplest degree of one property necessary for a person, perfect freedom is merely the absence of certain complicat-

[17] Contingent moral truths are ones made true by a conjunction of a necessary moral truth and a contingent non-moral truth. For example, it would be contingently true that I ought to pay you £20 if I have promised to pay you £20 (contingent non-moral truth) and people ought always to keep their promises (necessary moral truth). The contingent non-moral truths which, together with necessary moral truths, create contingent moral truths are normally truths about the past – truths about past commitments or truths about what past evidence shows is likely to happen in future. Hence a being who knew all truths about the past and all necessary truths would normally know all moral truths about what would be good for him to do now. But insofar as whether an action available to such a being who is also omnipotent and perfectly free (in my sense) is good now depends on what is yet to happen (and not merely on what present evidence shows about what is likely to happen), then such a being would predetermine the future in order to enable him to do what is good now. Hence only the kind of omniscience entailed by omnipotence is necessary for God's perfect goodness.

ing properties (causal desires) which yield a vast degree of knowledge of a simple although not the simplest kind. ('Simple' because it is all the knowledge possible for a perfectly free and omnipotent person.) On the other hand a person omniscient in the full sense which includes knowing everything which he would choose to do, could not be perfectly free. Indeed he could not be free to any degree. He would be causally predetermined at every moment of time to do every future action which he would do on each occasion when there is no unique best action for him to do. That has the consequence that on each such occasion he would be subject to a particular non-rational desire (determining and not merely influencing him) to do this rather than that action (when there would be no reason for doing this action rather than that one.) This would make God a very complicated person. So which is the way of understanding omniscience which makes an omnipotent being the simplest kind of person? The answer is clearly the restricted kind of omniscience[18] which allows him also to be perfectly free.

The other divine properties (understood in natural ways) follow from the ones analysed above.[19] For example, since God is omnipotent, everything else that comes to exist comes to exist because he causes it or allows it to exist; hence he is in a natural sense creator of all that is. Being omnipotent, he can make things happen anywhere and learn about things anywhere without depending on intermediate causes; so in a natural sense he is omnipresent, and so is not tied down to a body, and so is not physical.

God's omnipotence is a power and so an intrinsic property. God's omniscience (in the restricted sense) is not a power or a liability[20] but a categorical state necessary for omnipotence; and it might seem that it is

[18] For fuller discussion of this see *The Christian God*, 150-151.

[19] In deriving this restriction I am following the convention of calling a belief about the future true now iff in the future it will be true, even when its truth is not yet inevitable. This is a convention which we do not always follow when we talk of a belief 'not yet' being true. But if we do not count a belief whose truth or falsity is not yet inevitable, as not now being either true or false, then God's omniscience can be construed simply as having all true beliefs. I do not however think that it is any less simple to understand only beliefs whose truth value is inevitable as having a truth value, than to follow our more normal convention.

[20] Thus Aquinas: 'Knowledge in God is not ... a disposition *(habitus)*', *Summa Theologiae* Ia.14.ad 1.

not an intrinsic property of God but a relation to other entities because it involves possessing (having within oneself) certain objects – all true beliefs. If that was the right understanding of omniscience, then of course God would certainly not be simple. So all depends on what it is to have a certain belief; and it is not, I suggest, possessing an object within oneself. I suggest that having a belief that p is an intrinsic property and one too primitive to be defined. But we can show what we are talking about when we talk about S's belief that p by saying that it is the sort of thing S would acquire in a certain way (e.g. by S seeing that p, or by someone telling S that p) and which makes a certain kind of difference to S's behaviour (e.g. if it follows from p, that the way to get x is to do A (and not B), and it follows from not-p that the way to get x is to do B (and not A), and S has the purpose of getting x, he will do A.) We humans can have beliefs about matters about which we are not thinking. I have beliefs about what I did when young about which I am not thinking, but they are my beliefs in that – if prompted – I could bring them to consciousness. It is simpler to suppose that all God's beliefs are currently before his mind[21]; that he is currently aware of all his knowledge and so beliefs involved in his omnipotence follows from my account of his omnipotence, and these are the only ones which I am suggesting that he has. Beliefs need not be put into words; children have beliefs before they can put them into words, and when they acquire language they report that they had those beliefs. So I think that the best analogy for God's beliefs are the beliefs we acquire when we look at a scene before our eyes. Merely by looking we acquire innumerable beliefs about what objects there are, where they are, and what they look like. We are aware of these beliefs, but not as linguistic entities, and not as the brain states which causally sustain the beliefs in us. The beliefs are there in a fused pre-linguistic state out of which we can – if we choose – separate individual beliefs and put them into words (e.g. that 'there is a tree outside the window'.) We see things and acquire beliefs about them both as they are now, and (when we look at the stars) as they were thousands of years ago. Seeing involves categorizing: in seeing a tree I do not merely have a visual impression caused by a tree, but I see an object outside the window as a tree. And seeing an object thus

[21] Thus Aquinas: '[God] sees everything at once and not successively', *Summa Theologiae* Ia.14.7.

categorized inevitably involves seeing its powers – the tree has the power to grow, to resist pressure and so on. God's beliefs are in this way just like the beliefs of which we are aware, but they concern the whole of the universe and the insides of things us well as their outsides. He sees things as they are and as they were. While our beliefs may come to us by different causal routes, it is simpler to suppose (as the traditional picture supposes) that God's beliefs come to him only by one route, directly. So too God's wider fused pre-linguistic state of belief is one integrated state of himself. It does not consist of separate items within himself, but it is a property of himself.

God's perfect freedom is, to repeat, merely the absence of the property of being influenced by non-rational desires. It is always simpler to postulate an absence than a presence. It is simpler to suppose that God has the divine properties discussed so far essentially; otherwise it would be a vast accident that God continued for all time to exist and have these properties. By contrast an ordinary human person, although he needs some power etc in order to exist, does not need to have some particular amount of power in order to be the particular human he is. And it is simpler to suppose that God does not have thisness, which would be a particularizing feature additional to his properties. (A substance has thisness iff there could be instead of it another substance with all the same properties, intrinsic and relational, as it. God would not have thisness iff there could not be instead of the actual God a different God with all the same properties, and so all the divine properties discussed so far.) But a person who does not have thisness, and for whom not merely having power (plus freedom) but having a particular amount of power (plus freedom) is essential to his being the person he is, is a person unlike any other persons with whom we are familiar. To call such a 'person' a 'person' is to use the word in a somewhat analogical sense, but one significantly similar to the ordinary notion (like an ordinary person, God has purposes and beliefs) for God to count as a person in a wide sense.

My understanding of the divine properties is traditional, except in respect of his being everlasting rather than timeless, and being omniscient only in the restricted sense.[22] My arguments however have assumed that

[22] That God is timeless has been the dominant theological view from at least the fourth century onwards. However, in my view the biblical authors thought of God as everlasting, and God's eternity has not been the subject of dogmatic definition.

there cannot be a timeless person, and purport to show that the simplest kind of God would be omniscient only in the restricted sense. Persons other than God have more limited degrees of power and knowledge; their purposes are not influenced solely by their beliefs about what is good but by various desires for particular states of affairs whose strength is not always aligned with their beliefs about the goodness of these states. Other persons are very particular limited persons. God, defined (in the way I have spelled out) as an essentially everlasting omnipotent perfectly free person lacking thisness is a very simple substance, the simplest kind of person whose existence could explain the existence of the universe and its very general features. Hence the hypothesis of theism so construed is simpler than any hypothesis explaining the very general features of the universe in terms of a god construed in some other way. The sense in which I am claiming that God is simple is not quite the same as the sense in which later medieval theologians claimed that God is simple but it is not too far distant from it, at least on a familiar account of what that sense was.[23]

Nelson Pike's well-known book *God and Timelessness* (London: Routledge and Kegan Paul, 1970) concluded with his remark that he had not been able 'to find any basis for [the doctrine of divine timelessness] in biblical literature or in the confessional literature of either the Catholic or Protestant churches'. And it is disputable whether even all western theologians of the high middle ages were committed to an explicit doctrine of divine timelessness – see R. Fox, *Time and Eternity in Mid-Thirteenth-Century Thought* (Oxford University Press, 2006). As regards omniscience, although there are a number of biblical passages which – read in their natural sense – do imply that God does not know infallibly what God or humans will do (e.g. Genesis 6:6, Jonah 3:10, and Revelation 3:5 which implies that God may change what is written in the Book of Life), most biblical passages imply that God is omniscient in the more natural sense; and the vast majority of subsequent Christian tradition is committed to that view. However this matter has not been the subject of any definition binding on Orthodox, or any definition which might be regarded by Catholics as infallible apart from the statement of the First Vatican Council that 'all things are open and laid bare to [God's] eyes, even those which will be brought about by the free activity of creatures'. (See N.P.Tanner (ed.), *Decrees of the Ecumenical Councils* (London: Sheed and Ward, 1990), 806.) However, the Council authorised no anathema against those who held a rival view; and (as far as I can see) this view is not mentioned in the 1992 *Catechism of the Catholic Church* (London: Geoffrey Chapman, 1994) which 'aims at presenting an organic synthesis of the essential fundamental contents of Catholic doctrine' (p.9).

[23] For that familiar account, which may only be applicable to the later medievals, see my *The Christian God* (Oxford: Clarendon Press, 1994), ch 7. For these thinkers God's

The enormous simplicity and so prior probability of hypotheses postulating omni-properties can be illustrated by a simple scientific example. Newton's theory of gravitation had three laws of motion (which are largely in effect definitions) and its law of gravity affirming of every given body, that it has the power to attract every other body in the universe with a force proportional to mm'/r^2 and the liability always to exercise that power – from which it follows that each body also has the liability to be attracted by every such body with that force. These extensive powers and liabilities belong to the tiniest fundamental particle. So the power of such a particle extends over physical objects to the ends of the universe and that covers quite a range of the extent of God's power, though of course in no way comparable to it in strength; and its sensitivity to other physical objects, which we may compare to God's knowledge of them, also extends over quite a range of what God knows about. If there was no quantum indeterminism and we could make measurements with infinite accuracy, merely measuring the movements of one particle could tell us an enormous amount about the distribution of massive bodies throughout the universe; and measurements on several particles would tell everything about this (given the contingent truth of Newton's theory). The considerable probability of Newton's theory on the evidence available in 1689, far greater than that of the infinite number of rival theories which could have been postulated and would have predicted the evidence equally well, derives from the enormous simplicity of its omni-properties, ones which have considerable similarity to the omni-properties which theism attributes to God.

THE SIMPLEST INANIMATE EXPLANATION OF THE UNIVERSE IS LESS SIMPLE THAN GOD

Could there be a physical object as simple as God (understood in the way developed above) which could provide an inanimate explanation of the

simplicity was a matter of his not having parts, and all his essential properties being the same as each other and the same as God. I claim that God has no parts, and that (not having thisness) he is whatever instantiates his essential properties. I claim that God has just one essential property – everlasting omnipotence – together with the absence of a property. For a rather different account of Augustine's views on God's simplicity, see Brian Leftow, 'Divine Simplicity', *Faith and Philosophy* 23 (2006), 365-380.

existence of a universe with the very general features described above? The normal kind of 'first substance' postulated by physicists is an extended substance – a 'vacuum state' or a very compressed chunk of matter-energy. But such a state has parts and so is less simple than God. But physical cosmology could postulate one unextended substance, a particle, as that from which all else evolved (or on which all else depended, if the universe did not have a beginning). It would need to be a physical object of a certain kind in having certain properties, powers, and liabilities. The normal kind of physical substance would have powers of particular mathematical quantities (such as the attractive and repulsive powers of mass and charge) and liabilities to exercise them under certain physical conditions. These would need to be fairly specific powers and liabilities (some mathematically precise mass and charge for example) if they were to make it probable that it would bring about a universe with substances of few kinds with the same simple powers and liabilities as each other, of a human-body producing kind; and this specificity would make the hypothesis of the existence of such a substance more complicated than the hypothesis of theism. The singularity of the Big Bang would have to have all or at least (given a certain amount of indeterminism) most of the details of the future development of the universe built into it.

But could we not instead merely suppose that this physical object had the power to produce a good universe, and the liability always to exercise that power, and that would explain why we exist (because of the goodness of humans existing) and the other general features of our universe (the details of which were not predetermined but constitute one of the ways in which the universe would be good)? The power could be exercised by producing an appropriate sort of Big Bang. Given my earlier claim that laws of nature are to be analysed in terms of the powers and liabilities of substances, this is the nearest we can get to John Leslie's hypothesis that goodness has a propensity to exist.[24]

However even if such a physical object could do the explanatory work, it would not be nearly as simple as God. Like God, it could be essentially unextended, everlasting and lack thisness. The physical object's power to create the good would however be a limited power and so less simple than omnipotence (which can only be had by a being which can

[24] See his *Value and Existence* (Oxford: Basil Blackwell, 1979).

choose to act). And it would also need another power, the power to prevent any other substance bringing about the bad, and the liability always to exercise that power. Theism does not need a corresponding second power and liability, because it follows from divine omnipotence (plus perfect freedom) that every other substance only exists and produces effects insofar as God allows it to do so. The liabilities always to exercise these powers would be properties of the physical object additional to its powers. Yet being physical (and so public), it couldn't simply have *those* liabilities. A physical object (unless it is the only physical object) must have a location – there must be somewhere where it is (relative to other physical objects), and that means that there must be some detectable effects of it in one place which are not detectable elsewhere. (If a physical thing is here rather than there, here and not there is where it impedes the motions of other objects, or can be detected by a Geiger counter, or from where it emits light, or whatever.) So if the first physical substance were everlasting and so existed alongside other physical objects, its liability to produce effects must be manifested more or differently in one place than others, and that makes that property not a very simple property. (If however it ceased to exist after the creation of the physical universe, and so did not exist alongside other physical objects, it would not be everlasting and so would be less simple than God for a different reason.) So the properties of a first physical object (additional to being everlasting, unextended, and lacking thisness, which God also has) would make it far from being very simple. In addition to the properties common to both God and a first physical substance, God would be merely omnipotent and perfectly free, but the latter, as we have seen, is simply the absence of something – and absence is always simpler than presence. I conclude that the powers and liabilities which we would need to ascribe to the single particle hypothesis (the simplest kind of inanimate explanation of the orderliness of the universe there could be) would be less simple than the properties of God – essential everlasting omnipotence (plus perfect freedom), from which all the other properties follow. So the hypothesis of theism provides a simpler explanation of the very general features of the universe than does any inanimate explanation. And it will be evident that the primary reason for that is that moral beliefs motivate; and so a conscious being needs less in the way of properties than does an inanimate one to cause the same effects.

And even if the hypothesis of one physical first substance were just as simple as the hypothesis of God, its propensity to create the good could not explain certain more particular features of our universe which would only be good if they were brought about by God. Vast numbers of religious experiences apparently of God would not be good unless there is a God. Interactive prayer would be deceptive. And our universe contains so much suffering that it would probably be overall a good universe only if its creator were to suffer with his creatures[25]; and that is something that only a person can do. I conclude, for two separate reasons that God (understood in the way which I have developed) is the simplest kind of cause of our universe there could be.

[25] See my *The Resurrection of God Incarnate* (Oxford University Press, 2003), 44-45 for the obligation on a creator to share the suffering of those whom he causes to suffer for the sake of some great good. The argument which I use this point to develop here is that a good-producing physical object would have to produce a world with less suffering than would a good God who is prepared to share that suffering with creatures whom he makes to suffer for the sake of a great good.

GOD AND CHRISTIANITY ACCORDING TO SWINBURNE

JOHN HICK

Fellow of the Institute for Advanced Research in Arts and Social Sciences
University of Birmingham

Abstract. In this paper I discuss critically Richard Swinburne's concept of God, which I find to be incoherent, and his understanding of Christianity, which I find to be based on a pre-critical use of the New Testament.

Richard Swinburne has written extensively in both the philosophy of religion and latterly philosophical and biblical theology. In this paper I discuss a central theme in each. The difficulty in assessing Richard Swinburne's philosophy of religion, and the philosophical theology into which it merges, is that it is spread over a number of books, several now in revised second editions, so that it would definitely require a whole book to discuss it fully. Probably one day someone will do that, but the present article is much less ambitious. I want to look briefly and critically at Swinburne's concept of God, and at his understanding of Christianity.

The most obvious feature of Richard Swinburne's approach to the philosophy of religion is its highly abstract nature. He is concerned above all with religious beliefs or propositions, the probability of their being true and the rationality of believing them. Religious beliefs are beliefs about 'transcendent reality, including beliefs about whether or not there is a God or an after-life, beliefs about what properties God has (what God is like), and what actions He has performed'; and we want 'to have beliefs on these matters as probably true as we can get'[1]. In this he is part of a very prominent contemporary group of philosophers of religion which also includes Alvin Plantinga and the many influenced by him.

[1] Richard Swinburne, *Faith and Reason*, 2nd ed. (Oxford: Clarendon Press, 2005), 83.

They all also happen to have in common a presupposed highly conservative Christian conviction, though this is not something that I am concerned with here.

Since Swinburne has referred to one of my own proposals, I can first illustrate the excessively propositional nature of his approach from the way in which he summarises my position: 'that the ways of living commended by the major religions are of equal moral worth, and that the creeds of these religions are best understood as expressing the same eternal truth with the aid of different myths'[2]. It is his propositional approach that leads Swinburne to the notion of 'the same eternal truth', an idea that fundamentally misrepresents my position. My 'pluralistic hypothesis' is that the major world religions are very different human responses, formed in different culturally conditioned human terms, to the same ultimate transcendent reality, which can be called the Ultimate Reality or the Real. This is in itself transcategorial, beyond the scope of our human conceptual systems; and the beliefs of the different religions describe their own different experiences of the impact upon them of the universal presence of the Real. My point here is that, for me, it is not a *truth* but a *reality* that is eternal and ultimate. It is only when we come to particular doctrines that myths come into the picture. I suggest, for example, that the notion of divine incarnation is metaphorical and that the Christian doctrine of divine incarnation in Jesus of Nazareth is therefore mythological, a myth being an extended and often highly developed metaphor, often developed into a story.

GOD

Swinburne says that (1) 'God is a personal being – that is, in some sense *a person*. By a person I mean an individual with basic powers (to act intentionally), purposes, and beliefs'[3]. Further, says Swinburne, God is a unique individual, because he is (2) omnipotent – 'he can bring about as a basic action any event he chooses'[4]. (3) He is omniscient: 'whatever

[2] Ibid., v.

[3] Richard Swinburne, *Is There a God?* (Oxford University Press, 1996), 4, italics in original.

[4] Ibid., 5.

is true, God knows that it is true'[5]. And (4) he is '*perfectly free*, in that desires never exert causal influence on him at all'[6]. And (5) eternal: 'he exists at each moment of unending time'[7]. In addition, God is (6) bodiless[8], and (7) omnipresent[9]. Finally, (8) God is perfectly good[10]. This, in brief, is Swinburne's concept of God.

Some comments on some of these proposals. (1) God is in some sense a person. But in what sense? Surely, if this is to mean anything clear and distinct it must mean that God is literally a person. So Swinburne must mean that God is a person like ourselves, except for being infinite in power, knowledge, extension in time, and except also for being perfectly free and omnipresent and good. As Swinburne says, 'God is supposed to be like us, in having basic powers, beliefs, and purposes – but ones very different from ours'[11]. But does the idea of an infinite person make sense? We know what it is to be a person because we are ourselves persons. And to be a person is to be a particular person, distinct from other persons, each with our own boundaries. When two people are interacting with each other as persons, this is only because they have their own individual borders – otherwise they would not be two distinct persons. In other words, personhood is essentially finite, allowing for the existence of other persons. And so an infinite person is a self-contradiction. God cannot be both a person and infinite.

How might Swinburne reply to this? Possibly like this: God, the infinite person, allows finite persons to exist in a created realm, distinct from himself. So God is infinite, and we are finite. But this would not do. If God is omnipresent he must be present throughout the created realm. There cannot be both an omnipresent God and an area in which he is not present. And if, in the created realm, God interacts with finite persons (as recorded in the Bible), then both God and the other persons must have their individual borders. So Swinburne would have to defend the notion of an infinite person in some other way.

---

[5] Ibid., 6.
[6] Ibid., 7 (italics in original).
[7] Ibid., 9.
[8] Ibid., 10.
[9] Ibid., 10.
[10] Ibid., 12.
[11] Ibid., 5.

(4) Why does God's freedom require that desires have no causal influence on his actions? It is part of Swinburne's definition of a person that a person can act intentionally and purposefully, though, as he also says, 'God acts only in so far as he sees reason for acting'[12]. But what sort of reason might God have? Surely, only that he desires something to be the case. For example, God, in his goodness, decides to create a universe. He decides to create because, being good, he wants, i.e. desires there to be a created universe. Further, God might desire to love and to be loved and then decide whether to act on this desire by creating beings for him to love and to love him. But why would he ever do this unless he desires it? Surely if God is never caused by desires to act, he will never act. A perfect freedom which consists in being influenced by no desires would be a perfectly empty freedom. Reason without desires which one can decide whether or not to fulfil would never lead to anything. The picture of God as desireless reason creating a universe is incoherent. Without the divine desire to create there would be no creation.

However at this point Swinburne will perhaps say that God creates because it is good that he should create, and God always does what is good, in this case supererogatively good, i.e. good but not obligatory. But if creating is not required of God, then he does so because he wants to. There must be innumerable good but not obligatory things that God could have done but has not done – such as creating a different but equally good universe, or a million such, or within this universe, additional layers of angelic beings. So why has God done some but not other of the good things that he might decide to create? Must it not be because he prefers, i.e. wishes, to create what he has chosen to create?

(3) 'Whatever is true, God knows that it is true'. It is propositions that are true or false. So whatever propositions are true, God knows that they are true. But does God really think in propositions? Is this not a gratuitous assumption, arising from a presupposed anthropomorphic conception of God? It seems to me quite arbitrary, a picture of God as the Great Analytical Philosopher – created in the human philosopher's own image. Indeed, a great deal of Swinburne's thought about God rests upon this anthropomorphic image of God.

Further, in ordinary life, whilst we do know many propositions, our primary awareness is of things, both individually and, more usually, as

[12] Ibid., 43.

components of situations. When we see a table or a tree or a human body or a crowded street, or anything else, we see it, rather than 'knowing that it is true that there is a table'. Seeing is a very complex process, but nevertheless it is perception, not a knowing of propositions. And so God's omniscience will consist in his being simultaneously aware of every*thing*, not primarily in his knowing that hundreds of millions of propositions are true.

(5) God is eternal in that he exists at each moment of unending time. It would seem, then, that time is co-ultimate with God: time exists unendingly independently of God, who, as eternal, exists throughout it. For it is an essential attribute of God that he is eternal. So we have two ultimates: God and Time. On this view of time it would even be possible for time to exist without there being a God – though because he has necessary existence, this is not the case. So there cannot be a God without there also being Time, and Time is, like God, an ultimate brute fact.

There seems also to be, for Swinburne, a third ultimate in addition to God and Time. For, Swinburne says, God is perfectly good. 'His being perfectly good follows from his being perfectly free and omniscient. A perfectly free person will inevitably do what he believes to be (overall) the best action and never do what he believes to be an (overall) bad action'[13]. Socrates asked, do the gods love an action because it is good, or is it good because the gods love it? Translating this into monotheistic terms, Swinburne's answer is that God does what is good because it is good, rather than its being good because God commands it. In other words, morality is independent of God. For, 'if there are moral truths – truths about what is morally good and bad – an omniscient person will know what they are'[14]. And so, he says, 'I side with [Aquinas and Scotus] in holding that there are moral truths independent of the will of God. God can only enforce them, not alter them.'[15]. It follows that God can have moral obligations: 'God before he creates any other persons has no obligations, though it is a supererogatory good act for him to create many other persons including humans. If he does create them, he will then incur certain obligations towards them. Exactly what those are may be disputed, but the Christian tradition has normally maintained, for

[13] Ibid., 12.
[14] Ibid., 13.
[15] Ibid., 15.

example, that, if God makes promises to us, he is obliged to keep them'[16].

The third ultimate is thus morality – God, Time, and Morality. One is reminded at this point of A.N. Whitehead's metaphysical system, in which there are also three ultimates: God, Creativity, and Matter. This in turn reminds us that Swinburne is (perhaps without being aware of it) in the business of building his own speculative metaphysical system, and in this respect is unlike most of the other contemporary analytical philosophers I mentioned earlier.

Turning now to the probability of God's existence, Swinburne says that 'If, as theism maintains, there is a God who is essentially eternally omnipotent, omniscient, and perfectly free, then he will be the ultimate brute fact which explains everything else'[17]. In the light of Swinburne's own explanations we must amend this. According to him, there are three ultimate brute facts. For God cannot be eternal without Time, since to be eternal is to exist throughout all time. And he cannot be omniscient without Morality, since to be omniscient is to know all true propositions, including the truths of morality – such as that it is a supererogatory good deed to create other persons. The ultimate brute fact is thus a complex of God, Time and Morality.

However Swinburne's central argument for there being a God is that God is the simplest possible explanation of everything else. 'It is extraordinary that there should exist anything at all. Surely the most natural state of affairs is simply nothing: no universe, no god, nothing. But there is something . . . If we can explain the many bits of the universe by one simple being which keeps them in existence, we should do so – even if inevitably we cannot explain the existence of that simple being'[18].

There are two problems here. The first is that, according to Swinburne's Christianity God is not simple but is a Trinity of Father, Son, and Holy Spirit, and obviously a divine trinity is not maximally simple. But even if we waive this – though I don't see how Swinburne can waive it – God may be simple in himself, but not as the ultimate brute fact which explains everything else. For he cannot exist except as part of a complex of God, Time and Morality. This complex is less complex than the cre-

[16] Ibid., 16.
[17] Ibid., 19.
[18] Ibid., 48-9.

ated universe, which is virtually infinitely complex. So it may be argued that we should explain the more complex by the less complex. But this is a much weaker argument than the one Swinburne intends – that the ultimate brute fact is simple and thus provides the simplest possible explanation of the universe.

Further, if it is right to look for the simplest available brute fact, this is not the complex of God, Time and Morality. According to the Big Bang theory of the origin of the universe it began with the densest possible particle of matter, something as simple as matter can be. It seems that even this cannot have been absolutely simple. For the infinite complexity that came about with the evolution of matter through immense periods of time required some slight imbalance or complexity in the expanding universe, which must have been prefigured in the original particle. But that particle will nevertheless have been considerably less complex than the God-Time-Morality complex. So is not the original particle at least as good a candidate for the position of ultimate brute fact? Or indeed an even better one?

In short, it seems to me that Swinburne's argument for God – whose existence turns out to involve the co-existence of Time and Morality – is far from persuasive. If the Big Bang theory is correct, the original particle, and the universe which it has produced, is much more likely to be itself the ultimate brute fact. But perhaps the Big Bang theory is mistaken and the universe consists instead in a beginningless series of expansions and contractions. In that case the universe, in the enlarged sense of this oscillating series, will be the ultimate brute fact. But neither of these possibilities takes us beyond the physical universe to a God.

I conclude that Swinburne's concept of God is full of serious problems, and his basic argument for God's existence no more probative than all the other 'theistic proofs'.

CHRISTIANITY

For Swinburne, Christianity, like other religions, is a set of beliefs, a creed, together with a life style.[19] By a creed, in this context, he does not mean

[19] Richard Swinburne, *Faith and Reason* (Oxford: Clarendon Press, 2005), 161.

a formula – such, for example, as the Nicene creed – but a coherent body of beliefs.

In an important chapter of *Faith and Reason* (the second edition) Swinburne compares the creeds of different religions and concludes that the Christian creed is the most likely to be true. Comparing Christianity with the other 'Abrahamic' faiths of Judaism and Islam, he points out that the main difference between their understandings of God is the affirmation or denial of the doctrine of the Trinity. He then considers the probability of this. He claims that there is an *a priori* probability that a God would be 'tripersonalized':

> I believe that there are good a priori arguments in favour of the doctrine of the Trinity. But they were not available until that doctrine had become discussable by being part of the Christian Creed – they were, to my mind, first put forward in a satisfactory way by Richard of St Victor in the twelfth century. But they are arguments of some subtlety, and all Christians before Richard and almost all Christians after Richard needed revelation…to assure them of the truth of the doctrine.[20]

He refers in a footnote to his discussion of the a priori argument in his *The Christian God*. But I cannot in this article pursue his thought through other parts of his oeuvre. So at this point I simply note that he claims that there are good a priori arguments for the trinitarian nature of God although he does not present them here.

Swinburne then proceeds to the claimed revelation in Christ. What, he asks, is the a priori probability that a good and loving God would become incarnate on earth? It has often been argued (by many writers, including myself[21]) that the properties of humanity and deity are such that they cannot be combined at one time in a single individual: no one person can be, at the same time, omnipotent but not omnipotent, omniscient but not omniscient, omnipresent but not omnipresent, infinitely good but not infinitely good, creator of the universe but not creator of the universe. But, as Swinburne says, 'If Christianity is to be taken seriously, it has to be shown first that it is logically possible that God should

[20] Ibid., 235.

[21] John Hick, *The Metaphor of God Incarnate* (London: SCM Press, 1993).

become incarnate', referring in a footnote again to his *The Christian God,* and continues 'Let us suppose that that is shown'[22].

He does however list 'three reasons why, in virtue of his perfect goodness, God could be expected to become Incarnate:

> to make available atonement for our sins; to identify with our sufferings; and to reveal truths to us. How likely is it that God would become Incarnate for these reasons? A good God would certainly want to forgive the sins of His creatures, but He would also want them to take these sins seriously by asking God to accept a serious act of reparation for those sins. Yet every human has sinned and owes so much of his life to God anyway in gratitude for God creating him, and is inclined not to fulfil even minimum obligations. So none of us is well situated to make a proper reparation to God for our sins, let alone the sins of others. A good God might well be expected to help us by Himself making the atonement available . . . through coming to Earth and living a perfect human life. . . . We may reasonably think that, given the extent to which God (if there is a God) makes humans suffers, albeit for good reasons, the point has come where it is not merely good but obligatory that He should share that suffering. If that is so, then (since a perfectly good God will always fulfil his obligations), it follows that it is not merely probable but inevitable that God should become Incarnate for this reason. . . .[23]

All this seems to me extremely dubious. If a good God wants to forgive our sins, why should he require 'a serious act of reparation for those sins'? If he wants to forgive us, let him do so. Jesus taught us to pray, 'Heavenly Father . . . Forgive us our sins as we forgive those who have wronged us'. No act of reparation is expected, no atoning sacrifice required. To forgive those who have wronged us is itself a life-changing, a redeeming, act. So I find no force in Swinburne's first reason for God to become incarnate.

The second reason, to share our human suffering, is more plausible. Certainly Jesus suffered in many ways, and particularly in his excruciatingly painful death on the cross. But how does this benefit us? Swinburne would say that it shows us that, although suffering is inevitable in the world as God has created it, God sympathises with us and shows this by visibly sharing our human suffering. And we know this because we know that Jesus, who was crucified, was God incarnate. But who are the 'we'

[22] Richard Swinburne, *Faith and Reason* (Oxford: Clarendon Press, 2005), 235.
[23] Ibid., 235-6.

who know this? Seriously believing Christians who are comforted by the thought of God's suffering in Jesus constitute a very small minority of mankind. (By no means all the millions of inhabitants of officially Christian countries can count as seriously believing Christians). But this is the sort of concrete consideration that Swinburne does not notice, dealing as he does in pure theory and logic. So if God really wishes to share our human suffering, it is not nearly enough for him to become incarnate in only one individual, Jesus, at one time. If incarnation is his chosen method, he would need to become incarnate in a vast number of individuals in every part of the world. But Christianity does not teach this, but on the contrary would regard it as a heresy.

Swinburne's third reason for the Incarnation is to reveal truths to us. He does not say 'to reveal new truths', presumably because he knows that Jesus did not reveal any new truths: his teaching about God and his moral teaching, were already present in Judaism, and the Golden Rule, to do to others as you would have them do to you, is taught in all the major religions. So there was no need to become incarnate to teach what was already known within the people within whom he became incarnate, or to give moral teaching that had already been given in the religions that began before Christianity. So this is at best a very weak reason for divine incarnation as Jesus of Nazareth.

In arguing that the Christian creed is very likely to be true (and more likely than all other creeds) Swinburne now appeals to the biblical evidence. Here he is at his weakest. He treats the New Testament evidence selectively, ignoring everything that tells against his desired conclusion. Speaking of Jesus, he says:

> There is evidence to be expected if he founded a Church and taught that his life and death provided atonement for our sins – for example his saying that his life was a 'ransom for many'. There is some evidence of a kind to be expected if He believed himself to be God, and some evidence of a kind to be expected if the teaching of the later Church about this and other matters was a continuation of the teaching of Jesus.[24]

There is *some* evidence of all this; but it is heavily outweighed, in the opinion of very many New Testament scholars, by the counter evidence.

[24] Ibid., 238.

The evidence that Jesus founded a Church is in the passage in Matthew's gospel in which he says to Peter, 'you are Peter (*petros*) and on this rock (*petra*) I will build my church' (Matt. 16: 16). But there are two reasons to doubt the authenticity of this verse. One is that Jesus expected the end of the present Age, when God would intervene to establish his kingdom on earth, to happen quite soon: 'there are some standing here who will not taste death before they see that the kingdom of God has come with power (Mark 9: 3); 'There are some standing here who will not taste death before they see the Son of man coming in his kingdom' (Matt. 16: 28); 'this generation will not pass away till all these things take place' (Matt. 24: 34); 'there are some standing here who will not taste death before they see the kingdom of God' (Matt. 16: 28); and we see from Paul's earlier to his later letters how central this expectation was for the early church, but gradually faded as time passed and the end failed to come. But if the End was to come soon there was no point in Jesus establishing a continuing organization, such as he seems to speak about in the first quote. The second reason is that this quote contains a pun in Greek – *petros* and *petra* – and Jesus did not speak Greek but Aramaic. And so it seems more likely that the part of the church headed by Peter created this saying to validate his leadership.

There is evidence in the saying Swinburne quotes that Jesus intended his death as a ransom (*lutron*) for many. Ransom was a poignant idea in the ancient world. Great numbers of people were slaves because their nation had been conquered, its inhabitants becoming slaves. And to be ransomed was a supreme good. So ransoming was a powerful metaphor for deliverance – in the case of Jesus' teaching, deliverance from the power of demons, of sin and of death. But taking the metaphor literally the church asked to whom was the ransom paid? Origen gave the accepted reply: it could not be to God, so it was to the devil; Jesus' death was part of a deal with the devil to free humanity. (We see this again in C.S. Lewis' Narnia story). So there is good reason to doubt whether this saying should be taken literally.

Jesus probably shared the widespread Jewish belief that the death of a martyr somehow benefited Israel and assumed that this would be true of his own death, which was brought about because of the Romans rulers' fear of a would-be messiah leading an uprising against them.

It should also be noted that the early church did not give the idea of atonement the place that it came to have much later. Thus the Apostles' creed (although the apostles had nothing to do with it, for it originated as the Old Roman Creed in the fourth century) affirms God almighty, and Christ Jesus, his only son, and the remission of sins – not specified further – and the Holy Ghost. The Nicene Creed, also fourth century, likewise affirms God, the Father all-sovereign, and Jesus Christ, Son of God, of one substance with the Father, and the remission of sins by baptism – but with no mention of Jesus' death as an atonement.

Did Jesus believe himself to be God? It is only in the very late gospel of John, written towards the end of the first century, that Jesus is depicted as consciously divine. In the earlier synoptic gospels (Mark, Matthew and Luke) he is a charismatic healer and preacher, a prophet in the Old Testament tradition. Further, the term 'son of God' did not have the meaning that it has come to have in Christian theology. It did not connote divinity. The ancient Hebrew kings were enthroned as 'son of God' – we have the enthronement formula in Psalm 2: 7: 'He said to me, "You are my son, today I have begotten you". Indeed any outstandingly pious Jew could be called a son of God. So Jesus may well have been called a son of God in this metaphorical sense. But as Christian doctrine developed, the metaphorical son of God was transformed into the metaphysical God the Son, second Person of a divine Trinity. But the historical Jesus is reported to have taught, 'love your enemies, and do good, and lend, expecting nothing in return, and your reward will be great, and you will be sons of the Most High' (Luke 6:36) – obviously in a metaphorical sense of 'sons'. Again, Jesus is reported to have said, 'Why do you call me good? No one is good but God alone' (Mark 10: 18).

The evidence that Jesus did not teach that he was God is also evidence against the idea that the later teaching of the church about Jesus' divinity was a continuation of his own teaching. On the contrary, his teaching about God's love for us, and our call to love one another, were not central in most of the developing doctrines of the church, which were in second-order philosophical and theological language rather than first-order religious language.

Finally, returning to what I see as Swinburne's excessively intellectual and propositional approach, it should be noted that creeds, in his sense of belief systems, play a much smaller part in the religious life

than he seems to assume. For most practicing Jews the rituals are all-important, and propositional beliefs very much in the background. For most Buddhists, whilst there is a core of beliefs, their religion is much more a practice and an experience than a set of beliefs. The same is true of the various streams of Indian religion collectively labelled Hinduism. And for very many practicing Christians it is the rituals that are important. Indeed for some, the Quakers, beliefs are of little importance compared with the response in life to God's love: as their Advices and Queries says, 'Remember that Christianity is not a notion but a way'.

Swinburne could of course reply to all this that he is doing philosophy of religion, which is necessarily a second order discipline. But in my opinion it should be a second-order discussion of religion itself in all its dimensions, not merely of belief systems, important though these also are.

EUTHYPHRO'S "DILEMMA", SOCRATES' *DAIMONION*, AND PLATO'S GOD

TIMOTHY CHAPPELL

The Open University

Abstract. In this paper I start with the familiar accusation that divine command ethics faces a «Euthyphro dilemma». By looking at what Plato's Euthyphro actually says, I argue that no such argument against divine-command ethics was Plato's intention, and that, in any case, no such argument is cogent. I then explore the place of divine commands and inspiration in Plato's thought more generally, arguing that Plato sees an important epistemic and practical role for both.[1]

I.

The commonest use that most philosophers today make of Plato's *Euthyphro* is as the citation for what they call "the Euthyphro dilemma", which is supposed to be "an intractable difficulty" (to quote one of many instantly-googlable sources) or "a fatal objection" (to quote another) to "divine-command morality", which is said to be the view that what is good or right is what God loves or wills or commands. The usual story is that the sceptical Socrates meets the credulous Euthyphro, a rather sanctimonious divine-command theorist, and sets him this question: "Is what is good, good because God wills it? Or does God will it because it is good?" But Euthyphro—so the usual story goes—cannot take the first alternative, that what is good is good because God wills it. For then the

[1] Thanks for comments to Robert Audi, Chris Belshaw, Sarah Broadie, Peter Cave, Nicholas Denyer, Chris Emlyn-Hughes, John Gingell, Jakub Jirsa, Derek Matravers, Michael Morris, Mark McPherran, Jon Pike, David Sedley, Malcolm Schofield, Nigel Warburton, Robert Wardy, James Warren, Naoko Yamagata, and other members of audiences at the B Club in Cambridge, March 2009; at a departmental research conference in St. Edmund Hall, Oxford, April 2009; at the Open University Summer School in Bath, August 2009; and at the British Society for Philosophy of Religion in Oxford, Sept. 2009.

content of the good would be set by God's willing it, which would make the nature of the good arbitrary. Nor can Euthyphro take the second alternative, that God wills what is good because it is good. For then the nature of goodness is already set before God's will comes into the picture, and so God is not sovereign but subordinate to morality, and we do not need God's commands to know what goodness is.

This interpretation, it seems to me, has just two flaws. First, "the Euthyphro dilemma" is not in the *Euthyphro*. Secondly, "the Euthyphro dilemma" is not a dilemma. Let me take these flaws in turn.

"The Euthyphro dilemma" is not in the *Euthyphro*, because Socrates and Euthyphro are not talking about "what God wills". They are talking about "what the gods love": two crucial differences, since willing is obviously different from loving, and since, as Socrates points out (7b), different gods might love different things. Also, they are not talking about "what is good", but about "what is holy (*to hosion*)"; a crucial difference because, as Socrates points out (7e), "the holy" is only one kind of goodness or virtue. There are other kinds, "the just" and "the honourable" for instance, and the relation of "the holy" to these and to the good or virtue itself is central to Socrates' inquiry (see, e.g., 12a). (Still less are Socrates and Euthyphro talking about "what is morally right". Their argument is not concerned with the *orthotês*, rightness, of Euthyphro's prosecuting his father, but with its *hosiotês* or *eusebeia*, holiness or religious propriety. As many commentators have pointed out, it is anachronistic to read our concept of "the *morally* right" back into Plato.) Finally, Socrates is not arguing that "what is holy" is *not* "what the gods love": at *Euthyphro* 10e-11a he explicitly allows that it can be, provided the gods agree, and at *Euthyphro* 11b1 he calls "being loved by the gods" a *pathos* (a true description) of "what is holy". The point of his argument is that the mere description "what the gods love" cannot serve as a *definition* of "what is holy". But then, Socratic definition is a notoriously demanding business, so this may not be a very great restriction. In any case, it is certainly not a restriction that prevents it from being true that what is holy is loved by the gods.

And "the Euthyphro dilemma" is not a dilemma, an insoluble problem, because theistic ethicists can and do take either horn, or else refuse both horns. They can affirm with some plausibility that God's will (or, better, love) determines the nature of the good, provided they do not also

say that God's love is what we would call arbitrary (which any intelligent and civilised theist is hardly likely to say anyway). On this view it is true that *if* God had willed that infanticide be a good thing, then infanticide *would* be a good thing. But it doesn't matter: though we are in a way very much at the mercy of his divine pleasure, God is good and trustworthy, so we may be sure that infanticide is *not* his will. Or theists can with equal plausibility affirm that the nature of the good determines what God loves, provided they do not also say that this makes God subordinate to morality in some sense that undermines his sovereignty. On this view it is true that God has no more freedom to make infanticide a good thing than he has to make 2 + 2 equal 5, or create a stone so heavy that he cannot lift it. But it doesn't matter: God is not in the least impaired by these formal restrictions from doing anything that he actually wants to do.

Better still, theistic ethicists can refuse both horns: they can reject the whole idea that they need to choose between the God-to-good and good-to-God orders of determination.[2] If someone asks "Does equilaterality in triangles determine equiangularity in triangles, or is it the other way around?", we do not expect geometricians to accept this as the first line of a proof by dilemma that equilaterality and equiangularity cannot be connected features of triangles. Rather, we expect them to reject the question. One shouldn't imagine triangles *first* being equilateral, and *then*, as a result of this, somehow becoming equiangular too. We could say something parallel here: one should not imagine God's will existing *first,* and *then* shaping or creating morality; or morality existing first, and then shaping God's will. All such conceptions are misconceptions. When they put them aside, theistic ethicists will probably come to agree with

[2] Another possibility: Robert Audi has suggested (Audi 2007) a distinction between the command*ed* and the command*able* right. All sorts of things might be possible commands of God because they are *prima facie* morally right, while only some things are actual commands of God and therefore morally right *sans phrase.* The nature of the good determines the commandable right; God's choice determines the commanded right. This is not my picture, since it does seem to involve us in accepting antecedent determinants of God's will; but it is another way out of the dilemma. Again, could theists resist the dilemma by reading "because" in different senses of "because", e.g. constitutive vs causal, in "It's right because God commands it" and "God commands it because it's right"? Perhaps, though I find this combination hard to make sense of.

the "dilemma's" proponent that, in truth, neither horn can be affirmed. But not for the reason he thinks.

II.

"But surely there is *some* dilemma about divine commands and ethics in the *Euthyphro*, even if it isn't the one that is usually supposed to be there. Doesn't *Euthyphro* 10a1-2 set a dilemma? And can't we call that the Euthyphro dilemma?"

We can *call* it what we like. But Socrates' famous question at 10a1-2 ("Is the holy loved by the gods because it is holy, or is it holy because it is loved?"), while it certainly offers Euthyphro two options to choose between, is not very well described by calling it a dilemma. A true dilemma is, so to speak, a *modus tollens* with a disjunction in it: if p, then either q or r; but not q; and not r; so not p. But there is no suggestion that both the alternatives in front of Euthyphro are impossible for him to accept. On the contrary, Socrates shows Euthyphro which horn of the alleged dilemma to grasp. What Euthyphro should say—and confused and disgruntled though he is throughout the discussion, and ultimately discomfited at its end[3], he never shows any clear sign of disagreeing with this—is that the holy is loved by the gods because it is holy, and not *vice versa*.

If we insist on speaking of a dilemma, we should say that Socrates here[4] grasps its second horn, "subordinating" (if that is the word) the gods to what is holy. So neither Euthyphro nor Socrates is at a loss as to how to choose between these two alternatives, as we would expect in

[3] Mark McPherran (2003: 32-5) interestingly raises the question where Euthyphro goes at the end of the *Euthyphro.* He departs in a hurry (15e4-5): to proceed with the prosecution of his father? He does not say so, as we might expect him to. Rather it seems that Euthyphro, so to speak, gives up his 'place in the queue' at the door of the court: unlike Socrates, he no longer *diatribei peri ten tou basileôs stoan* (1a2-3). Evidently he lacks the constancy of character to bring an unjust prosecution—while Socrates does not lack the constancy to face one. So part of the irony of Socrates' *Hoia poieis, ô hetaire?* (15e6) is that Euthyphro's flight may be bad news and a big disappointment for one old gentleman, but is good news, and an unexpected reprieve, for another.

[4] Contrast *Republic* 597b, where we are told that God (*theon*) makes the Forms: an apparent subordination of the Forms to God. Perhaps Plato solves the "dilemma" in different ways for the gods plural and for God singular.

a true dilemma. Nor does Socrates suggest that his argument has refuted the whole idea of a divine-command ethics. Certainly it has refuted Euthyphro's attempted definition of the holy. But as already observed, Socrates does not argue by dilemma: he does not force Euthyphro to choose between two equally unpalatable alternatives, where the impossibility of both disproves Euthyphro's initial claim. Rather he refutes Euthyphro's definition by insisting, not entirely explicitly, on two familiar Socratic doctrines, both closely connected with his own rather specialised notion of definition. One is what we might (with what I hope is a harmless anachronism) call the logical priority of essence over accident and of activity over passivity (10b-c). The other is the doctrine that definitions must refer to the essences of things, not merely to their accidents or affections:

> And it looks, Euthyphro, as if—when you were asked, about the holy, exactly what it is (*ho ti pot' estin*)—you were unwilling to make its essence (*ousia*) clear to me. Instead, you told me about some affection (*pathos ti*) pertaining to it—something that this "the holy" has undergone, namely to be loved by all the gods. But its *being* (*ho de ti on*) you have not yet spoken. (*Euthyphro* 11a8-b2; I use my own translations throughout)

The point is not the downfall of any possible system of divine-command ethics. The point is not even that what is holy might not be truly described as "what the gods love". The point is simply that what is holy cannot be *defined* as "what the gods love", because "what the gods love" is a term in the wrong logical category to do this defining work. If we know only that the gods love something, we still need to know *why* they love it. It is only as we begin to get answers to that sort of question—if we do—that we can begin to get any sense of the real nature of what is holy.

Naturally, if the gods disagree on what they love (8d-e), then they will be little better as authorities about what is holy than Euthyphro's own warring family (4d-e). Taking views about what is holy will then be no more than taking sides in a feud, and our interactions with the gods will remain at the lowly level where Euthyphro's clearly are—the level of an *emporikê technê* of bartering temple-sacrifice for protection, perhaps even protection from other gods. The implied critique of Euthyphro's own beliefs about what the gods command, and of the popular religious

ethics of Socrates' own day that they no doubt represent, is obvious.[5] It is much less obvious that we are being told by Plato that *any* ethical beliefs based on divine commands, or *any* religious ethics, should be rejected. On this evidence, the moral of the *Euthyphro* might equally be the very different point that, in order to formulate an adequate divine-command ethics, we need a more adequate theology than Euthyphro's.

III.

There is another obvious reason why it was paradoxical to follow the usual reading of the *Euthyphro* and see Socrates as an opponent of divine-command ethics. This reason is that, as Euthyphro himself reminds us right at the start of the dialogue, Socrates himself lives by divine commands (*Euthyphro* 3a10-b5):

> EU. What things does [Meletus] say you are doing that corrupt the young?
> SOC. My fine fellow, they are absurdities, or at least they sound absurd. For he says that I am a maker of gods; that I make new gods, and do not respect the old ones. He has indicted me on behalf of the old gods, or so he says.
> EU. I understand, Socrates; it is because of the daemonic sign (*to daimonion*) that you say comes to you every now and then.

What is this "daemonic sign"? As Socrates very famously puts it in the *Apology*:

> The explanation of this [Socrates' way of life] is what you have so often heard me tell you about in so many contexts: it is that something divine and daemonic (*theion ti kai daimonion*) comes to me, the very thing that Meletus makes fun of me for in his indictment. From my childhood on it has come to me as a sort of voice (*phônê tis*), and whenever it comes, it always holds me back from something that I am about to do—it never pushes me forward… (31c10-d6; cp. 40b, *Phaedrus* 242c1, *Euthydemus* 272e)

[5] Though Xenophon manages, with even more than his usual obtuseness, to present Socrates as a conventionally-minded *defender* of this *emporikê technê*: *khrê oun mêden elleiponta kata dunamin timan tous theous tharrein te kai elpizein ta megista agatha* (*Memorabilia* 4.3.17). It is quite an achievement to make Socrates' views sound so dull.

We may contrast the directness of Socrates' "voice" with the most direct way in which, so far as we know, Euthyphro hears from the divine (5e-6b):

> EU. See, Socrates, what a great proof (*mega tekmêrion*) I give you that the law really stands this way—one which I have given others already… The fact is that men themselves believe that Zeus is the best and most just of the gods, and they agree that he chained up his father because he [Cronos] devoured his children, contrary to justice, and Cronos had mutilated *his* father [Ouranos] for other crimes of this sort. Yet they find fault with me for prosecuting my father when he has acted unjustly! So they are contradicting themselves in what they say about the gods and about me.
>
> SOC. And isn't this, Euthyphro, the very reason why I am being prosecuted—that when someone tells me this kind of story about the gods, I find it rather hard to accept?[6]

Euthyphro's rather boastful claim (5a2) to a unique level of accurate (*akribôs*) knowledge about everything to do with holiness is quite shown up here. Despite his confession of ignorance, it is clear that Socrates knows more about holiness and the gods than Euthyphro does.[7] Euthyphro is a sort of fundamentalist of the Greek myths. The basis of his belief that the gods command him to prosecute his father is that the myths

[6] Compare *Phaedrus* 229e-230a, where Socrates rejects the (Anaxagorean?) project of seeking rationalising explanations for stories of the gods, and even says that he does accept traditional theology: "I cannot yet know myself, in compliance with the Delphic inscription, and it seems ridiculous to me to inquire into unrelated matters when I am still ignorant about that. And so I pass by these questions about the gods, *and accept the conventional view about them* (*peithomenos de tôi nomizomenôi peri autôn*); as I said just now, I don't investigate them myself". Similarly he says at *Apology* 35d7 "I believe in the gods as none of my accusers does": *nomizô… hôs oudeis tôn emôn katêgorôn*. There is an obvious *double entendre* in *hôs*: Socrates' theism differs from Meletus' not just, as it were, in quantity, but in quality too. (Thanks for discussion to Mark McPherran. For his views on this issue see further McPherran 1997a.)

[7] *Cratylus* 396d ff. has Socrates claiming to have been inspired by his recent conversation with Euthyphro. But this is obviously ironic. The *Euthyphro* says nothing about etymologies, which is what Socrates is here claiming to speak about with inspiration. Nor is there any sign in the *Euthyphro* that Socrates learns anything at all from his conversation with Euthyphro. The passage proves little except that the *Cratylus* was written after the *Euthyphro*, which presumably we expected anyway.

say that Zeus did something horrible to his father because his father had done something even worse. And Euthyphro just accepts these myths as providing him with examples to justify his own actions by. It is hard to imagine a more rickety and second-hand basis for ethics. The central problem with Euthyphro's way of thinking is not so much that it commits him to relying on divine commands, as that he cannot give us any reason to agree with his belief that the bizarre and savage myths that he mentions here should have anything like the status of commands or examples for anyone. Of course Socrates too is not above appealing—in a rather more controlled fashion—to the Greek myths while considering ethical questions (see e.g. *Apology* 28c-d, praising Achilles' fearlessness of death). However, Socrates in the *Republic* explicitly and directly bans his hearers from providing a pretext for their own misdeeds by appealing to the gods' misdeeds, in a passage where Plato quite clearly alludes to *Euthyphro* 4b: "When a young man is listening, it is not to be said that it is nothing special for someone to do the very worst kind of injustice, or again to punish his father's injustice in the extremest possible way, because it is no more than what the greatest and first of gods have done already" (*Republic* 378b1-5). Moreover, Socrates also knows something directly of the gods, by inner experience.

Does Euthyphro too hear directly from God in his inner experience, as well as being an expert in the myths? Apparently he would claim that; he certainly claims to foretell the future, at 3c, and it looks as if such claims were at least often based on claimed direct experience of a 'divine voice' or some such experience. However, notice this: at *Euthyphro* 3e Socrates invites a prophecy of how his and Euthyphro's cases will go, and Euthyphro (3e3) replies "Well no doubt there will be no big deal, Socrates; you will contest your case according to wisdom (*kata noun*), as I think I will mine". So the only time in the dialogue that Euthyphro actually gives us a prophecy, he apparently gets it spectacularly wrong—certainly about his own case, and probably about Socrates' too. Since (we might say) Socrates will indeed conduct his case *kata noun,* perhaps Euthyphro does have hold of something genuine. But that something genuine is also something Delphic, and Euthyphro misunderstands the ambiguous phrase *agônisthai kata noun*.

The evidence could hardly be clearer that Socrates' report of his *daimonion* is a report of a direct religious experience of being divinely

commanded. We should not follow some recent writers in attempting to naturalise the *daimonion* into no more than a moral or rational hunch (so Vlastos), or the voice of reason within him (Nussbaum 1985: 234), or into the voice of Socrates' own subconscious (Weiss 1998: 19)[8]. Of course, if you are a philosophical naturalist, then (from an external, *de re* standpoint) you will want to naturalise Socrates' *daimonion* somehow. Be that as it may, there are no grounds for the corresponding internal, *de dicto* claim that *Socrates himself* saw his own *daimonion* this way. As Vlastos (1999: 57) puts it, we have here to face

> a fact about Socrates which has been so embarrassing to modern readers that a long line of Platonic scholarship has sought. . . to explain it away: Socrates' acceptance of the supernatural. . . . If we are to use Plato's and Xenophon's testimony about Socrates at all we must take it as a brute fact—as a premiss fixed for us in history—that, far ahead of his time as Socrates is in so many ways, in this part of his thought he is a man of his time. He subscribes unquestioningly to the age-old view that side by side with the physical world accessible to our senses, there exists another, populated by mysterious beings, personal like ourselves, but, unlike ourselves, having the power to invade at will the causal order to which our own actions are confined. . .how they act upon us we cannot hope to understand. But the fact is that they do and their communications to us through dreams and oracles is [*sic*] one of the inscrutable ways in which they display their power over us.[9]

Vlastos speaks here of Socrates' supernaturalism as "unquestioning" and wholly traditional. But even the earliest evidence, as far back as Meletus' indictment as cited above, says clearly that Socrates' theism had become

[8] "The *daimonion* is not. . .a voice independent of Socrates' own thinking and intuition that instructs him to contravene their guidance but rather a voice inspired by Socrates' thinking and intuition, by beliefs that are for the moment "subconscious"—if the reader will forgive the anachronism—a voice that gives him the strength to implement these "subconscious" beliefs when he is tempted to do otherwise. Indeed, when there is no tension between Socrates' imminent act and his deeper sense of what is right, when Socrates has no reservations, no qualms, about the course he is about to pursue, his *daimonion* is silent."

[9] Cp. McPherran 1997: 6: "by [Socrates' religious language] he is [on one line of interpretation] simply referring surreptitiously in the language of 'the many' to the 'divine' inner promptings of his utterly secular, completely human powers of ratiocination… In my judgement, this portrait of Socrates is the result of slighting and misinterpreting the evidence of our texts."

in important ways very untraditional, precisely as a result of Socrates' own questioning. (Contrast *Phaedrus* 229e-230a, cited in Footnote 5, for Socrates' willingness to leave traditionalism alone on grounds of ignorance, with *Timaeus* 40d7 ff., cited below, for Plato's much less eirenic attitude to traditional beliefs.)

At the level of Socrates' own self-understanding it is perfectly clear, both from Plato's evidence and from Xenophon's (*Memorabilia* 1.2-5), that Socrates takes the voice of the *daimonion* to be a form of supernatural guidance quite separate from the guidance that any of us might get, and that Socrates himself obviously sometimes gets, from doing some careful reasoning, or from trying to plumb the depths of our own minds. For if the *daimonion* were not, on Socrates' own conception of it, explicitly supernatural, what could be the point of his argument at Apology 27c that, since he believes in *daimonia*, he must also believe in *daimones*, and so cannot be accused by Meletus of atheism?

This restraining voice is not "virtually worthless" from a rational point of view;[10] and in any case, the restraining voice is not the only supernatural voice that Socrates hears. If we take *au pied de la lettre* his last remark just quoted, that his *daimonion* always speaks negatively, to forbid, and never positively commands, then the voice that Socrates talks about a little later in the *Apology*, as positively commanding him to philosophise, must be a different voice, or voices:

> This task [of philosophising], as I say, has been commanded for me by (the) God (*tôi theôi*), both by oracles, and by dreams, and by every way there is of commanding a divine destiny (*theia moira*) for any man. (*Apology* 33c5-8; cp. 37e6)

[10] *Pace* Brickhouse and Smith 1989: 253-4: "once we notice how little information Socrates gets from a daimonic alarm, we can see why Socrates could never be made wise by his *daimonion's* alarms. After all, when the *daimonion* tells Socrates that he should desist from what he is about to do, he can be completely certain that he must not continue what he was about to do. But this information tells him nothing about what it is that is wrong, when it is wrong, why it is wrong, and what it is to be wrong. The god does not lie to Socrates, but does manage to tell him next to nothing through the *daimonion*. What Socrates gets from his sign, therefore, is virtually worthless for the pursuit of the sorts of truth Socrates seeks philosophically—truth that explains and defines, and which thus can be applied to judgments and deliberations required for the achievement of the truly good life for men." More about this implicit contrast between seeking truth *philosophically* and in other ways towards the end of this paper.

("For any man": notice the generality of this. Socrates tells us at *Republic* 496c2-4 that he doubts that many others have experienced *to daimonion sêmeion*—though perhaps a few have. Even if the *daimonion* is almost unique to Socrates, the other sorts of religious experience and divine guidance that he also claims to have had hardly are. So I doubt we should infer, with Pierre Destrée (Destrée and Smith 2005), that Plato depicts Socrates as uniquely divinely guided because he is the only true philosopher. For one thing, Plato does not think that Socrates *is* a true philosopher: what Socrates has is the right way in to the true philosophy, Platonism, but not Platonism itself.)

It is not obvious that this *theos* who commands Socrates is straightforwardly Apollo, whose ambiguous oracle Socrates interprets as telling him to begin his quest for a wiser man than himself (*Apology* 21a-23b). For one thing, texts like *Euthyphro* 6a6-9 show us good reason to doubt that Socrates believed in the traditional Apollo at all, hence would have seen oracular utterances from an "Apollonian" source as really messages not from that Apollo, but at least from Apollo non-traditionally conceived, and/or from the being whom Socrates calls *ho theos*, which we may translate either "God" or "the god". Thus at *Apology* 21b7 Socrates asks himself what "the god can mean" by the oracle, and takes it for granted that the god cannot be lying—"for that would be against his nature". But a host of texts and traditions which Socrates must have known prove that lying and deception, like the rest of the anthropomorphic characteristics condemned in the *Republic* (e.g. at 378d, 379e), are not against *Apollo*'s nature, not at least on the traditional conception of Apollo: Aeschylus, *Agamemnon* 1228 ff. for a start, where Cassandra recounts how she double-crossed Apollo, and he double-crossed her back. Again, it may have been Apollo's oracle that started Socrates' quest, but we have no decisive evidence that Socrates thought that the many other portents that he mentions here were from Apollo, not at least if that means an Apollo anything like Aeschylus' Apollo.[11]

What Socrates did think is that these dreams, visions and voices came (more or less directly: perhaps the *daimonion* is an intermediary) from God, and that they came to him as divine commands. That makes it as

[11] For a different view see McPherran 2005. As McPherran has reminded me, Socrates is at least orthodox enough to compose a hymn to Apollo in prison (*Phaedo* 60d ff.)—though both Cebes and Euenus seem rather astonished to hear that he is doing this.

certain as it could be that, whatever else the *Euthyphro* is meant to be about, it cannot be intended as an attack on divine-command ethics. It is better understood, I suggest, as part of Plato's campaign against contemporary Athens' dominant theology of (as we might call it) chaotic polytheism, and in favour of something more like the ethical monotheism that is familiar to those of us who live in cultures shaped by the Abrahamic religions. Thus even when Socrates speaks of gods plural in the *Euthyphro*, what he says of them is that *ouden gar hêmin estin agathon, ho ti an mê ekeinoi dôsin* (15a1; cp. *Stm* 273b8)—almost like the Anglican Eucharistic affirmation that "all that we have comes from you, and of your own do we give you". The dialogue's deepest moral is not that divine-command ethics should be rejected. It is that Plato's ethical theism is preferable to chaotic polytheism, because it gives us the only possible context in which divine-command ethics can be sustained. For only if God is good can it be reasonable to hope that what God commands will converge with what is morally right.

> "But surely Plato can't give any respectable place in his thought to anything like Socrates' daimonion! Whatever the historical Socrates (or any other Socrates distinguishable from Plato) may have thought, Plato's sarcastic dismissal of manteia, inspiration, prophecy, and the like is one of the most frequent themes in his canon. How could Plato of all people, austere rationalist as he is, possibly retain such a relic of primitive religious irrationalism in his philosophy? Mustn't there be too much irony in his reports of Socrates' supposed religious experiences for us to take them seriously?"[12]

Briefly, my answer is No. I explain that answer in the next section.

[12] Or in Vlastos' words (1999: 69): "For Socrates diviners, seers, oracle-givers, poets are all in the same boat. All of them in his view are know-nothings, or rather, worse: unaware of their sorry epistemic state [unaware that they don't have the requisite sort of understanding], they set themselves up as repositories of wisdom emanating from a divine, all-wise source. What they say may be true; but even when it is true, they are in no position to discern what there is in it that is true. If their hearer were in a position to discern this, then *he* would have the knowledge denied to them; the knowledge would come from the application of *his reason* to what these people say without reason." Note the conflict between this Socrates-as-rationalist and the Socrates-as-traditionalist of the Vlastos quotation on my p.6.

IV.

At *Apology* 22b2-c8 Socrates is explaining to the judges at his trial how he discovered, to his surprise, that the poets were not wiser than him:

> . . . pretty well anyone present could explain those poems better than the very men who wrote them. I soon recognised that the thing about the poets was that they too [like the orators] did not do what they did by wisdom (*sophiai*). Instead they did it by some sort of natural instinct and by divine indwelling (*physei tini kai enthousiazontes*), like prophets or soothsayers (*theomanteis kai khrêsmôidoi*): who also say many fine things, but do not know anything of what they are talking about (*isasin de ouden hôn legousin*). . . .Their poetic gifts made them think that they were the wisest of all men in other things besides poetry as well; but mistakenly.

This is just the sort of passage that is easy to read as Plato engaging in a "sarcastic dismissal", as I called it above, of the claims of any sort of inspiration or revelation. There are plenty of other passages like it: *Ion* 533c-534e, for instance. The most strikingly similar—at times it is almost a doublet of the *Apology* passage—is in the conclusion of the *Meno* (99b-e):

> . . . it was not by wisdom (*sophiai*), nor by being wise, that men like these led their cities—men such as Themistocles and his circle, and those whom Anytus here mentioned just now. . . So if it was not by knowledge (*epistêmêi*), it must have happened by the only alternative, well-founded opinion (*eudoxiai*). That is what these politicians used to direct their cities; in respect of rational understanding (*pros to phronein*), their condition was no different from that of soothsayers and prophets (*khrêsmôidoi kai theomanteis*). For these too say many true things, but do not know anything of what they are talking about (*isasin de ouden hôn legousin*). . . those soothsayers and prophets whom we have just mentioned—and all who are artistically inclined (*poiêtikous*) as well—succeed in many great things in what they do and say, even though they lack mind (*noun mê ekhontes*). For this reason, we could rightly call them divine (*theious*). And we should say that the politicians are no less divine and divinely-indwelt, being inspired and possessed by god (*theious, enthousiazein, epipnous, katekhomenous ek tou theou*)… So [virtue seems to be] something that comes upon us by a divine dispensation without mind (*theiai moirai paragignomenê aneu nou*).

It is easy to read passages like these two as constituting clear and decisive evidence that Plato (or Socrates?) has no sympathy at all for the idea that anything might be learned from special experiences of apparently direct inspiration or revelation, whether those supposed revelations are religious, or ethical, or aesthetic. Since inspiration is just a kind of stumbling around in the dark, the "praise" of Themistocles and his circle that Socrates offers here is obviously (on this reading) no more than a veiled condemnation of them and their methods. They do not work by knowledge but at best by true opinion—and as the *Republic* shows at length (the reading continues), there cannot be a clearer condemnation of their ways than that. One deservedly influential recent interpretation of Plato, Nussbaum's in her celebrated 1986 essay "Madness, reason, and recantation in the *Phaedrus*", takes as read this early hostility to claims of inspiration, and assumes that it is Plato's uniform and consistent attitude until the time of the dialogue she focuses on, the *Phaedrus*, where Socrates' famous "recantation" (242d ff.) introduces a rather less austere and rationalistic approach to such claims.

But in fact even the earliest evidence is much more equivocal than such a reading suggests. For instance, the two passages I have cited above, one from a central early-period dialogue, the *Meno*, and the other from what may be Plato's very first philosophical publication, the *Apology*, are not, on a closer reading, merely "sarcastic dismissals" of the claims of inspiration. There are passages in Plato where it is right to hear sarcastic dismissiveness as the tone of the text (for instance Socrates' exchanges with Thrasymachus, Callicles, and Polus). But the sense of these passages from the *Meno* and the *Apology* is more nuanced.

Consider first these passages' contexts. Socrates at *Apology* 22b-c cannot be concerned only to dismiss with sarcasm the very idea that divine inspiration or revelation could have any authority. For Socrates himself appeals to inspiration throughout the *Apology*. Note his own prophecy of Athens' future at 38c. Of course this could be Plato retrojecting words into Socrates' mouth in the light of hindsight. But even then Plato is still committed to the claim that this was *the sort of thing* that Socrates said. Which we see anyway from the other most notable appeal to inspiration in the *Apology*, which introduces this very passage: at 22a4, he tells us that his examination of the poets was prompted by a divine command.

Similarly, Socrates' aim in *Meno* 99b-e cannot just be to mock politicians, poets, and priests for their dependence on the illusions and fatuities of supposed divine inspiration. For the deepest and most important teaching of the *Meno* is introduced by a passage in which it is possible to see Socrates himself as being overcome by a divine inspiration (81a ff.). Even on the most modest interpretation of this passage, Socrates here appeals to the authority of certain priests and priestesses "about divine matters" (81a4-5). It is quite impossible to square this way of proceeding with Plato's alleged "austere rationalism". (These wise ones have, intriguingly enough, succeeded in their studies of their own enterprise to a point where they are able to "give an account" of it. Tantalisingly, however, this is an account which Socrates does not pass on to Meno and us.)

Similar structural and contextual remarks can be made about other early-period texts which ought, on a reading like Vlastos' or Nussbaum's, to be straightforward exemplars of a high-minded rejection of the claims of inspiration. Euthyphro's initial attitude to Socrates (3c) is to see him as a kindred spirit; if Euthyphro is laughed at for offering prophecies in the *ekklêsia*, the kind of things that bring Socrates too the Athenians' derision and slander do not seem to him very different. Even the *Republic* has at least one myth in it (614b ff.), and even the *Crito* (44b) reports a prophetic dream of Socrates', while the *Phaedo* (60e-61b) reports multiple dreams and a divine command to him—a command, what's more, to practise poetry.

Again, in the *Apology* passage, notice Socrates' embarrassment (*aiskhunomai*, 22b6) at relating his discovery about the poets. Is his talk of embarrassment here mere affectation of what he does not actually feel, a clumsy attempt to make what he is saying more acceptable to his listeners? That is how we must read it if we think (like Vlastos) that Socrates' real point here, ironically concealed, is merely that poets talk a lot of irrational rubbish, which true philosophers in their superiority will shun. On that reading, similarly, when Socrates says that the poets *legousi polla kai kala* (22c4; cp. *Euthyphro* 13e12), we will have to take these words as nearly the opposite of his real view. It is more natural to take both remarks at face value. Socrates really is embarrassed, and the source of his embarrassment is that he really does think the poets' works fine productions—yet cannot square their admirable qualities with the chaotic and irrational way in which, it seems to him, all poetry comes to be.

Similarly with *Meno* 99b-e, the point of equating poetry (and the political and priestly arts) with *eudoxia*, well-founded belief, as opposed to equating them with *epistêmê*, knowledge, is not to dismiss poetry, politics, and priestcraft as completely worthless. Rather, the point is the same as the point of the famous comparison between knowledge and true belief—the image of the road to Larisa—that Plato's Socrates has just offered us at *Meno* 96e-98b. Knowledge and true belief can produce the same good results; the difference between them lies not in their results, but in the *unaccountability* of the good results of true belief. With mere true belief, there is by definition no explaining how we get the good results we do; we just *do* get those results, and that's all we can say (*Symposium* 202a4-9). This difference between knowledge and true belief may have grave consequences in some areas—for example, as the *Meno* and *Protagoras* and *Laches* all stress, it is a difference that makes true belief impossible to convey to another by teaching, or at any rate by rational teaching. But it does not abrogate from the genuine admirability, indeed the divine quality, of the good results that true belief can have. Both Socrates' mission in the *Apology*, and Themistocles' political skill in the *Meno*, come *theiai moirai*, by divine allotment. So when Plato's Socrates describes politicians such as Themistocles as *theious* he is not merely being snide. His admiration for their achievements is perfectly sincere. What complicates his attitude to them is not an undertone of sarcasm, but of puzzlement at how they can be so divine in their achievements, and yet so innocent of any touch of real understanding of what they are about.

Compare Socrates' puzzlement, in the *Ion*, about how Ion can be (532c) so expert in understanding and expounding *Homer*, yet so dozy and incompetent when it comes to *Hesiod*. The answer is that his understanding of Homer is a *theia dunamis*, not a *tekhnê* or *epistêmê*, *Ion* 533d2 (it too comes *theiai moirai*: *Ion* 534c1). And the present point generalises in a way that shows up something wrong in W.R.M. Lamb's comment, in his introduction to the *Ion* in the Loeb edition (p.403), that in that dialogue Plato insists "that no art… can be of real worth unless it is based on some systematic knowledge." The point is rather that art like Homer's clearly *is* of real worth—there is no irony in Socrates' description of Homer as *aristos kai theiotatos tôn poiêtôn*, *Ion* 530b10—*even though it isn't* based on systematic knowledge. The puzzle, and it is a deep one, is how this can be.

We might feel the same puzzlement about Socrates himself. How can Socrates' philosophical inquiries have such remarkable results, when he himself knows nothing? Famously, Socrates' own explanation is that he is a sort of midwife of ideas: he produces no children of his own, but helps bring others' conceptions to the light (*Theaetetus* 148c-151d). Mightn't something similar be true of Themistocles' sort? Mightn't they too be able to bring about in others a knowledge or expertise that they cannot attain to themselves? The trouble with this suggestion is that Socrates, for all his looking, never finds anyone who actually possesses any such *politikê tekhnê.* Come to that, we never in the early dialogues see Socrates' midwifery bring about any successful labours. Given that so many of those dialogues end in *aporia*, it is hard to see who Plato thinks are Socrates' successful patients. (Perhaps he has his own philosophical school in mind: see Sedley 2005. More about that suggestion later.)

Given the scant results of Socrates' own rationalistic endeavours, the dialogues' attitude to politicians like Themistocles is generally not one of simple denunciation. The attitude is summed up by the question posed by the *Meno* quotation above: how can they be so "divine" (*theious*) when they are also "lacking in mind" (*noun mê ekhontes*)?[13] And this use of the word *nous* seems not to be an accident, since both for Socrates in the *Phaedo* (97c1), and for Plato in the *Timaeus* (39e8) and *Laws* (897c6, d9), Mind, *Nous*, is apparently a name of God. Its use here presses what struck Plato as a forceful paradox.

V.

What is the paradox? The parallel may seem outlandish, but it is almost Pascal's: *le coeur a ses raisons que la raison ne connaît point*.[14] It is that inspiration or revelation or intuition seems to be a possible route to the truth—as it clearly is for Socrates, for instance. Yet we cannot rationally explain *how* it gets us to the truth. By all the stern rules of philosophical

[13] In the *Ion* God not only bypasses but actually *takes away* the *nous* of rhapsodes*: dia tauta de ho theos exairoumenos toutôn ton noun*, 534c9-10.

[14] *Le coeur a ses raisons, que la raison ne connaît point. On le sent en mille choses. C'est le cśur qui sent Dieu, et non la raison. Voilà ce que c'est que la foi parfaite, Dieu sensible au cśur*. – Blaise Pascal, *Pensées*, IV.277 (p. 458 in Brunschvicg).

rationality, that can only make inspiration suspect. Plato (and Socrates) can be read as regarding its directness and certainty as a kind of *cheat*: when we should be earning our certainties by the long and arduous road of the dialectical education that the *Republic* lays out for us, simply to claim to *see* the truth has all the advantages of theft over honest toil. We might say, with a second and even more outrageous parallel, that it is as bad as getting a computer to find the meaning of life for you:

> "You just let the machines get on with the adding up," warned Majikthise, "and we'll take care of the eternal verities, thank you very much. You want to check your legal position, you do mate. Under law the Quest for Ultimate Truth is quite clearly the inalienable prerogative of your working thinkers. Any bloody machine goes and actually finds it and we're straight out of a job aren't we? I mean what's the use of our sitting up half the night arguing that there may or may not be a God if this machine only goes and gives us his bleeding phone number the next morning?" (Adams 1979: 129-30)

Unlike Majikthise, Plato takes no pleasure in prolonging intellectual struggle and bewilderment for its own sake (still less for cash). But he is convinced that we cannot claim real knowledge unless we can not only see the truth, but also explain what makes it the truth, and how each truth connects to the others. If we are so lucky that intuition or revelation or inspiration *gives* us the truth, we should not be ungrateful for that; but we should not be satisfied either. Beyond true belief, there is still the long road of justification to travel. And what we are likely to have by the end of this journey is so different from what we started with, that we should not speak of knowledge as simply an *upgraded* version of true belief at all. Most obviously in the *Republic* and *Theaetetus*, knowledge is apparently so different from true belief or perception that true belief and perception are not even ingredients of it.

However, there is a further twist of paradox in the tale. The ultimate aim of the *Republic*'s dialectical education is itself a sort of direct, revelatory perception, acquaintance, or intuition: "True Being… is visible (*theatê*) only to *Nous*" (*Phaedrus* 247c8; cp. *theôrôn* at *Symposium* 210d4). Hence we find Plato apparently denouncing this-worldly perception in almost the same breath as he exalts the perception of the Forms (*Rep* 517b1-9). Knowledge, as the *Theaetetus* insists, is not perception in any ordinary or mundane sense. And yet at the end of the philosopher's laborious ascent,

by way of reasoning and hard thought and logical work, what we find is that the ultimate knowledge is itself something so like perception in its directness, immediacy, and non-discursive simplicity that Plato never finds a better or more illuminating image to describe it by.

The worry underlying Socrates' Dream (*Theaetetus* 201e-202d) is about this similarity. The worry is that, for all the disanalogies between Plato's own view and the empiricist theories that (on my reading, which is controversial, but not to be defended here: see Chappell 2005) the *Theaetetus* is devoted to attacking, still there is a deep structural parallel between empiricism and Platonism. Both rest all discursive, propositional, rational knowledge on a foundation of non-discursive, non-propositional, surd acquaintance. Hence both raise the paradox that the ultimate objects of what we would like to call knowledge are themselves not *reasoned about* or *known*, but *perceived* (202b6-7).

Of course, given the differences between physical and intellectual perception that Plato also stresses, these parallels between the two kinds of perception would not amount to a revalidation of physical perception and other less-than-ideally-rational forms of intuition, such as inspiration and religious experience, were it not for this small point about the vision of the Form of the Good: it's impossible. Or at least, that vision is an ideal limit of the understanding. It is what all our mortal, body-imprisoned attempts at knowledge and wisdom strive towards. But we have little reason to think that any of us can actually *reach* the godlike state of contemplation of the Form of the Good that Plato urges us to imitate (*Theaetetus* 176b2, *Timaeus* 90b-c) so long as we remain in this life; or that we can know beyond all possibility of error that we have reached it, even when we have.

Certainly *Socrates*—whether that means the historical one, or Plato's character—makes no claim to have attained this state. On the contrary, he explicitly disavows any such knowledge (*Rep* 506c1):

> "What then?" I said, "do you think it justice for someone to speak about what he doesn't know, as if he did know? …Have you not noticed that all opinions without knowledge are things of shame? The best of them are blind. Or can you see any difference between blind men who take the right road [surely an allusion to *Meno* 96e ff.], and those who have a true belief without *nous*?"

The highest and most metaphysically ambitious doctrine—that of the *Republic*—that the wisest of all men—Socrates—can offer us: even this is not knowledge, but only true belief, an image (*eikona*, 509b1), a comparison or analogy (*apeikasia,* 514a1). It is not the good itself that he offers us, but only *an interest-payment* on the good. (The Greek word is *tokos*, literally 'child' (507a2); Plato means something derived from the good, by way of the good's own generative powers, which shares its properties and reveals them in small compass.) Even at his highest pitch Socrates is not a knower but a true-believer, with all that that implies. So in the *Euthyphro* he possesses, unwillingly, an even scarier skill than Daedalus', destabilising not only his own products but others' as well (11e2). Very commonly in Plato anything like positive doctrine on Socrates' lips is presented as a dream, or a vision, or a mere image, or the report of some inspired prophet or *sophos*. No doubt too it is Socrates'—or indeed Plato's—lack of knowledge that answers the old question why Plato writes dialogues rather than treatises like, say, Plotinus'.

The moral is clear.[15] Since creatures of imperfect knowledge like us cannot hope to attain perfect truth by the rational route of knowledge, we should not neglect the possibility of attaining truth by routes of less rational purity. In fact, if *Phaedrus* 244a8-10 is to be believed, we should actively *court* madness: for "the greatest of goods come to us through madness, provided it is given by divine gift" (*theiai dosei*; cp. *theia moira* above).

In that famous speech Socrates goes on to distinguish three kinds of *mania*: *mantikê*, foretelling of the future, a name which he etymologically connects with *mania* (244b); *prophêteia*, divining of *miasma* (244e);[16] and poetry (245a). He then argues (245b8 ff.) that *erôs* should

[15] Vlastos (1999: 66) explicitly opposes my reading: "however plausible it may seem", the view that Socrates "would look to the intimations of his *daimonion* as a source of moral knowledge apart from reason and superior to it" "is unsupportable by textual evidence and is in fact inconsistent with it". However, Vlastos' chief argument against the reading is that what Socrates gets from his *daimonion* (and other supernatural sources) is not *knowledge.* But that is not in dispute: my claim is only that Socrates gets true beliefs and reliable guidance from such sources. Anyway, Socrates does not get knowledge from his reasoning either.

[16] Is the diagnosis of *miasma* what is meant? 265b suggests rather a sort of religious transport that lifts its subject out of a sense of *miasma* and other kinds of trouble in the soul—perhaps a little like the sense of relief or forgiveness of sin sometimes reported by

be distinguished as a fourth kind. This catalogue of kinds of madness is open-ended: there is nothing to stop Socrates from adding the voice of his *daimonion* as a fifth kind, or the positive commands that he hears through dreams as portents as a sixth kind, or indeed more than one kind, of *mania*. Perhaps Socrates also thinks that he already experiences the first three kinds of madness anyway. We have already noted his claim to tell the future at *Apology* 38c (cp. *Phaedrus* 242c), and his practice of poetry in *Phaedo* 603-61b; and maybe finding the hole in a bad argument (or character) is as much like divining a *miasma* as it is like the activity of a midwife (notice too the reference to the *daimonion* at *Theaetetus* 151a3, helping Socrates to decide whom to associate with). In all of these ways and others we can hope, by the gift of the gods, to attain truth. How churlish to refuse that gift, where we cannot hope to attain perfect knowledge anyway.

VI.

Something like this, we might conclude, must be the "more adequate theology" on the basis of which Euthyphro—and Socrates—might have constructed their divine-command ethics. On the Platonic conception (we might now suggest), Euthyphro's crude exchange-and-mart cultus is replaced by *homoiôsis theôi*, the project of becoming God-like by getting as close as we can to God's or *Nous*'s own activity, the philosophical contemplation of the Forms, perhaps even of the Form of the Good itself. We might suggest that one way in which this getting-close can happen—and always the best—is by way of philosophical understanding and reasoning. Where that gives out, however—and it gives out frequently—we are not entirely without resource. Inspiration and revelation is possible too. And this can be a way to the truth, even when it cannot, on its own, be a way to the understanding that goes with knowledge. Such forms of insight may not come rationally, *by way of* reasoning, but that does not make them irrational, contrary to reasoning. *Pace* Vlastos (1999: 56), there is simply no clash between the Socrates who describes himself as

---

Christian revivalists, e.g. in Wesley's journals. The lack of match between the two lists may just be an anomaly.

a follower of divine commands and the Socrates who describes himself as "the sort of man who is persuaded by nothing in me but the proposition which appears to me to be the best when I reason about it" (*Crito* 46b). After all, the proposition which "appears best" to me can easily be: "Obey the god".

More generally, there can be perception of truths that we cannot arrive at by reasoning.[17] We can have reason, as Socrates has reason, to trust and obey what we take to be a revelation of God's will for us—even where we do not fully understand that revelation. *Timaeus* 70d-72b tells us that alongside the part of the soul that deliberates, and the part that only cares for food and drink and other bodily desires—indeed, physically in between them—there lies the liver, which receives confusedly, a little like a mirror, the clarities of the rational mind, and conveys them to the appetites in a vivid form that they will understand. That, Plato tells us, is why the liver is the bodily seat of divination, and a rational man should take its promptings seriously, as the indirect evidence that they are of the dictates of reason (*Timaeus* 71e8-72a2): "no one who is in his right mind (*ennous*) attains true and divinely inspired divination, but only when the power of his understanding is fettered by sleep, or when he is disinhibited through illness or some divine visitation (*dia tina enthousiasmon*). But it is the part of a sane man (*emphronos*) to remember what has been spoken and what has been recollected by divination".

The conception still needs a bit more refining; for even though (as this *Timaeus* passage shows) it has some support in Plato's own writings, still, in Plato's own terms, this conclusion has something importantly unsatisfactory about it. At least we have—I hope—shaken off the prejudice that anyone who claims to hear God speak is "simply insane or seriously deluded" (Long 2006: 65). Still, on this conception, divine commands, inspirations, and revelations turn out to be a kind of *pis aller*, a quick and dirty route to truth that is permissible for us only because we cannot manage the longer and purer route of dialectic. Perhaps we should hold such revelations at arm's length if we can; perhaps we should "investigate the concept of God" *"no further* than is needed to bring it in line with" our ethical views (Vlastos 1999: 60)? But recall the final message of Soc-

[17] See Chappell 2008 for an argument that the reasoning and perceiving alternatives may both be available, in principle at least, in ethics.

rates' palinode in the *Phaedrus*, which is not merely that inspiration gets you where reasoning would also have got you—less logically and rationally, but faster—but rather that there are many places that *only* inspiration or ecstatic vision can get you to, places that reason alone will be no more able to arrive at than a frigid speech like Lysias' or Socrates' first will win a lover worth winning.

For complex-psyche-d creatures like us—it might be different for the Demiurge—a purely calculative or ratiocinative grasp of the Form of the Good, with no wonder or joy or love or exaltation involved in it, would not be a *grasp* of the Form of the Good at all; no more than Mary the Colour Scientist can know everything that there is to be known about redness, just by fully understanding the physics of redness. The fourth kind of madness is the madness of the lover *and* the philosopher (*Phdr.* 249a1, d5), the man whose vision of beauty "down here" (*têide*) causes him to recollect the *true* beauty: "he sprouts wings and longs to take the upward flight, but when he cannot, he gazes upwards like a bird, neglecting things below" (*Phdr.* 249e7-9).

Having once achieved (some measure of) an experiential grasp of the good, it is not just natural, but irresistible, to try and rationalise the experience, to try and make sense of it or spin a theory out of it. Of course this effort can bear fruit, by helping us to understand what has happened to us; and of course experience needs to be subject to the jurisdiction of reason, because experience can, and notoriously often does, lead us astray when we misinterpret its frequently ambiguous oracles.

When experience leads us astray, a dilemma can certainly emerge. This dilemma is a genuine one, but the problem it poses is not (just) theoretical but practical—about whether to follow the experience, or the reasoning that suggests that the experience is misleading—and it would be better called Abraham's dilemma than Euthyphro's. And anyway, in our enthusiasm to register the truth that experience can lead us astray, we should not miss the equal truth that reasoning can lead us astray too—as William James points out, in a different context:

> Personal religion will prove itself more fundamental than theology or ecclesiasticism. Churches, when once established, live second-hand upon tradition; but the *founders* of every church owed their power originally to the fact of their direct personal communion with the divine. Not only the superhuman

> founders, the Christ, the Buddha, Mahomet, but all the originators of Christian sects have been in this case [as too, we can now add, was Socrates]; —so personal religion should still seem to be the primordial thing, even to those who esteem it incomplete. (William James, *Varieties of Religious Experience,* p.30; quoted from Hedley 2008: 103)

No doubt experience without rationality is (that word again) blind. But it is equally true that rationality without experience—rationality bereft of the kinds of roots in a foundation of experience that Socrates' Dream describes—is empty. What Plato at his best (e.g. in the *Phaedrus*) points us towards is not merely the slightly condescending moral that revelation or inspiration or divine command has its part to play, given our unavoidable cognitive deficiencies, in getting us towards truths which are, however, *best* grasped by systematic reasoning. Rather the moral is *experience first*; for it is only once you *have* experienced that you can have anything worth systematising.

How far such systematising should go, and how much it can in fact add to the cogency of the original vision (whether religious, or artistic, or intellectual, or ethical), is a question for another occasion. My own growing conviction is that most contemporary philosophers, in their understanding of religion, of ethics, and indeed of Plato, tend to try and persuade each other and us of the cogency of a theory or a generalisation, when what is really cogent—what really persuaded them—is not the theory or the generalisation at all, but the particular experience from which the theory is an extrapolation, the generalisation an *over*-generalisation. They tend, in short, to overrate the value of system, and to underrate or even ignore the value of epiphany.

> To seek our Divinity merely in Books and Writings, is to *seek the living among the dead*; we do but in vain seek God many times in these, where his Truth too is not so much enshrined, as entomb'd: no; *intra te quaere Deum*, seek for God within thine own soul; he is best discerned… as *Plotinus* phraseth it [*Enneads* 5.3.17], by an *Intellectual touch of him*… the soul itself hath its sense as well as the Body. (John Smith, *Discourses,* Cambridge 1660, quoted from Hedley 2008: 93; cp. Taliaferro and Teply 2004: 158)

BIBLIOGRAPHY

Adams 1979: Douglas Adams, *The Hitchhiker's Guide to the Galaxy.* London: Pan.

Audi 2007: Brickhouse and Smith: Thomas Brickhouse and Nicholas Smith, *Socrates on Trial*. Princeton: Princeton UP.

Chappell 2005: Timothy Chappell, *Reading Plato's Theaetetus*. Indianapolis: Hackett.

Chappell 2008: Timothy Chappell, "Moral perception", *Philosophy* **83**, 421-438.

Destrée and Smith 2005: Pierre Destrée and Nicholas D. Smith, edd. *Socrates' Divine Sign: Religion, Practice, and Value in Socratic Philosophy.* Special issue of APEIRON: A Journal for Ancient Philosophy and Science, Vol. 38, no. 2 (June 2005). Kelowna, BC: Academic Printing and Publishing, 2005.

Dodds 1951: E.R.Dodds, *The Greeks and the Irrational.* London: University of California Press.

Hedley 2008: Douglas Hedley, *Living Forms of the Imagination.* London: T. & T. Clark.

James 1899: William James, *The Varieties of Religious Experience* [1899]. Available online at http://www.psychwww.com/psyrelig/james/james4.htm#65 .

King 1986: Martin Luther King, "Pilgrimage to Nonviolence", in J.M.Washington (ed.), *A Testament of Hope.* San Francisco: Harper Collins, 1986.

Krailsheimer 1966: A.Krailsheimer, *Pascal: Selected Works.* London: Penguin.

W.R.M. Lamb and H.N. Fowler, *Plato: Statesman, Philebus, Ion.* Loeb Classical Library. Cambridge, MA: Harvard UP, 1925.

McPherran 1997: Mark McPherran, *The Religion of Socrates.* University Park, PA: Pennsylvania State UP.

McPherran 1997a: Mark McPherran, "Recognizing the Gods of Socrates," in M. McPherran (ed.), *Wisdom, Ignorance and Virtue: New Essays on Socrates, Apeiron*, Supplemental Volume 28 (1997): 125-139.

McPherran 2003: Mark McPherran, "The Aporetic Interlude and Fifth Elenchos of Plato's *Euthyphro*," *Oxford Studies in Ancient Philosophy* 25 (2003), 1-38.

McPherran 2005: Mark McPherran, "Introducing a New God: Socrates and His *Daimonion*," in Destrée and Smith 2005: 13-30.

Nussbaum 1985: Martha Nussbaum, "Commentary on Edmunds," in Cleary, ed., *Proceedings of the Boston Area Colloquium in Ancient Philosophy*, Vol. 1 (University Press of America, 1985), 231-240.

Nussbaum 1986: Martha Nussbaum, "Madness, reason, and recantation in the *Phaedrus*". In her *The Fragility of Goodness* (Cambridge: Cambridge UP, 1986), 200-233.

Blaise Pascal, *Pensées et opuscules,* ed. L.Brunschvicg. Paris: Classiques Hachette, 1961.

Sedley 2005: David Sedley, *The Midwife of Platonism*. Oxford: Oxford UP.
Taliaferro and Teply 2004: Charles Taliaferro and Alison Teply, edd., *Cambridge Platonist Spirituality*. New York: Paulist Press.
Vlastos 1999: Gregory Vlastos, "Socratic piety", pp.56-77 in Gail Fine, ed., *Plato 2: Ethics, Politics, Religion, and the Soul*. Oxford: OUP, 1999. (Reprinted from Vlastos' *Socrates: Ironist and Moral Philosopher*, Cornell UP 1991.)
Weiss 1998: Roslyn Weiss, *Socrates Dissatisfied*. New York: OUP 1997.
Xenophon, *Memorabilia*. Edd. E.C. Marchant and O.J. Todd, in the Loeb Classical Library, no.168. London: Harvard UP.

THE RULE OF ST. BENEDICT AND MODERN LIBERAL AUTHORITY

LINDA ZAGZEBSKI

University of Oklahoma

Abstract. In this paper I examine the sixth century *Rule of St. Benedict*, and argue that the authority structure of Benedictine communities as described in that document satisfies well-known principles of authority defended by Joseph Raz. This should lead us to doubt the common assumption that pre-modern models of authority violate the modern ideal of the autonomy of the self. I suggest that what distinguishes modern liberal authority from Benedictine authority is not the principles that justify it, but rather the first order beliefs for the sake of which authority is sought by the individual, and the degree of trust between the authority and the subject.

I. INTRODUCTION

One of the ideas embraced by the West during the Enlightenment was the idea that the only authority that needs no justification is the authority of the self over the self. Any other kind of authority is derivative. Philosophers recognized that we cannot live without political authority, so an important project of the early Enlightenment was to show that the authority of the state can ultimately be justified by the authority of the self. Other kinds of authority were permitted to disappear, particularly moral and religious authority, both of which are still commonly dismissed as incompatible with personal autonomy. My purpose in this paper is not to critique modern ideas of autonomy or the self. Instead, I want to re-examine the common assumption that pre-modern models of authority violate the modern ideal of the ultimate authority of the self.

My focus will be a document written by Benedict of Nursia almost 1500 years ago describing an authority structure that has been followed in all subsequent centuries and on every continent. I have several reasons

for undertaking this unusual study. First, I want to show that authority as described in the *Rule of St. Benedict* satisfies Joseph Raz's theses of authority, one of the most influential contemporary accounts of modern liberal authority. This has some interesting implications. It shows that Benedictine authority need not refer to anything more than well-known post-Enlightenment principles for its justification. It also shows that what distinguishes the modern liberal state from Benedict's pre-modern authority structure is not the principles by which the authority is justified, but something else. The difference, I will suggest, is in the content of the first order beliefs for the sake of which authority is accepted, and the degree of trust binding together authority and subject.

My second reason for looking at the Rule is that I believe a good way to understand authority is to examine structures of authority that have proven successful. This has some advantages over the alternative of making up principles of authority *a priori*. I am not denying the usefulness of the latter, but I think there has been insufficient attention to the former.

Third, the political state is not a good model for authority in communities in which there are shared ends and a high degree of trust within the community. If we examine these communities closely, we might see that they have features that apply to communities that are not monastic, so the Benedictine model may have wider implications for the nature of authority in general.

II. JOSEPH RAZ'S THESES OF AUTHORITY

In an influential account of political authority, Joseph Raz proposes several theses that he believes apply to authority *simpliciter*, and he then uses them to justify the authority of the modern state.[1] In this section I will briefly summarize Raz's theses. I will not defend them since for the purposes of this paper I am relying upon their influence, not their truth. But I think they are all reasonable principles, given a modern sensibility that detaches authority from dependence upon God, and which attempts to derive authority from the more basic authority of the self.

[1] Joseph Raz, *The Morality of Freedom* (Oxford: Clarendon Press, 1986).

We begin with the assumption that authority is a relation between the person or institution in authority and the subject, in which the former has a special kind of normative power over the latter. The authority's command or directive gives the subject a reason to act in the way directed by the authority. Raz proposes that an authoritative directive has two constitutive features, both of which come from H.L.A. Hart.

The first is **content-independence:**

An authoritative utterance gives the subject a reason to follow the directive which is such that there is no direct connection between the reason and the action for which it is a reason.

That is, within certain limits, the authority's authority[2] does not depend upon *what* the authority directs. The subject has a reason to obey regardless of the content. For instance, the authority can command me to drive on the left side of the road or on the right. The authority can command military conscription or it can repeal it. Content-independence is necessary for a directive to be authoritative, says Raz, but it is not sufficient since threats and advice are also content-independent utterances, but they are not instances of authority (pp. 35-37).

The second thesis is the **pre-emption thesis:**

The fact that an authority requires performance of an action is a reason for its performance that replaces other relevant reasons and is not simply added to them.

This thesis is the distinguishing feature of authoritative directives. The subject's reason for following the directive is that it is a directive of the authority, not that the authority's directive gives the subject additional reasons for doing the act directed. The latter would not be a case of obeying authority, but a case of taking advice (pp. 42, 57-59).

Raz argues that there are cases in which a person ought to obey a directive pre-emptively by offering two other theses on authority which assume the primacy of the subject's own reasons for action: the Dependency thesis and the Normal Justification thesis. The first places limits on the kinds of reasons that an authority may legitimately use in giving its

[2] The term "authority" is ambiguous. Sometimes it refers to the person or institution that has authority, as when we say "The authorities closed down the streets around the bank after the heist." Sometimes instead, it refers to the normative power that a person or institution has, as when we say, "Political authority differs from religious authority". I will use the term in both senses and rely upon context to disambiguate it.

directives. The second concerns the type of argument that would justify a claim to authority.

Raz's **Dependency Thesis** is the following:

All authoritative directives should be based on reasons that already independently apply to the subjects of the directives and are relevant to their action in the circumstances covered by the directive.

The authority will consider more reasons for the directive than the reasons that already apply to the subject in the circumstances, but a constraint on his authority is that he must intend the directive to take the subject's reasons into account (p. 47). For instance, I have reasons for wanting affordable, high quality health care that I cannot lose, but medical care providers, insurers, and drug companies are also subjects, with different reasons which the authority must take into account in passing a law governing the way medical care is provided and paid for. Raz makes it clear that although the directive of legitimate authority must be intended to reflect the reasons the subjects have, it need not correctly reflect those reasons. The legitimacy of its authority is not dependent upon its issuing the right directive.

The normal way to justify authority is the **Normal Justification thesis (NJ thesis):**

> The normal way to establish that a person has authority over another person is to show that the alleged subject is likely better to comply with reasons that apply to him if he accepts the directives of the alleged authority as authoritatively binding and tries to follow them, rather than by trying to follow the reasons that apply to him directly (p. 53).

The idea here is that even an autonomous agent sometimes has reason to follow an indirect strategy in order to act on her own first order reasons. She is doing what she judges she should do on the balance of reasons, but in some cases the balance of reasons dictates obeying the authority. For example, I want to build a squirrel-proof bird feeder in my back yard, but I do not know how to do it. If I find an expert who offers his assistance, I have reason to follow his directions since I am more likely to act in a way that serves my ends if I do so than if I try to figure it out myself.

Raz proposes that the NJ thesis can justify the authority of the state. I have no position on whether he is right about that, but the NJ thesis

does plausibly link the reasonableness of obeying authority with reasons one has independent of the authority. A political authority is justified if it helps me to act on my duties to others. An extension of the same principle would justify religious authority if it helps me to act on my duties to God. Similarly, epistemic authority is justified if it helps me get the truth, thereby aiding me in satisfying a duty to myself. Since I have these duties anyway, it is reasonable for me to accept the authority as authoritative in a specific domain. And if the authority's legitimacy is justified by what it is reasonable for me to accept, then the authority's legitimacy can be justified in this way.

Authorities clearly differ in the degree of authority justified by the NJ thesis. It is reasonable for me to follow the directives of the bird feeder expert if I want a good bird feeder, but it is not reasonable for me to agree to be punished if I do not obey. It does not help me act on my first order reasons if I agree to such a strong kind of authority. But in other cases it does. I may do a better job of acting on some moral duties if I not only agree to obey a given authority, but also agree to be punished if I disobey. The authority of the state needs to be of this stronger kind. It must be strong enough to legitimate punishment for infraction of the laws.

There is one other thesis of authority that Raz considers, but rejects. He says that the Dependency thesis is sometimes confused with what he calls the **No Difference thesis:**

> The no difference thesis asserts that the exercise of authority should make no difference to what its subjects ought to do, for it ought to direct them to do what they ought to do in any event (p. 48).

Raz says this is false. If Parliament passes a tax law, it does not follow that the citizens had reason to pay the tax before the law was passed. If my Church commands me to go to Mass on Sunday, it does not follow that I had reason to do so anyway. If the bird feeder expert tells me to buy a certain kind of bracket, it does not follow that I had reason to buy a bracket like that in advance.

If we combine the NJ thesis and the pre-emption thesis, and formulate it from the point of view of the subject, we get the following **general thesis of authority**:

When I have reason to believe that following the directives of a putative authority makes it more likely that I will act on my first order reasons

than if I try to act on those reasons directly, I should do what the authority tells me to do pre-emptively.

My reason for accepting the authority does not depend upon the content of the authority's directive (content-independence), and the authority's legitimacy is constrained by the principle that it must intend to reflect my first order reasons in its directives (dependency thesis).

I will end this summary with an objection to pre-emption and Raz's reply. Someone might argue that the reasonable thing to do when an authority satisfies the condition of the NJ thesis is to weigh the authority's directive more heavily than my other reasons, but still take my other reasons into account. Why isn't the fact of the authority's directive simply one more reason to act that I put into the mix of my total set of reasons? The answer is that when I have reason to think that an authority satisfies the NJ thesis, I will do better at acting on my first order reasons if I follow the authority's directive pre-emptively than if I do not. Raz says:

> Suppose I can identify a range of cases in which I am wrong more than the putative authority. Suppose I decide because of this to tilt the balance in all those cases in favour of its solution. That is, in every case I will first make up my own mind independently of the 'authority's' verdict, and then, in those cases in which my judgment differs from its, I will add a certain weight to the solution favoured by it, on the ground that it, the authority, knows better than I. This procedure will reverse my independent judgment in a certain proportion of the cases. Sometimes even after giving the argument favoured by the authority an extra weight it will not win. On other occasions the additional weight will make all the difference. How will I fare under this procedure? If, as we are assuming, there is no other relevant information available, then we can expect that in the cases in which I endorse the authority's judgment my rate of mistakes declines and equals that of the authority. In the cases in which even now I contradict the authority's judgment the rate of my mistakes remains unchanged, i.e. greater than that of the authority. This shows that only by allowing the authority's judgment to pre-empt mine altogether will I succeed in improving my performance and bringing it to the level of the authority. Of course sometimes I do have additional information showing that the authority is better than me in some areas and not in others. This may be sufficient to show that it lacks authority over me in those other areas. The argument about the pre-emptiveness of authoritative decrees does not apply to such cases (pp. 68-9).

I think Raz's argument is sound. The decision to obey pre-emptively is the best strategy given the assumption that I have identified a person who is more likely to be right than I am about the best way to act on my own reasons for action. But we need models of the way to identify such a person. We will get to one way to do that in the next section.

Raz says he offers his work as a contribution to the literature on political freedom, and he aims to justify political authority within the framework of political liberalism (p. 1). He begins with the general realm of authority over action in order to present the contours of authority as he understands it, but he does so with an eye to applying his theses to the political domain. I suspect that this means he is operating with certain constraints.

One constraint is the desire to maximize political freedom and to minimize political authority. It is interesting that most modern political thought is motivated more by fear of bad authority than by desire for good authority. The idea is that it is more important to devise an account of authority that prevents tyranny than to give the bearer of authority the function of assisting the subjects in pursuing their individual and collective good.[3] With such an aim, it is reasonable to devise principles to restrict authority as much as possible, compatible with having a tolerably smooth-functioning society.

A second constraint is that the account must be applicable to authority over large populations with no presumption of personal trust between authority and subject. The strong personal bonds that exist in small communities such as the family or the village cannot be assumed when the authority is distant from the subjects and there is no personal interaction between them.

I am interested in the contours of authority in domains in which these constraints may not apply. Small communities with a great deal of personal interaction and bonding between authority and subject do not have the protection of the subject from the authority as a primary goal, which is not to say that subjects should not be protected. But I find it interesting to test the way in which Raz's principles would be applied in small, tightly-knit communities where the authority is stronger and

[3] Two exceptions are Yves Simon, *A General Theory of Authority* (University of Notre Dame Press, 1980), and Jean Bethke Elshtain, *Sovereignty: God, State, and Self* (N.Y.: Basic Books, 2008).

covers a much greater range of the subjects' lives than political authority does. If the community has a religious basis, is there something distinctively religious in the principles that justify authority in such communities? Could a modern liberal recognize the authority structure of these communities as legitimate? We are used to thinking of authority in the modern world as not only dependent upon certain principles deriving from the authority of the self, but as having a certain *result* – minimal intrusion in the lives of the subjects. In the next section I will turn to a model of authority that suggests that common assumptions about the difference between modern and pre-modern authority are false.

III. THE RULE OF ST. BENEDICT

St. Benedict wrote his Rule in the first half of the sixth century. More than a millennium before the advent of Constitutional government, Benedict understood the need for a combination of the authority of persons and the authority of a written rule known to all. This combination is important to the justification of authority in a Benedictine community, and it indicates one of the ways that modern liberal authority has separable elements. There are successful structures of authority that have some of those elements and not others.

Benedict's first words are "Listen, carefully, my son, to the master's instructions, and attend to them with the ear of your heart. This is advice from a father who loves you." (trans. Timothy Fry).[4] Think of the nature of this appeal. Benedict does not say, "I am an authority and I claim that you have a duty to obey me." He sets up a relationship of trust. "My words are addressed to you especially, whoever you may be, whatever your circumstances, who turn from the pursuit of your own self-will and ask to enlist under Christ, who is Lord of all, by following him through

[4] *The Rule of St. Benedict in English*, trans. by Timothy Fry (Collegeville, Minn: Liturgical Press, 1982). I use three translations of the Rule in the following discussion. The translations by Timothy Fry and Leonard Doyle follow Benedict in wording the Rule for communities of men. Since there were communities of women following the Rule from the beginning, the Barry translation explicitly words the Rule in a way that is applicable to these communities as well. My discussion will primarily focus on communities of monks, but will vary with the context of the translation I am using.

taking to yourself that strong and blessed armour of obedience which he made his own on coming in our world." (trans. Patrick Barry).[5] In the next few paragraphs Benedict reminds the monk of the first order ends they share. He says, "Let us open our eyes to the light that can change us into the likeness of God," and in the following paragraph he quotes the psalm, "Who is there with a love of true life and a longing for days of real fulfillment?" (Barry). In the following paragraphs Benedict reviews the aims of the Christian who follows the Gospel. "Brothers, now that we have asked the Lord who will dwell in his tent, we have heard the instruction for dwelling in it, but only if we fulfill the obligations of those who live there." (Fry). And in the concluding paragraph of the Prologue he says,

> Therefore we intend to establish a school for the Lord's service. In drawing up its regulations, we hope to set down nothing harsh, nothing burdensome. The good of all concerned, however, may prompt us to a little strictness in order to amend faults and to safeguard love. Do not be daunted immediately by fear and run away from the road that leads to salvation. It is bound to be narrow at the outset (Fry).

Notice that there are ends stated at the beginning of the Rule, and Benedict knows that each monk or seeker testing his monastic vocation has a reason to act for those ends. Within the space of a few paragraphs, Benedict appeals to the monk's first order reasons for living, and the monk's second order reasons for thinking that living as a monk under the Rule will help him live in the way he aims to live. The rules of Christian life are detailed in Chapter Four. These are rules that apply to the Christian whether or not he decides to enter a monastery. Many readers have a second order reason to take the authority of the Rule as a better way to act on the reasons given in Chapter Four than if they attempted it on their own.

Chapter One presents alternatives to the authority structure of a monastic community under a rule with an Abbot (or Abbess) and gives reasons for rejecting them. This is a smart way to begin. Benedict is aware that there are monks who either live without authority or live without

[5] *The Rule of St. Benedict,* trans. by Patrick Barry (1997), in *The Benedictine Handbook*, introduction by Anthony Merritt-Crosby (Collegeville, Minn: Liturgical Press, 2003).

a rule, and he knows his readers will be aware of them. He considers four kinds of monks. The anchorites (hermits) can be self-reliant and do not live under an Abbot, but that is only because they have already learned what they need to learn in a monastery. Otherwise, they will not be able to go it alone. For the sarabaites the law is whatever they want to do, whatever strikes their fancy. They are evidence of what happens when a monk follows his own will. The gyrovagues are the worst. They drift from place to place, taking from others, and never committing to a community. Since they move so frequently, their misdeeds are easier to hide. The crucial point of this chapter is that it is not enough for a monk to take a rule and attempt to live by it without living under an authority. The reason for living in a community as a cenobite is partly that living without authority tends to lead to gyrovagues and sarabaites, and partly that the tools of the spiritual craft require living in community. The goal of the individual monk is personal salvation, but Benedict calls the reader to trust him and his experience in judging that living under the authority of an Abbot according to the Rule is a better means to that end than following the monk's own will.

The Rule describes the qualities of the Abbot as well as the ordinary monk and those with particular roles like the Cellarer, the Porter, and the Prior. The Rule sets out most of the structure of the monastic life, including the hours of the Divine Office and the psalms that shall be chanted, the amount of food and wine to be consumed, the treatment of the sick, punishment for infractions of the Rule, and so on. It describes the way the Abbot is elected and the responsibilities of the Abbot to the monks, as well as the responsibilities of the monks to the Abbot, and the responsibilities of the monks to each other.

The Abbot is a teacher and spiritual mentor as well as a manager of the Abbey's affairs. A person who is good at one might not be good at the other, but it is clear that Benedict thinks that teaching is the more important role. The Abbot or Abbess "should give a lead to their disciples by two distinct methods of teaching– by the example of the lives they lead (and that is the most important way) and by the words they use in their teaching." (ch. 2, Barry). Benedict directs the Abbot to personalize his teaching and his discipline for the needs of the individual monk. One of his most touching pieces of advice to the Abbot is that he should "arrange everything that the strong have something to yearn for and the

weak nothing to run from." (ch. 64, Fry). Benedict clearly intends the Abbot to be trustworthy, and the monk's trust in him is grounded partly in the Abbot's behavior, but for the most part it is grounded in the personal qualities that make him an exemplar of a life lived by the Rule.[6]

The Rule satisfies and expresses all of the theses about authority defended by Raz. Benedict clearly implies that the Abbot's directive preempts the monk's other reasons for performing an act. Further, the command should be obeyed immediately. "It is, in fact, almost in one single moment that a command is uttered by the superior and the task carried to completion by the disciple." (ch. 5, Barry).

Raz's Dependency thesis states that the directives of the authority should be based on reasons that independently apply to the subjects of the directives and are relevant to their action in circumstances covered by the directive. Benedict implies acceptance of this thesis in a number of places. For instance, in the chapter on the election of an Abbot (ch. 64) he says that "They [the Abbot or Abbess] must understand that the call of their office is not to exercise power over those who are their subjects but to serve and help them in their needs." (Barry). It was radical in sixth century Rome to present an image of authority that exists for the sake of the subjects. The Rule restricts the purpose for which an Abbot may issue directives, and it reminds the Abbot several times that he is under the Rule himself and will have to answer to God: "And let the Abbot be sure that any lack of profit the master of the house may find in the sheep will be laid to the blame of the shepherd." (ch. 2, Doyle).[7] The Abbot may issue directives for purposes of his own, but that is compatible with the Dependency thesis, as Raz mentions. The crucial point of the thesis is that the authority must intend to take into account the reasons that apply to the subjects individually in giving his directives.

When the monk agrees to live under the Rule, he is under authority because he has accepted it, and his acceptance is grounded in reasons

[6] I believe that exemplars of virtue are not only crucial in moral training, but can play the central role in a form of moral theory I call exemplarist virtue theory. I outline such a theory in "Exemplarist Virtue Theory," forthcoming, *Metaphilosophy*, and give a detailed example of the way such a theory can be developed in *Divine Motivation Theory* (Cambridge University Press, 2004).

[7] *The Rule of St. Benedict*, trans. by Leonard Doyle (1948), (Collegeville, Minn: Liturgical Press, 2001).

he had prior to his acceptance of the Rule. Benedict offers the monk a second order reason to follow the Rule, and the Rule gives the monk a reason to obey the Abbot, and to accept the Abbot's authority to inflict the punishments given in the Rule for violations of the rule, including public apology for certain minor offenses, eating and praying alone for more serious offenses, and excommunication for the most serious offenses. So both the authority of the Rule and of the Abbot satisfy Raz's Normal Justification Thesis. Benedict says:

> It is love that impels them to pursue everlasting life; therefore, they are eager to take the narrow road of which the Lord says: Narrow is the road that leads to life (Matt 7:14). They no longer live by their own judgment, giving in to their whims and appetites; rather they walk according to another's decisions and directions, choosing to live in monasteries and to have an abbot over them." (ch. 5, Fry).

The Normal Justification thesis also implies conditions under which an authority loses legitimacy. Because of the modern suspicion of authority, this is an issue that now gets considerable attention, but Benedict has a different worry. His concern is that the fact that the Abbot is elected may not be enough to establish the legitimacy of his authority since the whole community might conspire to elect one who will consent to their "evil ways" (ch. 64). If that happens, it is the responsibility of the local bishop or ordinary Christians living nearby to intervene "to prevent so depraved a conspiracy." But a modern reader will ask, what if the Abbot is abusive and violates the Rule? There does not seem to be a formal procedure for such cases, but Benedict discusses visitations from other abbeys (ch. 61) and the fact that the Abbot will have to give an accounting to God for the way he governed the monastery. What Benedict expected was that the Abbot would become lax rather than tyrannical, and the history of monasticism indicates he was right. The many reforms over the centuries usually resulted in the monasteries becoming more disciplined, not less so.[8]

Benedict gives very clear directions on the qualities that an Abbot should have. He does not say that the Abbot loses his authority if he lacks these qualities, but he warns the Abbot again that there is a higher

[8] It is interesting that the only monastic order that has never needed reform is the Carthusians, which is also the strictest.

authority to whom he will one day have to answer. The Abbot has a strong motive to follow the Rule, and the Abbot is chosen from among monks who have already been following the Rule for many years. It is interesting that the Rule is regularly read aloud (ch. 66). Not only does the Rule specify that the Abbot must submit to the Rule, but everyone in the monastery knows that and is regularly reminded of it.

The Rule says that if the Abbot commands the monk to do something harsh, he should still do it. "The fourth step of humility is to go even further than this [obedience] by readily accepting in patient and silent endurance, without thought of giving up or avoiding the issue, any hard and demanding things that may come our way in the course of that obedience, even if they include harsh impositions which are unjust." (ch. 7, Barry). In a later chapter he gives the steps a monk should take if the Abbot commands something too harsh (ch. 68). But there is no doubt that the Abbot does not lose authority by giving such commands, and the monk's duty is to obey. Moreover, it is unlikely that this conflicts with the modern liberal notion of authority. Raz argues that an authority does not cease being an authority when the directives do not reflect the subjects' first order reasons for acting. It is only necessary that the authority attempt to do so (p. 41), and Raz makes it clear that even this is an ideal. It is sufficient if the authority attempts to reflect the subjects' reasons often enough to justify the power of the authority (p. 47). Similarly, an Abbot might not always attempt to issue directives that reflect the monks' first order reasons, and Benedict does not say that he loses authority in such cases. But egregious abuse of power would normally be detectable.

It is significant that the monk can leave virtually any time he wants. (If he does, Benedict permits him to return up to three times). The NJ thesis gives grounds for authority in reasons the subject has for thinking that accepting the authority is likely to have a certain consequence. This does not mean that the authority loses its authority if the anticipated consequence does not occur, but in monastic communities and other voluntary communities, the subject has the option of rejecting the authority by leaving the community. That gives him a stronger basis for accepting authority because casting his lot with the authority is not an irrevocable decision. If at some point in time, the authority's behavior leads him to reject his second order judgment about the authority, he is free to reject it by leaving.

Benedict, like Raz, rejects the No Difference thesis. Authority is not authority if it does not create new obligations. The point of the new obligations for Benedict is to create a special community in which Christian objectives and the following of the commandments of Christian life are easier to do. But it is important that obedience to authority in areas that do not directly pertain to his spiritual end as stated at the beginning of the Rule make it more likely that the monk will reach the spiritual end. So the monk may not see the connection in the particular case between the directive of the authority and the end he seeks. This might be a way in which the virtue of obedience in a monastic community differs from obedience to political authority. If the political authority issues a directive that seems to have nothing to do with their function, many of us would not trust the authority sufficiently to accept the directive as legitimate. For example, if the City Council issues an order that all persons must refrain from parking on the street on the first Tuesday of the month, citizens typically expect to have a reason to think that the directive is within the scope of authority they have agreed to assign the council– for instance, that that is street-cleaning day.[9] The citizen is under the council's authority in this respect even if the citizen disagrees with the need to refrain from parking for that reason, but the citizen needs reason to think that the council is issuing rules within the domain of its authority. In contrast, the authority of the Abbot extends to directives the relevance of which cannot be determined by the monk. But that difference, if there is one, is not in the conditions for legitimate authority, but rather in the trust that the subjects have in the authority. If the citizens trust the members of their city council to issue directives in the appropriate domain, their attitude towards those directives will not differ from those of the monk towards the Abbot.

Raz's theses of authority are general enough that they can ground both political authority in a modern liberal state and the tightly controlled community life governed by the Rule of St. Benedict. As far as I can see, there is nothing in the modern view of authority that would prevent a Razian liberal from becoming a Benedictine monk or nun. The liberal

[9] Presumably most of us would obey anyway to avoid a fine, but the question I am raising is whether we would think that we are obeying legitimate authority when we do so.

will be more aware of the limits of authority than we see expressed in the Rule, but I see no reason to think there is an incompatibility between the Rule and Raz's theses. Obviously, the conditions for agreeing to live by St. Benedict's Rule include much more than taking a certain line on authority, but it seems to me that those conditions have nothing to do with the view of authority *per se*. What is distinctive about Benedictine authority is acceptance of the first order reasons for living to which the Rule refers, and the degree of trust necessary to have a second order reason to accept the Rule as it is followed in a certain community as the best way to live by those reasons. What makes authority in these communities much stronger than political authority is the trust that binds authority and subject together, not the way in which the authority is justified or exercised. I see no reason to think, then, that political liberals should deny the possibility of a community that satisfies the conditions for liberal authority but has the radically pre-modern structure of a Benedictine community. The difference is in the degree of trust between authority and subject and the substance of the beliefs they share.

There is an objection to the justification of monastic authority that is sometimes made in the literature on autonomy, and is indicative of the common view that monastic authority is anti-modern. Thomas May discusses monasticism in the context of a discussion of whether a person can autonomously choose slavery.[10] He says the life of a monk in a rigorously disciplined monastery "closely approximates the life of a slave." (p. 138). The problem as May sees it is that the monk is not the "helmsman" in the determination of action.

> The voluntary slave does not determine his action. Rather, he only determines that his action will not be determined by himself. While in cases such as the monk's this determination *may* perfectly well be justified, he nonetheless surrenders his autonomy to whatever authority he is obligated to obey." (p. 139).

[10] Thomas May, *Autonomy, Authority, and Moral Responsibility* (Dordrecht: Kluwer, 1998), 138-9.

May continues:

> While it is true that the voluntary slave may "steer" toward slavery, once in slavery the slave is eliminated from the determination of his behavior. To voluntarily enter into slavery is tantamount to placing oneself under the direction of an "automatic pilot"; one may autonomously place oneself in this condition but cannot be said to "steer" while under this condition (p. 139).

May believes that the monk gives up his autonomy when he becomes a monk, and he seems willing to go farther than that and to say that the monk has agreed to become something that closely approximates a slave. The stronger claim surely is too strong since the monk's lack of control over the structure of his daily life hardly puts him in the category of a slave in his daily acts. It is not as if he decided to hand over to someone else each day the decision to tell him when to rise, when to chant the Divine Office, when to do his individual work, and so on. The times for each of these activities are prescribed by the Rule and the practice of the individual monastery, and since they are invariant, they were predictable at the time he became a monk. At that time he decided to get up at 5:30 (for instance), chant Vigils at 6, breakfast at 6:30, and so on. So he is hardly subject to another person's changeable whim.[11]

But May finds lack of autonomy in being on "automatic pilot." That suggests that he thinks of the monk as someone who mindlessly goes through a routine, thereby saving himself the trouble of having to make up his mind about what to do at each hour of the day. But that is a misunderstanding of monastic life. Nobody can be the "helmsman" for her every act, so presumably the autonomous person exercises direct control over some of her acts, and indirect control over others, and a person who acts on authority justified by the Normal Justification thesis exercises indirect control over the acts that fall under that principle. The issue is which acts are direct and which are indirect, not whether there are any instances of the latter.

The answer to that question depends upon what a person's first order reasons for action are. He might have reasons to act that are important

[11] Actually, the monk's level of direct control over his daily activities is at least as great as the control persons have in the typical eight-hour daily job. The hours in which the monk does not have direct control are just distributed differently.

to him, and he might judge that he is much more likely to act upon them if he does *not* exercise direct control over many day-to-day acts which, according to his own beliefs, are of trivial importance. In the chapter from which the above passage is taken (ch. 6), May defends the position that authority is compatible with autonomy, and he accepts the Normal Justification thesis.[12] So suppose a person has first order ends the probable attainment of which requires acting on authority in a domain including such things as when one rises, what and when one eats, and what prayers are said on certain days. Then, given that May accepts the NJ thesis, he should agree that that person would be justified in acting on authority in those cases. And since he believes that authority is compatible with autonomy, he should agree that the person would be acting autonomously. It seems to me, then, that May can deny that the monk is autonomous only if he thinks that the monk's first order ends themselves contradict autonomy. That would mean that autonomy is not a principle or value that says an agent should be governed by his own reasons; it precludes the possession of certain reasons. I find this implausible if autonomy is something valuable, and in any case, it goes counter to the presumption of political liberalism that persons are permitted to determine their own ends.

VI. MODERN AUTHORITY, MONASTIC AUTHORITY, AND RELIGIOUS AUTHORITY

Our foray into the Rule of St. Benedict has some interesting implications for the relationship between modern and traditional authority, and I think it can help us understand the justification of authority in the modern world outside the walls of a monastery, including communities that are neither monastic nor political.

Treatments of authority virtually always interpret modern authority as differing from pre-modern authority in (a) structure and degree, and (b) the principles used to justify it. Indeed, it is generally assumed that the contrast in (a) can be explained by a contrast in (b). That is, the

[12] May, 136-7. May argues for a revision of Raz's Dependency thesis in that chapter.

explanation for the striking change in the strength and range of authority in the modern period is that the modern liberal interprets authority as deriving from the individual's authority over herself, whereas nobody would have justified authority that way before the modern era. Instead, pre-moderns believed that all authority comes from God. I think that the Rule of St. Benedict shows that this assumption is mistaken. The structure of Benedictine authority is obviously pre-modern, yet it can be justified on the same principles that justify the modern liberal state. The differences in the outcome of the application of those principles must therefore be due to other features of the community over which the authority governs. I have proposed that the difference is in the content of the first order reasons for the sake of which authority is accepted, and the degree of trust between authority and subject. If I am right about that, liberal authority is no more justified by reference to the authority of the individual than is Benedictine authority. I am not, of course, suggesting that Benedictines should think of authority in their communities as justified by Razian principles. But those persons who accept such principles should judge that Benedictine authority is legitimate for those with the relevant first order ends. As I mentioned, they might personally disagree with those ends, but a mark of political liberalism is the presumption that individuals have the right to set their own ends. With that assumption, it is very hard to see on what grounds the political liberal can claim that authority in a Benedictine monastic community is illegitimate.

In spite of the fact that Benedictine authority is considerably stronger and has much greater scope than the authority of the modern state, it has some of the central features of a constitutional democracy. Authority exists for the sake of the governed, the Abbot is elected, and the Rule has a function similar to that of a constitution. We are used to thinking of these features as accompanied by other features of a representative democracy, and so it may be surprising to see how unlike a democracy an authority structure with these features can be. The difference in the trust between authority and subject is no doubt related to other differences, in particular, the fact that the constraints on political authority mentioned at the end of section II are not applicable in many communities. Political authority must apply to large and often diverse populations, and there is no escape from it. It cannot be assumed that the subjects have common ends, with the exception of the recognition of basic human rights, but

some of these are designed to protect the subjects from the authorities rather than from each other. Historical memories are of abuses of power, and so modern political thought tends to be motivated more by fear of bad authority than by desire for good authority. With these constraints it is no wonder that modern accounts of political authority attempt to not only ground authority in the rational will of the individual subject, but they also aim to minimize political authority as much as is feasible. In contrast, authority in a monastic community is not accepted as a necessary evil, but as a means to obtaining goods for the subjects, and the historical memory is not dominated by stories of tyranny. Because the communities are small and voluntary, they do not need to operate with the constraints necessary for governance over large and diverse populations. But if I am right, these features do not affect the principles justifying authority, but rather the form it takes in different communities.

This suggests that since Razian principles do not dictate that authority has a particular structure– e.g., a democracy, it is possible that there are other communities whose authority structure is justified on modern liberal principles, but which differ significantly from the structure of governance of the modern state because of a difference in the ends of the subjects and the level of trust within the community. Given that a subject can be reasonable in obeying pre-emptively the directives of a trusted individual, it also seems possible that a subject could have a reason to obey a trusted institution whose historical tradition is one with which she identifies, and which she reasonably believes is a better guide to her ends than she is when she is acting on her own. That is, she might reasonably trust the institutional embodiment of a tradition with an authority structure that issues directives governing some part of her life more than she trusts herself in that domain. If so, Raz's NJ thesis justifies her in acting pre-emptively on the commands of the institutional authority.

Institutions such as the Catholic Church have some of the features of political authority and some of the features of monastic authority. There are obviously many shared first order reasons for action within the Church, but the number of subjects is huge and the level of trust varies. But I suspect that authority in some religious institutions such as the Catholic Church might be justified on Razian principles even though the structure of authority in these institutions is nothing like that of the

modern liberal state. In any case, the way in which the structure and strength of authority is separable from its justification is worth further investigation for the case of religious authority.

I think that these considerations also suggest the desirability of further work on the connection between authority and autonomy, particularly in the case of religious authority, and probably also moral authority. Kant thought that the autonomous person needs self-discipline in order to avoid having a heteronomous will, a will determined by inclination– either one's own or that of a dominant other. The autonomous person is rational, and it takes self-control to be rational. Sometimes the autonomous person needs to have the self-discipline to obey when her own reason tells her to do so. At least, I have not yet seen a convincing argument that a monastic does not have such a reason, nor that monasticism takes away autonomy. If it does not, that leaves open the possibility that there are other forms of authority that are compatible with autonomy.

IS RELIGION UNDERMINED BY EVOLUTIONARY ARGUMENTS?

LOUIS CARUANA

Heythrop College, University of London

Abstract. I examine three major anti-religious arguments that are often proposed in various forms by cognitive and evolutionary scientists, and indicate possible responses to them. A fundamental problem with the entire debate arises because the term "religion" is too vague. So I reformulate the debate in terms of a less vague central concept: faith. Referring mainly to Aquinas on faith, I proceed by evaluating how the previously mentioned cognitive and evolutionary arguments fare when dealing with faith. The results show that some aspects of the concept of faith are in principle beyond the range of evolutionary explanation and some other aspects are not. Nevertheless, an evolutionary account merges smoothly with faith's theological dimensions.

This paper deals with the epistemological foundations of religion. The basic question is: "Does evolutionary explanation undermine religion?" To answer this question, I will start by illustrating how evolutionary explanation has been extended from biology, its natural habitat, to religion, its final frontier, according to some. Current studies in this area proceed in their overall approach on the paths already traced by various predecessors that go all the way back to Charles Darwin himself.[1] I will focus mainly on the current situation. In the first section, I will argue that underneath the arguments issuing from these evolutionary accounts, whether for or against religion, there lies a major problem that often goes unmentioned. This has to do with the very notion of religion. Could it be that more progress becomes possible in this area if, instead of

[1] Typical current studies include Boyer (2001), Atran (2002), Plantinga (2002), Sloan Wilson (2003), Schloss & Murry (2009). Major precedents include Teilhard de Chardin (1955), and Darwin himself in Darwin (1874).

working with the central notion of religion, we work with the narrower notion of faith? To explore this sub-question, I dedicate the second section to a clarification of the notion of faith, building up on the foundations set by Aquinas. Then, in the third section, I will determine the extent to which the evolutionary accounts mentioned before apply with more profit to the understanding of faith. My hope is that the analysis will allow me to articulate a plausible version of the notion of faith consistent with evolutionary anthropology.

I. EVOLUTIONARY ACCOUNTS OF RELIGION

Although the literature is vast, one may safely say that the core ingredient of this research program is simple. Since evolutionary principles have been successfully applied to explain human biology, they should also be applied to explain religion. These principles include three main elements. In brief, we need *random mutation* of a *hereditary* trait that is crucial for *survival*. Once these three elements are present within a self-replicating system, natural selection occurs in the long run. Of course, for any given organism, there can be many traits that satisfy this triple condition; consequently, natural selection may occur simultaneously at various fronts as regards the same species. For religion to fall within the range of this kind of explanation, it must have some feature that satisfies the triple condition. Hence, evolutionary psychologists interested in offering an evolutionary explanation of religion must first locate an aspect of their object of study that is evolutionarily relevant in this sense. Once they fix this, they will be able to tell a story whereby the presence of this aspect can be seen as the product of natural selection. An evolutionary explanation of religion would thus be available. Two broad camps will be considered: one holds that some aspects of religion are genuinely located within the range of evolutionary explanation. These aspects are adaptive, in the sense that they confer a survival advantage on the organisms that have them. The other camp holds that there is no such evolutionarily relevant aspect of religion. All aspects of religion are neutral. Religion confers neither advantages nor disadvantages for the survival of the organism.[2]

[2] Other kinds of argument are available. For instance, it is also possible to argue that religion confers serious disadvantages for survival, but that these disadvantages are

Let's start with the adaptive camp. The evolutionarily relevant aspects of religion most commonly referred to in this context are two: one involving supernatural agents, and the other involving super-knowing judges. The first one is the one most widely discussed. Child psychologists have discovered that small infants perceiving moving objects or dots on a screen readily attribute agency to these items. They see purpose everywhere. This is taken to indicate an intrinsic mental procedure constitutive of human nature. To refer to this procedure, researchers use the expression "hypersensitive agency detection device", HADD for short (Atran 2002; Barrett 2004). This expression is somewhat loaded with materialist overtones, especially evident in the term "device." To remain neutral with respect to thorny debates in the philosophy of mind, it is better to use "disposition" rather than "device", with the acronym remaining the same. With this in view, it is not difficult to see how this cognitive feature enhances fitness. Organisms that have it show a strong tendency to believe in the existence of agents when they detect certain kinds of stimuli. They are thus wary of various circumstances – wary of circumstances that are genuinely dangerous and wary also of circumstances that are not. Although this cognitive feature delivers a lot of false positives, it definitely represents the safer strategy. In the long run, it is better to err on the side of being cautious than on the side of being reckless. Some researchers working on HADD proceed then by affirming that religion is essentially constituted by this disposition. It is an extrapolation of HADD. Humans have religious beliefs because they have been naturally selected according to HADD. Those of our ancestors with religion survived. Those without it were filtered off.

Another aspect of religion that is considered in this context is associated with the belief in invisible but super-knowing judges, in other words, the belief that great superhuman forces are observing and judging all that people do. This stance is associated with the origins of morality and cooperation. As has been shown, cooperation within groups can be the result of natural selection when two principles are in operation: kin-selection and reciprocity. This explanation is vulnerable because of one problem. A cooperative group remains vulnerable to free-riders or

---

counterbalanced by other traits. For lack of space, this kind of argument is not being considered in this paper.

cheaters: those who receive benefits from the group without contributing. Such instability undermines the survival value of cooperation. Being a cheater, of course, needs the capacity to deceive, and this capacity varies depending on the intelligence of the organisms under consideration. For the hominid family, therefore, it seems that cooperation should have disappeared early on, undermined by ever more effective cheating. The fact that it didn't shows that there's something else in the equation, something that trumps the effect of cheaters. And this is precisely the element that interests us here. The amazingly high degree of cooperation between humans shows that there is some pressure that acts against deception. So we postulate a cause for this pressure: it comes from a belief that, even though other humans are not aware of the cheating, punishment will still be delivered. There are superhuman forces that observe and judge everything people do. Selfishness and cheating must therefore be curbed for fear that such forces will deliver their punishment. The deliverer of such punishment is referred to by various terms, including, for instance, gods, witches, and dead ancestors.[3] As a corollary, this view entails that ill fortune is a sign of hidden moral misconduct. The bottom line is that humans are religious as an outcome of their being cooperative, and they are cooperative because of an innate primordial belief in the existence of super-knowing agents who ensure that justice is done.

So up to now we have mentioned two possible candidates for an adaptationist view of religion. As I said, defenders of these two candidates form one camp. The other camp consists of those who contest the adaptationist view. They argue that religion in itself offers no survival advantage whatsoever.[4] They argue that religion is an epiphenomenon, a by-product that in itself has no adaptive value but is associated with some traits that do. The term often used in this context is "spandrel." This is an architectural term referring to a structure that arises within a building not because it is needed but because of other structures that are. An example of a biological spandrel is the sound of the heartbeat. What is evolutionarily relevant as regards the heart is the way it pumps

[3] In this section, I draw from, Johnson & Bering (2009). The argument is not that this evolutionarily-relevant aspect is the only feature that ensures cooperation. There may be other factors that block cheating or enhance cooperation. The argument, however, does say that the idea of a super-knowing punisher is the major relevant feature.

[4] A very clear case is made in Boyer (2001).

blood efficiently, not the sound it makes. Hence the production of sound gets a free ride, as it were, all along the evolutionary development of the heart. It is not filtered off because, as regards natural selection, it is invisible. In the same way, religion may be a free rider. It may be a phenomenon that exists simply because it "rides" on other human features that confer survival advantages, while itself being irrelevant from the point of view of natural selection. Some researchers have tried to gather evidence and build a case for this possibility, but expert scientific opinion shows no clear agreement on whether religion is indeed epiphenomenal. One might always object, for instance, that new evidence may turn up showing that some survival advantages are indeed associated with the sound of heart-beat. And this shows that all epiphenomena are judged to be so only because of current limitations in evolutionary knowledge. Nevertheless, in line with the argument of this paper, I would like to take this view as the third major possible account. I think it is reasonable to say that religion *could* be an epiphenomenon.

These three possible evolutionary accounts of religion are often presented with anti-religious strings attached. Researchers working in this area often present their results as undermining the validity of religion in some way or other. They take their results to show that religious claims are unacceptable and that religious practice is either detrimental to society or utterly futile. How do they do this?[5] Their critique takes various forms. Consider the first typical argument:

(A1) Religion is unacceptable because it is essentially a consequence of HADD,
a mental operation that exaggerates the detection of agency.

The force of this critique arises from the fact that HADD is unreliable. It exaggerates. And it is precisely because of this exaggerating tendency that it had been selected. On closer scrutiny, this objection loses a lot of its impact. Its weakness arises mainly because the critique makes no reference to the obvious fact that agent-detection happens all the time in the life of a normal human. Moreover, there is no clear dividing line between cases where agent-detection is readily verifiable and cases where

[5] For further details on these arguments, see Murray (2009).

agent-detection is not readily verifiable. There is no clear dividing line between the use of HADD as regards not-yet-observed entities and its use as regards unobservable entities. Calling HADD unreliable should make us abandon it even for everyday scenarios. But abandoning HADD is not possible. We have stipulated at the very start of the argument that this disposition is advantageous for survival. So argument A1 should not worry defenders of religion too much.

Consider another typical criticism:

(A2) Religion is unacceptable because its central beliefs are not caused by the entities they talk about; they arise from mechanisms that function independently from these entities.

This critique has bite only within the context of epistemological externalism. Within this context, knowledge claims and beliefs must have a proper causal relationship with the mind-independent world they talk about. Argument A2 presses the point that religion, as described by evolutionary theories, is not caused by the objects it talks about. And this undermines belief about them. Those who want to block this critique have at least one clear option. They can press the point that, if we assume that religious beliefs have been naturally selected, we are saying that they confer an advantage on those who have them. Take vision as an example. Vision confers evolutionary advantages on those who have it. It is not just vision that confers these advantages. It must be truthful vision. Mutations that result in organisms that see things where there are no things to be seen are dead ends. Such organisms are filtered off. Hence, vision's adaptive attribute is related to the realism involved in seeing. This example can be extended to reason and to the mind in general. Correct reasoning confers evolutionary advantages because inferring and deducing correctly makes good reasoners survive where confused reasoners die off. It should be evident, therefore, that once we assume at the start that having religious beliefs is adaptive, we are committed to some kind of efficiency with respect to these beliefs. If kidneys and hearts are effective in their domain, it is very plausible to hold that vision and intelligence are also. And, if religion is a special expression of affective and intelligent engagement with the world, then religion can be considered acceptable on these grounds.

A third critique of religion arises from the evolutionary account that takes it to be a useless by-product. In short, the objection could be phrased as follows.

(A3) Religion is unacceptable because it is epiphenomenal.

The basic point is that religious beliefs haven't proved their mettle by going through the filtering processes that shaped hominid evolution. They are just irrelevant fossils of primordial, mental meanderings that never washed away. Can a defence of religion be mounted against this argument? A possible move could start from the fact that A3 dismisses much more than religion. It obliges us to consider all non-instrumental thinking epiphenomenal. And this leads to all kinds of problems. What constitutes human nature is vastly determined by non-instrumental, symbolic thinking – by poetic and metaphorical language, art, music, literature and sculpture. Once we accept A3, it becomes obvious that evolutionary theories can account for only a very small section of the human phenomenon. It would be naïve to claim that such an account exhausts all there is to say about human beings. It would therefore be equally naive to argue that religion is unacceptable because it lies beyond the range of evolutionary explanation.

None of the arguments and counter-arguments discussed so far is a knock-down argument. They all invite further discussion and analysis. The main reason to include them here is not primarily to give the reader a taste of what is going on in this area of philosophy. It is rather to highlight an overriding problem with all such arguments. They all assume that the object of discussion, namely religion, is clear and precise enough to allow fruitful debates about it. But this assumption is misleading. To see why this is so, I propose excavating below the surface of these arguments so as to uncover what is happening at the level of the very definition of religion.

Let me characterize the individual steps taken by typical anti-religious evolutionary psychologists that end up arguing in the form of A1, A2 or A3.

(S1) They first acknowledge, consciously or unconsciously, a "received concept" of religion, call it R (it is often taken to be

a complex agglomeration of attitudes, practices, rituals, habits, and so on, many of which can only be described by using vague predicates).

(S2) They extract from R an aspect that is evolutionarily relevant; call this Q.

(S3) They show how natural selection successfully explains Q.

(S4) They claim that the success of S3 shows either that the really significant aspect of R is Q; or that we are now, finally, able to define R properly, namely as Q (hence, all there is to R is Q).

The basic idea behind this sequence is clear. The general strategy is to move from what may be called a hazy knowledge of something to a clear knowledge of it. What we have here is a refining process. But does the success of S3 really justify the reduction occurring in S4? I argue that it does not.

To see why, consider first, as an example, the love between husband and wife. Since time immemorial, love has been of central importance in human existence: as is evident in literature, drama, art, philosophy and religion. In the nineteenth century, experimental work in France and Germany established that special chemicals are responsible for a kind of communication between different organs within an animal. At the Royal College of Physicians in London, the word "hormone" was then launched as the standard term for these chemicals, and since then immense interest has been generated in various related research fields. As regards the specific issue of human love, the excitement and arousal love involves became understandable as the effect of hormones rushing from cell to cell along the blood stream, coordinating the action, perception and feeling of different parts of the body. Now, if one were to suggest that, because of these discoveries, this empirical account should be enshrined as the real meaning of love, not many would be convinced – and with good reason. The depth of meaning associated with love remains to a very great extent completely unaffected by these empirical discoveries, important as they are. When love between a husband and wife is described as real, nothing is being said about the chemicals involved in that experience. What is being said is situated at another level. The scientific discovery changes the broad concept and experience of love only at one

tiny spot, if it does at all. And it makes no sense to have reductionism smuggled into the story.

The case of religion is similar. Suppose that evolutionary psychology has established to everyone's satisfaction that religion is definitely either an instance of HADD or an extrapolation of the idea of a super-knowing punisher, as explained above. Should this claim oblige us to reduce the concept to this? Should we say that now we have discovered what religion *really* is? On the strength of the analogy with the example above, the answer is no. Nothing stops us from accepting both the newly discovered aspect and the other associated meanings of the term not in direct contradiction with the newly discovered aspects. This is as far as we can go with the analogy.

One disanalogy with the example of love is very important. Whereas the empirical substrate of love can be considered a genuine part of the study of love, the case of religion is different. The concept of religion seems to be situated within our conceptual scheme so as to enable us to refer to what people do when they reach out beyond what is empirically graspable. There are, of course, empirically graspable religious body movements and social practices. But the essential feature in religion seems to be the *meaning* attributed to these empirically graspable features, not the features themselves. Moreover, whereas the concept of love can perhaps be seen as one single concept, usable and reusable in various contexts, the concept of religion is not straightforwardly unified. We need to admit that the term religion is vague in the logical sense. It has no clear boundaries. What we call religions indeed share similarities between them but there is no guarantee that a core of features is shared by all. It is very probable that there is no such common core. The more data cultural anthropology delivers, the more the idea of a common essence of religion seems to drift away; and it is certainly not useful to cover this up by looking away from religions as they are and speculating on how they should be. The concept of religion is like the Wittgensteinian concept of a game, or even worse – worse because some religions involve a very high degree of self-reflection and self-adjustment, which implies an ongoing transformation of their very nature.

Should we give up hope then as regards trying to understand religion with the various modes of inquiry at our disposal? Not necessarily. I suggest the following strategy. Since religion is such a vast and elusive

category, a safer method is to carve out a smaller, more manageable unit and work with that. Faith is one of these units. The aim of this inquiry, therefore, becomes to determine the insights that can be drawn from evolutionary psychology as regards faith. The first step in this new venture will take us right across to the other end of the spectrum. From the scientifically charged explanations discussed so far, we move to theologically charged accounts of faith. This preliminary discussion will establish a good working definition of faith. Only when we have this can we determine whether evolutionary explanations can be successful in this context or not.

II. THE NATURE OF FAITH

To answer the question "What is faith?" we can adopt various strategies. One of them is to ask typical religious persons for their own version of the nature of faith. This method delivers a phenomenology of faith, and what these people report would constitute, as it were, first-hand data with respect to which theories may be constructed and evaluated. The result of such a phenomenology of faith, however, suffers from its high degree of subjectivity and context-dependence. Another strategy is to determine the structure of the concept of faith. One does this by analyzing the grammar of faith-involving linguistic expressions. We determine where and how these expressions behave properly and where and how they break down. This is the typical method of analytic philosophers.[6] Yet another method is to look back and see what the major thinkers of the past have written on this topic and start from there. This strategy is not completely different from the other two. The major philosophers of the past, of course, were themselves employing phenomenology or conceptual analysis, or both. In this paper, I am adopting the third strategy, concentrating on Aquinas, with the occasional allusion to the second, especially when indicating how technical terms of past centuries translate into those in current use.[7]

[6] For examples of this method, see Audi (2008), Wolterstorff (1990).

[7] Some evolutionary psychologists would, of course, find this method unacceptable. They present their views as *the* correct account of religion. They insist that what practitioners of religion like Aquinas say about it is irrelevant (e.g. Boyer 2001, p. 262-3). But

Thomas Aquinas builds his account of faith on two foundations: Aristotelian psychology, involving the categories of will and intellect, and Augustinian theological ideas, summed up in the formula "to believe is nothing other than to think with assent".[8] I will summarize the Thomist view in three points. First, he accepts Aristotle's view that the human intellect can sometimes be determined by the will. This does not mean that, for Aristotle, I am entitled to believe the world to be so just because I want it to be so. He is referring to cases when we assent to a proposition even though we do not have enough evidence for it. In this context, we can take assent to mean, quite simply, agreement with or compliance to a proposal.[9] For Aquinas, faith is such an act. It is a case of the will determining the intellect. In normal circumstances, will-motivated assent is partial or conditional. We say, "I assent to this claim because it's probably true" or "I assent to this claim because it is a logical consequence of another claim that I take to be true." In the case of faith, however, the assent is complete and unconditional. Believers declare, "We believe in God". They don't declare, "We take it that there's probably a God." In what way does the assent of faith differ from normal assent? There are two main differences. The first difference concerns intellectual dissatisfaction. This kind of dissatisfaction arises when a person assents to a claim that, as far

---

this is mistaken. It is as mistaken as its opposite, namely that only practitioners of religion can explain religion. The complexity of religion should make us beware of simplistic reductionism. Truth should be welcome whatever its source. The more viewpoints on the issue, the better. Moreover, we should not forget that Boyer and his colleagues are themselves explainable by sociology of science in ways that may perhaps surprise them.

[8] "Credere, nihil aliud est, quam cum assensione cogitare". St Augustine, *On the Predestination of Saints*, Book I, chapter 5 (*Patrologia Latina*, vol. 44, col. 963).

[9] This key-term "assent" is certainly related to the term "acceptance," which is quite prominent in current philosophical discussions. Their meaning, however, is not the same. Accepting that *p* means endorsing *p* deliberately, or taking *p* to be true for the sake of the argument, even if you do not believe it. Assenting to *p* has wider scope; you can be assenting to *p* even if you are not at present conscious of doing so. Further analysis is needed to situate assent with respect to acceptance and belief, but this lies beyond the scope of this paper. Audi (2008) is a good starting point. Moreover, I here take the will to refer to a person's faculty of choice – certainly not to a part of the person's body. Evolutionary psychologists would perhaps want to undermine such an Aristotelian category by breaking it down into small parts, perhaps by endorsing some form of modularity of mind. This will not affect the account of faith I describe here. Whatever the background structure of willing, it is undeniable that people sometimes will to believe a given proposition.

as that person knows, doesn't enjoy sufficient evidential support. How do believers deal with this? For normal cases of assent, this dissatisfaction is appeased by the believer seeking more evidence. For faith, however, the dissatisfaction is pacified by the believer acknowledging that he or she is acting out of trust. The believer believes not because of evidence, but because of the authority of God who is the guarantor of revelation. In faith, the will directs the intellect to assent because of its trustful submission to God who is recognized as infallible and omniscient. There is a categorical difference then between the certitude associated with natural sources of knowledge and that associated with faith. The former is the fruit of evidence and demonstration; the latter the fruit of risk that one deliberately takes on because of a loving relationship.

The second point to highlight is the way Aquinas exposes the tension that exists in the very concept of faith. This point is fundamental for my argument in this paper, because in cognitive accounts of religion it is often neglected. Aquinas insists that faith involves not only the person who has it but also God whose extra help in this matter is indispensable. The object of the assent involves a set of truths that go beyond the grasp of natural reason or of science as we understand it today – Aquinas is here referring to truths like the truth of the Triune nature of God. Consequently, the act of assent must involve a contribution that comes from beyond the natural capacities of humans. Moreover, there is no neat, two-tier structure within the idea of faith. Aquinas denies that there is a purely human contribution, presumably explainable by evolutionary psychology, and then an added divine push. To be consistent with biblical texts, he insists that God is responsible for the entire act of faith:

> The Pelagians held that this cause [i.e. the internal cause that moves man inwardly to assent to matters of faith] was just man's freewill, and consequently they said that the beginning of faith is from ourselves, inasmuch as it is certainly in our power to be ready to assent to things which are of faith, but that the consummation of faith is from God, from whom are proposed the things we have to believe. But this is false. Since man, by assenting to matters of faith, is raised above his nature, this must be added to him from some supernatural principle moving him inwardly; and this is God. Therefore faith, as regards the assent, which is the chief act of faith, is from God moving man inwardly by grace.[10]

[10] In this paper, quotes from Aquinas are my translation, using the online version of *Summa Theologiae* in *Corpus Thomisticum* (Fundación Tomás de Aquino, 2000-2009),

The third important point to highlight becomes evident when Aquinas moves beyond the idea of faith as involving assent to the idea of faith as a virtue. Up to now, I have focused on how he equates faith with the act of believing particular truths of revelation. This is only part of the story. For a comprehensive grasp of his view, we need to add his distinction between *fides* and *credere*. *Fides* is a virtue while *credere* is the act of assenting to a proposition. In general, a virtue is an excellence of a human faculty or habit directed towards the good. Faith is a twofold excellence. It involves the excellence of the intellect as it seeks absolute truth, namely God; and it involves also the excellence of the will as it seeks absolute good, again God. He writes: "For since to believe is an act of the intellect assenting to the truth at the command of the will, two things are required for this act to be perfect. One of these is that the intellect should infallibly tend to its good, which is the true; while the other is that it should be infallibly directed to the last end, by which the will assents to the true. And both of these are to be found in the act of mature faith."[11]

With these three points in mind, we can summarize Aquinas's overall view using his own words: "to believe is an act of intellect assenting to the divine truth by virtue of the command of the will as moved by God through grace".[12]

It is useful here to digress a little by supplementing the above points with some ideas from the works of Martin Luther. Although the major, overarching debate in Luther's writings on faith concerns justification,

---

abbreviated henceforth as *ST*. For hidden nuances, the reader may want to consult the original Latin, included in footnotes. This text is from *ST* 2a2ae Q6 a1: "Hanc autem causam Pelagiani ponebant solum liberum arbitrium hominis, et propter hoc dicebant quod initium fidei est ex nobis, inquantum scilicet ex nobis est quod parati sumus ad assentiendum his quae sunt fidei; sed consummatio fidei est a Deo, per quem nobis proponuntur ea quae credere debemus. Sed hoc est falsum. Quia cum homo, assentiendo his quae sunt fidei, elevetur supra naturam suam, oportet quod hoc insit ei ex supernaturali principio interius movente, quod est Deus. Et ideo fides quantum ad assensum, qui est principalis actus fidei, est a Deo interius movente per gratiam."

[11] "Cum enim credere sit actus intellectus assentientis vero ex imperio voluntatis, ad hoc quod iste actus sit perfectus duo requiruntur. Quorum unum est ut infallibiliter intellectus tendat in suum bonum, quod est verum, aliud autem est ut infallibiliter ordinetur ad ultimum finem, propter quem voluntas assentit vero. Et utrumque invenitur in actu fidei formatae." *ST* 2a2ae Q4 a5.

[12] *ST* 2a2ae Q2 a9: "credere est actus intellectus assentientis veritati divinae ex imperio voluntatis a Deo motae per gratiam, et sic subiacet libero arbitrio in ordine ad Deum."

a debate which is primarily theological, one can identify a valuable core that is essentially philosophical and complementary to the Thomistic views expressed above. First of all, for Luther, faith is in stark contrast to works. His major claim is that it is wrong to attempt to worship God according to works only. Consider a typical claim of his: "this is the reason why our theology is certain: it snatches us away from ourselves, so that we do not depend on our own strength, conscience, experience, person or works but depend on that which is outside ourselves, that is, on the promise and truth of God, which cannot deceive."[13] We have here a clear expression of the idea that faith in the Word of God is not within our power. Faith demands that we let go of what our human knowledge delivers so as to cling only to the testimony of Christ. Secondly, Luther emphasizes the fact that faith offers freedom from human laws and traditions. The coercion and constraint of human laws make individuals obey through fear, concerned mainly with themselves. Faith has the effect of liberating such people so that they may obey and indeed engage in good works in selflessness and love. Luther writes: "the just man [i.e. the one with faith] lives as though he had no need of the Law to admonish, urge, and constrain him; but spontaneously, without any legal constraint, he does more than the Law requires."[14] Thirdly, Luther defends a notion of faith that is not concerned primarily with believing a number of propositions. He retrieved the original idea of faith as used by St Paul, namely the idea that faith is a characteristic of the entire person who is open to God. It refers to a state of a person rather than an attitude towards a set of propositions. The person of faith, on this account, is faithful to God, does God's will, hopes in God, believes what God says, and so on. Faith refers to the entire package. For Luther, "faith is God's work in us, that changes us and gives new birth from God […] Faith is a living, bold trust in God's grace, so certain of God's favor that it would risk death a thousand times trusting in it."[15]

Luther's focus is essentially on the lived faith. He is highlighting the third aspect of Aquinas' view, namely *fides* as distinct from *credere*. Both

[13] Luther (1955-1967), vol. 26, p. 387. See also Zachman (1993).

[14] Luther (1955-1967), vol. 27, p. 96.

[15] *Vermischte Deutsche Schriften*, ed. J.K. Irmischer Vol. 63 (Erlangen: Heyder and Zimmer, 1854), pp.124-125, trans. R.E. Smith (1994), URL: http://www.iclnet.org/pub/resources/text/wittenberg/luther/luther-faith.txt

authors converge on this point, a fact that highlights its importance. Of course, the brief sketch of faith presented up to now can barely do justice to Aquinas and Luther, and it is too short to include the various refinements their positions received since they were written. Nevertheless, even the points mentioned here are enough to indicate the basic, essential characteristics of faith, characteristics that ensure the right kind of coherence with biblical texts. So I will proceed by focusing on the area where the two views overlap. Whether expressed in terms of virtue or in terms of a way of life, faith for both can be described as follows.

(F) Faith is a state or attribute of the person as a whole, involving an exercise of freedom and a special divine initiative.

I will build on this formula, adding here just one further clarification: that the state referred to in this statement can be manifested in various ways. It can be manifested when the person assents to explicit propositions. It can be manifested when the person participates in organized celebrations or liturgies. It can be manifested when the person makes decisions. Moreover, if there is no manifestation of a person's faith, one cannot draw the conclusion that it is absent. All these observations follow from the idea of habit, of which faith is a special kind.

III. EVOLUTIONARY ACCOUNTS OF FAITH

I will now bring in the three evolutionary explanations discussed in section one: first, the account of religion as an instance of HADD; secondly, the account of religion as an extrapolation of the idea of a super-knowing punisher; and third, the account of religion as a fortuitous by-product with no relevance for fitness. Are these accounts valid and useful when considering faith as defined in statement F? There are three elements within this definition. It says that faith is an attribute of the person as a whole; it mentions freedom; and it mentions God's initiative.

The first element seems to be the most readily amenable to an evolutionary account. It speaks of faith as an attribute of the person as a whole, and thus it includes the various dispositions the person can acquire over time. Such dispositions, of course, can be of the individual person, or of

the particular group or tradition the person belongs to, or even of the species as a whole. As regards this aspect of faith, therefore, an evolutionary account can have great significance. It is certainly plausible to argue that HADD is responsible to some extent for a disposition to believe in the existence of a cause of the universe. Whether this disposition needs to be labelled with the term "hypersensitive", as if it involves an unjustified excess of something, may be disputed. But the point here is that there is no obvious contradiction between the evolutionary account involving HADD and this aspect of the notion of faith. The same thing can be said as regards the second evolutionary explanation, the one invoking the belief in a super-knowing punisher. It is plausible to argue that such a belief in a super-knowing punisher, which is a fruit of natural selection, is, to some extent, responsible for a disposition to believe in the afterlife and in the last judgment. There is nothing intrinsically contradictory in making this claim, as long as we do not add the reductive clause that faith is this disposition and nothing else. If we move on now to those evolutionary psychologists who want to take the third account and argue that religion emerged as an epiphenomenon, we can see also that they will have no direct clash with the concept of faith. They are just saying that faith is not a biological phenomenon, in the strict sense. Presumably, they would say the same thing as regards Beethoven's composing of the 9th Symphony, Michelangelo's painting of the Sistine Chapel, and Darwin's conceiving of evolution through natural selection. They would acknowledge that what happens at the level of culture lies largely beyond the explanatory reach of evolution by natural selection. This is in line with the way some philosophers of biology explain the specificity of *Homo sapiens*, whose appearance represents a crucial junction: the point where evolution gave rise to a system that is no longer within its range, a system that floats freely (Sober 1992). So, all in all, it seems plausible to hold that the first ingredient of definition F can merge smoothly with all three evolutionary accounts.

We move on now to the second and third elements within F. They are: the claim that faith involves an exercise of freedom and the claim that it involves a special divine initiative. Can these two elements of faith be accounted for by an evolutionary explanation?

The short answer is no to both. The first element introduces the idea of freedom; and freedom refers to the capacity of rational agents to

choose consciously one course of action from among various alternatives. This point introduces one of the major distinctions between the empirical and the human sciences. The objects of study for the human sciences differ radically from those for the empirical sciences. The objects of study for the human sciences are themselves epistemic agents. They can be influenced by their own awareness of the results of the very inquiry that is allegedly describing them objectively. To further describe this phenomenon, I will use the established term "reflexivity" (Geuss 1981; Rosenberg 2007). In general, reflexivity occurs when a theory is such that the dissemination of its results will jeopardize its confirmation. Reflexivity occurs, for instance, when a theory predicts that oil prices will fall; that very prediction then makes agents in society act to falsify it by raising prices.[16] Such reflexivity embedded within the human sciences has given rise to the idea that these sciences should not be conceived of on the model of physics. They are inherently different. They are useless for prediction but indispensable for emancipation. They present before the human agents they study a mirror in which these agents can see themselves together with the ideologies that determine their action. The agents are thus enabled to look objectively on what till then had been unconscious, and to accept, reject or change their guiding principles. As regards faith, my claim is that reflexivity is central. It is not just present. It is constitutive of faith's very nature. Any hint that individuals have been coerced to be persons of faith, undermines their status of being persons of faith.[17] Since faith is an instance of freedom, evolutionary explanations can, at best, only be partial.

If we consider the other ingredient of the notion of faith, namely the requirement that it involves divine initiative, the situation vis-à-vis any

[16] It is interesting to recall that, in some domains, physics itself had to undergo revisions because of this very point. Paul Dirac explained the origin of indeterminacy in quantum physics by resorting to a similar scenario. There is a level at which the photons used to measure the position of a micro-particle displace the micro-particle in the very process of measuring. Hence classical physics needs radical revision, at least at this level.

[17] The centrality of freedom here points towards an interesting corollary about the role of reason. Individuals who feel obliged to believe that God exists because they consider themselves coerced by the necessity of logical arguments are not people of faith. Faith is possible only through freedom and trust in God.

possible evolutionary explanation is the same. The requirement here is that a correct understanding of faith includes God's action as the first mover of the person's process of assent. The reason for this, let us recall, is that, by faith, the person has access to truths that are not accessible naturally. Aquinas concludes: "therefore faith, as regards assent, which is the chief act of faith, is from God moving man inwardly by grace." Notice, first of all, that this element indicates how the concept of religious faith is not a mere extension of the general notion of faith, evident when we talk of someone, say, who has faith that a colleague will succeed. The notion of religious faith is to some extent *sui generis* because persons who have such faith have it on trust – trusting the very "object" of their faith. They do not first seek a convincing argument for the existence of God and then start having faith by trusting him. There is only one act involved. And God is acknowledged to be both its source and end. Now, it is clear that any evolutionary explanation can only range over what humans do. It cannot cover also what God does. So one might want to call a halt here to the entire inquiry. One might object to the idea of including God within the very definition of faith. Once God is in there, explanation is jeopardized. In other words, the very idea of explanation should be seen as requiring that God not be part of the explanandum. Various reasons could be brought forward for this requirement. Invoking God is not part of empirical science; invoking God is essentially explaining the unfamiliar by referring to something that is even less familiar; and invoking God offers false satisfaction, as it tends to undercut the motivation to seek the structure of secondary causes. This objection cuts deep. It shows that the evolutionary psychologist disagrees here with the very definition of faith we started with. At this juncture of the debate, there is apparently little more that can be done. The believer has one definition of faith. The evolutionary psychologist has another. They often seem to be talking about the same thing, but in fact they are not.

Is it really as bad as that? I would like to remove this impasse by drawing some insight from the clause Aquinas adds, apparently casually, in the statement just quoted. The clause in question is the qualification regarding the assent of faith. Aquinas writes: "faith, as regards assent, which is the chief act of faith, […]." He calls the assent the chief act of faith, *principalis actus fidei*. I take him to indicate thereby that there are other aspects of faith apart from the one he calls principal, other aspects

that he would call secondary. And I propose that among these other aspects there are some features that are perfectly explainable via evolutionary accounts.

To support this claim, I offer a sketch of a synthetic view, merging an evolutionary account with the theologically charged understanding of faith described in statement F. This can be derived from an analogy between faith and morals. As regards acting morally, one is sometimes predisposed to do the good by a spontaneous desire or emotion, as when a mother helps her child automatically. But acting in this way doesn't show the heights of moral virtue. In fact, acting in this way reduces the moral value, the praiseworthiness, of the act. The mother can be said to be fully mature in her moral virtue when she helps her child wilfully, after a deliberate acceptance of her natural and primordial, affective drive. She helps her child even, say, when he has grown up and is not responding favourably to her anymore. Notice that the predisposition of a mother to help her child is indeed explainable via evolutionary considerations – in the same way as the same disposition of many other animals towards their offspring.

Now what has been said as regards morals can also be said as regards faith. The human person can be predisposed to believe truths about the divine and to live accordingly. Such predispositions can be explainable as arising out of features like HADD, or out of the idea of a super-knowing punisher, or via some other feature not yet discovered. When the person has beliefs of this kind and lives accordingly, we can legitimately call that person a person of faith, but only in a very superficial or rudimentary sense. Such a person acts religiously because compelled to do so by dispositions for which he or she is not responsible. Being religious in this way certainly doesn't show the heights of the virtue of faith, and some would even say that, strictly speaking, the person here would only have a shadow of faith, not faith itself. Be that as it may, the crucial step occurs when the person starts to accept deliberately the general drive these dispositions suggest, and finally arrives at accepting the divine with the right motivation, namely love of God.[18] My argument here resembles the

[18] I am not suggesting that there is literally a definite stage in a person's life when desires change into *conceptualized* desires, allowing the person to choose to act in line or not in line with them. The transition is gradual as the individual grows to maturity; and, even then, desires and dispositions become conceptualized only when they are the focus

thinking of Aquinas, but is not exactly like his. Consider his concise way of putting it:

> […] in moral virtues, a passion which precedes choice diminishes the praiseworthiness of the virtuous act. For just as a man ought to perform acts of moral virtue on account of the judgment of reason, and not on account of a passion, so ought he to believe matters of faith, not because of human reason, but because of Divine authority.[19]

Aquinas is here using the analogy between morals and faith. He is effectively drawing the analogy by saying that passion is to reason (for morals) as reason is to trust in God (for faith). So, in his version, passion, as such, does not appear on the side of faith. I am suggesting that it could. A passion may indeed be involved in faith. Of course, the term "passion" here is to be understood in the medieval sense, namely as a received disposition. I am suggesting that primordial dispositions, hammered into the human species through millennia of natural selection, are the rudiments of faith. I concede that the dimension I am introducing could be contestable. The main contention could arise because the preambles of faith, *praeambula fidei*, have traditionally been associated exclusively with a set of truths, truths that natural reason can know of God independently of revelation.[20] They were not taken to include dispositions. My suggestion is in fact saying that these preambles include not only a set of truths but also some basic somatic dispositions. In other words, I'm urging that some bodily habits are conducive to the faith, and are presupposed by it. Such bodily habits are the material, as it were, on which faith flourishes. The suggestion is justified to the extent that faith, as expressed in statement F, is a state or attribute of the person as a whole: body, mind, community. If this final step in my reasoning is correct, the overall bottom line is that evolutionary explanations can indeed

---

of attention. It is arguable that this change from being desire-driven to being reason-driven occurs also gradually all along the evolution of the *Hominidae* family.

[19] *ST* 2a2ae Q2 a10: "[…] passio praecedens electionem in virtutibus moralibus diminuit laudem virtuosi actus. Sicut enim homo actus virtutum moralium debet exercere propter iudicium rationis, non propter passionem; ita credere debet homo ea quae sunt fidei non propter rationem humanam, sed propter auctoritatem divinam."

[20] The nature of the *praeambula fidei* is not a completely settled issue. For current contrasting views, see McInerny (2006), Wippel (2000).

merge smoothly with a theological understanding of faith, even though they account for some of its aspects and not all.

CONCLUSION

My original aim in this paper was to answer the question: "Does evolutionary explanation undermine religion?" I started with an overview of the way evolutionary explanation is being extended beyond biology to cover also the area of religion. I highlighted the three major arguments that are often used to undermine religion, but I argued that each of the three attacks has a possible counter-argument that neutralizes it. The fundamental problem beneath these arguments and counterarguments is that "religion" is too broad a term to allow valuable analysis. So I proceeded by reducing some of the vagueness inherent within the inquiry. I did this by concentrating on the notion of faith rather than religion. To secure a working definition of faith, suitable for the kind of inquiry engaged in here, I extracted its major ingredients from Aquinas and proceeded by evaluating how the previously mentioned evolutionary accounts of religion fare when dealing with faith. The results showed that, although parts of the concept of faith are in principle beyond the range of evolutionary explanation, other parts are not. I supported this further by sketching an account of faith wherein an evolutionary explanation merges smoothly with its theological aspects. My original question therefore has not been answered but transformed. And as regards its new version "Does evolutionary explanation undermine the notion of faith?" the answer is no.

Acknowledgments. I thank the following for their comments on previous drafts of this paper: Gerard J. Hughes, Justin Barrett, Kelly James Clark, Dani Rabinowitz, and members of the Centre for Philosophy of Religion at the University of Notre Dame.

BIBLIOGRAPHY

Atran, Scott. *In Gods we Trust: the Evolutionary Landscape of Religion*. Oxford: Oxford University Press, 2002.

Audi, Robert. "Belief, faith, and acceptance." *International Journal for the Philosophy of Religion*, 63 (2008): 87-102.

Barrett, Justin. *Why Would Anyone Believe in God?* Walnut Creek, CA: AltaMira Press, 2004.

Boyer, Pascal. *Religion Explained. The Evolutionary Origins of Religious Thought.* New York: Basic Books, 2001.

Darwin, Charles. *The Descent of Man and Selection in Relation to Sex*, 2nd ed. New York: American Home Library, 1874.

Geuss, R. *The Idea of a Critical Theory*. Cambridge: Cambridge University Press, 1981.

Johnson, D. & Bering, J. "Hand of God, Mind of Man: punishment and cognition in the evolution of cooperation." In: Schloss & Murray (2009), 26-43.

Luther, Martin. *Luther's Works*, ed. J. Pelikan & H.T. Lehman, Philadelphia: Fortress Press, 1955-1967.

McInerny Ralph M. *Praeambula Fidei: Thomism and the God of the Philosophers.* Washington, D.C.: Catholic University of America Press, 2006.

Murray, M.J. "Scientific Explanations of Religion and the Justification of Religious Belief." In: Schloss & Murray (2009), 168-78.

Plantinga, Alvin. "Evolutionary argument against naturalism." In: *Naturalism defeated?* edited by J. Beilby, 1-12. Ithaca, NY: Cornell University Press, 2002.

Rosenberg, A. *Philosophy of Social Science.* Westview Press, 2007.

Schloss, J. & Murray, M.J. (eds). *The Believing Primate.* Oxford: Oxford University Press, 2009.

Sloan Wilson, David. *Darwin's Cathedral: Evolution, Religion, and the Nature of Society.* Chicago: University Of Chicago Press, 2003.

Sober, Elliott. "Models of Cultural Evolution." In: *Trees of Life: Essays in the Philosophy of Biology*, edited by P. Griffiths, 17-40. Dordrecht: Kluwer, 1992.

Teilhard de Chardin, Pierre. *Le Phénomène humaine*. Paris: Éditions de Seuil, 1955; trans. B. Wall, *The Phenomenon of Man.* London: Collins, 1959.

Thomas Aquinas. *Summa Theologiae*. In: *Corpus Thomisticum*. Fundación Tomás de Aquino, 2000-2009. URL: http://www.corpusthomisticum.org/

Wippel, J.F. *The Metaphysical Thought of Thomas Aquinas.* Washington, D.C.: Catholic University of America Press, 2000.

Wolterstorff, N. "The assurance of faith." *Faith and Philosophy* 7(4) (1990): 396-417.

Zachman, R. *The Assurance of Faith: Conscience in the Theology of Martin Luther and John Calvin*. Minneapolis: Fortress Press, 1993.

EPISTEMOLOGICAL FOUNDATIONS FOR KOONS' COSMOLOGICAL ARGUMENT?

GRAHAM OPPY

Monash University

Abstract. Some people—including the present author—have proposed and defended alternative restricted causal principles that block Robert Koons' 'new' cosmological argument without undermining the intuition that causation is very close to ubiquitous. In 'Epistemological Foundations for the Cosmological Argument', Koons argues that any restricted causal principles that are insufficient for the purposes of his cosmological argument cause epistemological collapse into general scepticism. In this paper I argue, against Koons, that there is no reason to suppose that my favourite restricted causal principle precipitates epistemological collapse into general scepticism. If we impose the *same kinds* of restrictions on causal epistemological principles and on principles of general causation, then we cannot be vulnerable to the kind of argument that Koons develops.

Koons (2008) argues for the very surprising conclusion that 'any exception to the principle of general causation [i.e., the principle that everything has a cause] that is narrow enough to avoid a collapse into global scepticism about empirical knowledge is also narrow enough to permit the construction of a successful proof of God's existence' (p. 106). While Koons supposes that there are two ways in which a 'principle of general causation' could be connected to the possibility of empirical knowledge—namely (i) as an objective fact needed as the ground for the reliability of our cognitive processes, and (ii) as a subjectively required presumption needed for immunity to internal defeaters—he does little more than sketch the beginnings of the development of an argument of the first kind, reserving almost all of his attention for the development of an argument of the second kind. We shall follow his lead.

I.

Here are the definitions, propositions, lemmas and theorems that make up Koons' argument for Theorem 2, i.e., for 'the main result' (p. 121) of his paper:

Definition 0: A person's *knowledge-net* consists in all of his belief states, together with their objects and those states, if any, that both cause one of those belief states and intermediate causally between it and its object (or between it and the common cause of it and its object).[1]

Definition 1: A proposition that q is a *rebutting defeater* of a proposition that p for an agent S *iff* S believes that q and the proposition that q provides S with adequate grounds for judging that p is false, even when combined with S's evidence for p.

Definition 2: A proposition that q is an *undercutting defeater* of a proposition that p for agent S *iff* S believes that q and the proposition that q provides S with adequate grounds for judging that it is not highly likely that the processes that led to his* disposition to believe that p are warrant conferring.[2]

Definition 3: A proposition that r is a *neutralising defeater* of the proposition that q in relation to the proposition that p for S *iff* S believes that r and that q, and the proposition that q is a defeater of the proposition that p for S, and the conjunctive proposition that r and q is not a defeater of the proposition that p for S.

Proposition 1: S knows that p *only if* every rebutting or undercutting defeater of the proposition that p for S is neutralised for S.

---

[1] Although Koons does not say this explicitly, I take it that, in his definition of a knowledge-net, he means to refer only to belief states that are also knowledge states: there are, for example, no false beliefs in one's knowledge net. The example that Koons gives to illustrate his definition might be taken to confirm this point: 'If S has perceptual knowledge of the fact that p by vision, then S's knowledge net includes his belief that p, that fact that p, and those states that are causally intermediate between these two, such as the reflection of light by the objects involved in the fact that p, the transmission of that light to S's eyes, the occurrence of nerve signals between S's retina and brain, and S's visual impressions as of the truth that p.' (p. 111)

[2] Koons adopts Castañeda's convention of using an asterisk to indicate *de se* attributions of attitudes. I follow this same convention throughout my paper.

Proposition 2: S knows that p *only if* S is in a position to believe that p with internal justification.[3]

Proposition 3: S is in a position to believe with internal justification that p *only if* S is in a position to believe with internal justification that it is very unlikely that q is true, for every available proposition that q that would, if believed, be an unneutralised undercutting defeater for S of the proposition that p.[4]

Proposition 4: S is in a position to believe with internal justification that p only if S is in a position to believe with internal justification that it is highly likely that his* belief that p is warranted.

Proposition 5: If S's belief that p is not strongly *a priori* justified or self-verifying—i.e. if S's belief that p is an ordinary empirical belief—then S is in a position to believe with internal justification that p *only if* S is potentially in a position to believe with internal justification that it is highly probable that his* belief that p is warranted—i.e. formed by a normal and alethically reliable process—in such a way that S's belief that p would depend for its internal justification on the justification of the latter belief.[5]

Proposition 6: It is evident that—with the possible exception of strongly *a priori* justified beliefs and self-verifying beliefs—any belief that is uncaused or whose epistemic grounds are uncaused is not warranted, because such a belief is not then formed by a normal, alethically reliable process. Moreover, the proposition that some or all of his* beliefs are uncaused is available to S.

Lemma 1: If S's belief that p is an ordinary empirical belief, then S knows that p *only if* S is in a position to believe with internal justification that it is

[3] Koons says 'is in a position to believe that p' rather than 'believes that p' because he wants to allow that people can have knowledge in cases in which they don't actually have internal justification for believing that p but in which they do have what it takes to have internal justification for believing that p.

[4] Koons does not say what it is for a proposition to be 'available'. I assume that what he has in mind is that internal justification for belief is not compromised by inability to deem unlikely propositions that one cannot even grasp.

[5] Koons tells us that 'a belief is justified in a *strongly a priori* way iff the belief is justified without reference to any kind of experience or inclination whatsoever, whether sensual or purely intellectual' (p. 115). On Koons' estimation strongly *a priori* justified beliefs are intrinsically immune to undercutting defeat (pp. 116, 122).

highly probable that his* belief that p and the grounds for his belief that p are caused in such a way that S's belief that p depends for its justification on the justification for his* belief that p and the grounds for his belief that p are caused. (From Propositions 1-6.)

Proposition 7: Let R be a relation whose range is the set of propositions belief in which S is in a position to be internally justified in having. Let R hold between the propositions that p and that q, just in case S is potentially in a position to be in a noetic state in which S's belief that p depends for its internal justification on S's belief that q. Then R is a partial well-ordering: well-founded, transitive and irreflexive.

Lemma 2: If S's belief that p is an ordinary empirical belief, then S knows that p *only if* there is a noetic state n and a proposition that q of such a kind that (i) S is in a position to be in n, (ii) in state n, S's belief that p depends for its internal justification on S's belief that it is highly likely that his* belief that q is caused, and (iii) in state n, S's belief that it is highly likely that his* belief that q is caused does not depend for its internal justification on any ordinary empirical beliefs of S's. (From Lemma 1 and Proposition 7.)

Lemma 3: If S's belief that p is an ordinary empirical belief then S know that p *only if* there is a proposition that q such that S is potentially in a position to be strongly *a priori* justified in believing that his* ordinary empirical belief that q is caused and in believing that the epistemic grounds of his* belief that q are very likely caused.

Proposition 9: Necessarily, if S's belief that p is an ordinary empirical belief, then S is potentially in a position to be strongly *a priori* justified in believing that it is highly likely that his* belief that p and the grounds of his* belief that p are caused *only if* S is in a position to be strongly *a priori* justified in believing that it is highly likely that *any* of the situations in his empirical knowledge-net is caused.

Lemma 4: Necessarily, if S's belief that p is an ordinary empirical belief, then S knows that p only if S is in a position to be strongly *a priori* justified in believing that it is highly likely that *any* of the situations in his empirical knowledge-net is caused. (From Lemma 3 and Proposition 9.)

Definition 4: γ is a *principle of general causation* iff γ takes the form: it is nomologically impossible for a situation of type T to be actual in the absence of a cause. For such a principle, T is γ's *range of application*.

Definition 5: γ is a *qualified principle of general causation* iff γ takes the following form: the objective probabilities are of such a kind that, for every possible situation of type T, the probability of s's occurring uncaused is vanishingly low—i.e., so low that, no matter how unlikely the caused occurrence of s might be according to a possible noetic state, its uncaused occurrence is much more unlikely.[6]

Proposition 10: S is potentially in a position to be strongly *a priori* justified in believing that it is highly likely that any of the situations in his* empirical knowledge net is caused *only if* S is strongly *a priori* justified in believing that it is very likely that there is some type T such that (i) some principle—or qualified principle—of general causation γ holds with T as its range of application, and (ii) it is self-evident to S that nearly all of the situations in his* empirical knowledge net fall within T.

Proposition 11: S is potentially in a position to be strongly *a priori* justified in believing that it is highly likely that any of the situations in his* empirical knowledge-net is caused *only if* there is some type T such that S is strongly *a priori* justified in believing that (i) it is very likely that it is nomologically impossible for situations of type T to be actual in the absence of a cause, and (ii) it is self-evident to S that nearly all of the situations in his* empirical knowledge net fall within T. (From Proposition 10.)

Theorem 2: If S's belief that p is an ordinary empirical belief, then S knows that p only if there is some type T such that S is strongly *a priori* justified in believing that (i) it is very likely that it is nomologically impossible for situations of type T to be actual in the absence of a cause, and (ii) it is self-evident to S that nearly all of the situations in his* empirical knowledge net fall within T. (From Lemma 4 and Proposition 11.)

II.

Clearly, there are many questions that could be raised about the argument for Theorem 2. Koons himself acknowledges that 'there is a great deal more work to be done on the nature of immunity to defeat, on the nature of the related dependency relations between propositions, and

[6] The 'objective probabilities' to which Koons refers here are 'subjective in nature, but correspond to the probability judgments of an ideal rational agent' (p. 115).

on the possible scope of strongly *a priori* beliefs' (131)—and others will surely want to raise objections against the basic epistemological assumptions that help to drive the argument. However, I propose to focus on just one difficulty that I see in the argument for Theorem 2, namely the acceptability of Proposition 9.

Shorn of irrelevant complications, Proposition 9 tells us that, in order to be justified in believing that one's belief that p and the grounds for one's belief that p are caused, one needs to be justified in believing that it is highly likely that *any* of the situations in one's knowledge net is caused. However, it is clear that this is not so—and, indeed, as we shall see, Koons himself tacitly concedes that this is not so.

Recall that, according to Koons, there are three kinds of elements in a knowledge-net: belief states, objects of beliefs, and events that mediate between belief states and their objects. In any particular case of empirical belief, given the other assumptions that Koons makes, it is clear that we can concede that, in order to be justified in believing that one's belief that p and the grounds for one's belief that p are caused, one needs to be justified in believing that the state of one's believing that p and the events that mediate between the state of one's believing that p and the object of one's belief that p are caused—but, at least for all that we have been given so far, we have no reason at all to concede that we so much as need to believe that the object of one's belief that p is caused.

When Koons himself introduces the relevant considerations, he writes as follows:

> In the case of empirical knowledge, one must be justified in believing that there is a high objective probability that any of the situations making up one's knowledge-net—one's belief states that constitute ordinary empirical knowledge and any of the epistemically mediating events (events belonging to the causal chain that connects those belief states with their objects, or to the common cause of the states and their objects)—are caused in an epistemically appropriate normal way. (115)

The omission here of any consideration of the objects of states of knowledge is telling: for it is clear that there is nothing in reliabilism, or proper function theory, or any other contemporary epistemological theory of the kind that Koons allows in play that requires that, in order for some-

thing to be an object of knowledge for a given subject, it must be very likely that that *object* of knowledge has a cause. Suppose that the fact that p is uncaused. So long as my belief that p is caused, and there is an appropriate chain of epistemically mediating events that connects my belief that p to the fact that p, there is nothing in the kinds of contemporary epistemological theories that Koons favours that rules that it is simply out of the question that I know that p.

Consider the case of an ideal agent who has total empirical knowledge. Following Koons, we might suppose that such an agent will have a knowledge-net that embraces the whole of the Cosmos (119). However, even if we suppose that the Cosmos is a causal plenum, it will still be the case that we are free to suppose that the Cosmos has a boundary, and—for all that has been argued thus far—that the boundary of the Cosmos has no cause. Moreover, and more importantly, there is nothing in the kinds of conditions on empirical knowledge to which Koons is appealing that requires us to suppose that the boundary of the Cosmos has a cause: for, plainly, the elements on the boundary of the Cosmos can only figure in our knowledge-nets as objects of knowledge. (The elements on the boundary of the Cosmos cannot be belief states that are constituents of empirical knowledge states, since—on the assumptions now in play—there cannot be empirical knowledge states without causally anterior objects; and the elements on the boundary of the Cosmos cannot be mediating states for empirical knowledge, since they are not causally posterior to anything in the Cosmos.)

If these considerations are on the right track, then the most the Koons can derive from the various assumptions that he makes is not Theorem 2, but rather something like Theorem 2*:

> **Theorem 2***: If S's belief that p is an ordinary empirical belief, then S knows that p only if there is some type T such that S is strongly *a priori* justified in believing that (i) it is very likely that it is nomologically impossible for situations of type T to be actual in the absence of a cause, and (ii) it is self-evident to S that nearly all of the situations in his* empirical knowledge-net *that do not lie on the boundary of his* empirical knowledge-net* fall within T.

III.

Koons (1997) defends the following cosmological argument (here I follow the presentation in Koons (2008)):

Axiom 1: x is a part of y iff everything that overlaps x overlaps y.
Axiom 2: If there are any φ's, then there exists a sum of all of the φ's; for any x, x overlaps this sum iff x overlaps one of the φ's.
Axiom 3: x=y iff x is a part of y and y is a part of x.
Axiom 4: Situations necessitate the actual existence of their parts.
Axiom 5: The actual existence of all of the members of a sum necessitates the actual existence of the sum.
Axiom 6: Causation is a binary relation between actually existing situations.
Axiom 7: Causes and effects do not overlap (i.e. have no parts in common).
Axiom 8: For any given wholly contingent situation x, there is a (defeasible) presumption that x has a cause.

Theorem: If there are any contingently existing situations, then there is a necessarily existing situation that is the cause of the Cosmos, i.e. of the sum of all wholly contingent situations.

For present purposes, the key feature of this argument is Axiom 8. Shorn of the considerations about defeasibility—i.e. with the deletion of the words 'there is a (defeasible) presumption that'—this axiom would be the following assumption about causality:

Cause 0: All wholly contingent situations have causes.

As Koons (2008) notes, there are many alternative causal principles that might be proposed which are such that, when we appropriately add back in the words 'there is a defeasible presumption that', and insert the amended principle into the above argument in place of Axiom 8, will not permit the derivation of the Theorem. In particular, Koons mentions all of the following principles, which he claims to be unacceptable as general principles of causation:

Cause 1: All non-first situations have causes.
Cause 2: All situations with finite temporal duration have causes.

Cause 3: All situations that don't occur at a first moment of time have causes.
Cause 4: All situations that don't include temporal regresses have causes.
Cause 5: All situations that aren't both extremely simple and cosmic in scale have causes.
Cause 6: All situations that could (*de re*) be caused have causes.
Cause 7: All situations that could possibly have wholly contingent causes have causes.

Moreover, Koons also mentions the following principles, which he claims to be acceptable as general principles of causation, and which he claims are such that, when we appropriately add back in the words 'there is a defeasible presumption that', will permit the derivation of the Theorem if inserted into the above argument in place of Axiom 8:

Cause 8: All situations that are composed of parts, for each of which it is metaphysically possible that there exists a situation that approximately duplicates it and has a cause, have causes
Cause 9: All situations that are natural—i.e., occurring in space and time, or involving finite powers and dispositions—or that involve acts or states of consciousness that are finitary in content, or that involve acts or states of consciousness that are composed of parts that are finitary in content, have causes.
Cause 10: All situations that do not involve metaphysically simple acts of consciousness with infinitely rich content, have causes.

Before he tries to use his Theorem 2 to separate the sheep from the goats, Koons gives a list of what he takes to be criteria of acceptability of principles of general causation. These criteria are as follows:

Criterion 1: If one is *strongly a priori* justified in believing that one's knowledge-net falls within the range of application of a principle of general causation, then it is metaphysically necessary that any knowledge-net of any person with humanoid consciousness falls within the same range.
Criterion 2: It must be plausible to suppose that it is self-evident that the range of application of a principle of general causation encompasses our knowledge-nets.
Criterion 3: An epistemologically acceptable principle of general causation must be one whose range of application specifies a set of intrinsic properties of situations.

> **Criterion 4**: If belief in a principle of general causation is to be strongly *a priori* justified, then the boundaries of the range of application of that principle must be non-arbitrary and metrically isolated.
> **Criterion 5**: The range of application of a principle of general causation must be closed under proper parthood.
> **Criterion 6**: An epistemically acceptable principle of general causation must be sensitive to the fact that human cognition includes an open-ended, highly general form of abductive inference.

Given these criteria, and given his Theorem 2, Koons undertakes to show that none of Cause 1 through Cause 7 is an acceptable principle of general causation, while each of Cause 8 through Cause 10—and, I guess, also Cause 0—is an acceptable principle of general causation.

IV.

Here is Koons' argument against Cause 1:

> How could I be SAP justified in believing that my current belief-state is a non-first situation? In order to do so, I would have to know that there were situations that preceded my current belief-state in time, but my knowledge of the past consists entirely in ordinary empirical beliefs (including memory and testimony), all of which presuppose (as I have argued) belief in the applicability of the causal principle to my current belief state. Hence we cannot have a non-circular justification of immunity to defeat relying on [Cause 1]. (124/5)

What Koons has in mind here is, I think, this. In order to satisfy the demands of his Theorem 2, it has to be the case that one is strongly *a priori* justified in believing that it is self-evident that nearly all of the situations in one's empirical knowledge-net are non-first situations. Hence, in particular, it has to be the case that one is strongly *a priori* justified in believing that it is self-evident that one's current belief state is a non-first situation. Or, more simply, it has to be the case that one is strongly *a priori* justified in believing that one's current belief state is a non-first situation. But if one is strongly *a priori* justified in believing that one's current belief state is a non-first situation, then one is strongly *a priori*

justified in believing that there have been situations that have preceded one's current belief state. However, on Koons' account, one is not strongly *a priori* justified in believing that there have been situations that have preceded one's current belief state: one's belief that there have been situations that have preceded one's current belief state necessarily rests exclusively upon ordinary empirical beliefs—memory, testimony, and the like—all of which presuppose that there have been situations that have preceded one's current belief state. So Cause 1 fails to meet the demands of Theorem 2.

It is clear where our criticism of this argument will begin. As we argued above, Theorem 2 is not supported by the considerations that Koons advances on its behalf. The most that we get from the considerations that Koons advances is Theorem 2*. But, in order to satisfy the demands of Theorem 2*, it only needs to be the case that one is strongly *a priori* justified in believing that it is self-evident that nearly all of the situations in one*'s knowledge-net that do not lie on the boundary of one's empirical knowledge net are non-first situations. And, of course, this is true: in fact, one is strongly *a priori* justified in believing that it is self-evident that *all* of the situations in one*'s knowledge net that do not lie on the boundary of that net are non-boundary situations. So, of course, Cause 1 does succeed in meeting the demands of Theorem 2*.

Of course, even if Cause 1 does succeed in meeting the demands of Theorem 2*, that does not entail that Cause 1 is an epistemologically acceptable principle of general causation. For all that we have argued thus far, it may be that there is some other way in which Theorem 2 can be established. Moreover, for all that we have argued thus far, there may be some other way in which we can argue directly that Cause 1 is not an epistemologically acceptable principle of general causation. All that we have argued, thus far, is that Koons has failed to establish that there is something unacceptable about Cause 1.

But we can do more. Suppose that we accept Cause 0—i.e., suppose we accept Koons' claim that all wholly contingent events have causes. Does it follow from my acceptance of Cause 0 that I am strongly *a priori* justified in rejecting the claim that my current belief state was directly caused by a fallen angel? We suppose that the God of theism exists, and that the God of theism created angels with the capacity to make inde-

pendent domains of contingently existing things. One of these angels has rebelled, and has chosen to make an independent domain of contingently existing things in which my current belief state is an initial state, i.e. a state that is not preceded by other elements in that independent domain of contingently existing things. Since Cause 0 is clearly satisfied in this scenario, it cannot be that acceptance of Cause 0 justifies the assignment of a low probability to the claim that my current belief state was directly caused by a fallen angel. But then, by Koons' lights, how could I be strongly *a priori* justified in rejecting the claim that my current belief state was directly caused by a fallen angel?

Perhaps Koons might try to argue that we are strongly *a priori* justified in believing that the God of theism could not make an angel with the capacity to make independent domains of contingently existing things. But that looks highly implausible. On the one hand, the God of theism is omnipotent: it can't be that God is unable to make such an angel because God lacks the power or ability to do so. On the other hand, even though the God of theism is perfectly good, theists will typically find it hard to deny that the goodness of the God of theism is consistent with the creation, by that God, of creatures that are free with respect to decisions that have very significant consequences. For these reasons, it seems that theists will typically not even be justified in believing that the God of theism could not make an angel with the capacity to make independent domains of contingently existing things, let alone strongly *a priori* justified in so believing.[7]

However, if we grant that the God of theism could make an angel with the capacity to make independent domains of contingently existing things, then how can one rule out the possibility that one's current belief state was created directly by a fallen angel? In order to establish that one's current belief state is not an initial state of a 'natural' world, one would need to show that there were situations that preceded one's current belief-state in time. But, on Koons' own reckoning, 'my knowledge of the

[7] As a referee pointed out to me, there is more to say here. Some medieval philosophical theologians argued on *a priori* grounds that no creatures could create *ex nihilo*—and those arguments might be thought to cast some doubt on my case. I think that these arguments, even if cogent, are clearly beside the point: for nothing in my case requires my angel to create *ex nihilo*. Surely God could give my angel raw materials from which to construct physical universes containing contingently existing creatures!

past consists entirely in ordinary empirical beliefs (including memory and testimony), all of which presuppose … belief in the applicability of the causal principle to my current belief state'. So, on Koons' own reckoning, it seems, we cannot have a non-circular justification of immunity to defeat relying on Cause 0.

If the argument of the last two paragraphs is granted, then it seems that we can conclude that the kind of argument that Koons' makes against Cause 1 can be repackaged to knock out almost any principle of general causation. In particular, it is obvious that the same kind of consideration will extend to Cause 9 and Cause 10: if the argument that Koons makes against Cause 1 is good, then the same kind of argument that we have made against Cause 0 will knock out those causal principles as well.

V.

I think that it is natural to suppose that, given the framework that he develops earlier in the paper, Koons' argument against Cause 1 goes wrong in supposing that 'knowledge of the past consists *entirely* in ordinary empirical beliefs—including memory and testimony—all of which presuppose belief in the applicability of the causal principle to current belief states'. What is required to defeat the hypothesis *that* my current belief state is a first situation—i.e. that my current belief state is a situation that does not have a natural or empirical past—is justified belief *that* there have been natural or empirical situations prior to my current belief state. But it seems to me that the very reasons that Koons gives for believing that one is strongly *a priori* justified in believing that nearly all of the situations in one's empirical knowledge-net have causes extend to reasons for believing that one is strongly justified in believing that there have been natural or empirical situations prior to one's current belief state.

Before I can explain why I suppose that this is so, I need to make some preliminary observations about knowledge-nets and belief-nets. As we noted earlier, Koons takes knowledge-nets to have three kinds of constituents: for a given subject S, the knowledge-net for S consists of (i) all of the belief states that p for which it is true that S knows that p, (ii) all of the situations that p for which it is true that S knows that p, and (iii) all of the situations that causally mediate between the situation that p and

S's belief that p in cases in which S knows that p[8]. Thus, a knowledge-net for a given subject has two rather different parts: on the one hand, the knowledge-net contains all of the belief states of that subject that are states of knowledge; on the other hand, the knowledge-net contains a (most likely proper) part of the Cosmos that consists of (a) all the parts of the Cosmos that are objects of knowledge of the subject, and (b) all of the parts of the Cosmos that belong to appropriate kinds of chains that causally connect those objects of knowledge to the subject.

From a third-person perspective on a subject S, it is in principle easy to distinguish between those of S's beliefs that constitute knowledge (and hence that belong to S's knowledge-net), and those of S's beliefs that do not constitute knowledge (and hence that do not belong to S's knowledge-net). However, from a first-person perspective on oneself, it is more difficult to make out this distinction. True enough, if we suppose that only beliefs that are held with sufficient firmness can be knowledge, then one can identify one class of one's own beliefs that are not candidates to be knowledge. But sensible intellectual modesty requires one to hold that, even amongst the beliefs that one holds with sufficient firmness to make them candidates for knowledge, one has some false beliefs. Yet, for any *particular* belief that one holds with sufficient firmness, it seems that one can only accept that that belief does not amount to knowledge by either reducing the firmness with which one holds that belief, or by giving it up altogether. (If I suppose that it is in doubt whether it is true that p, then I simply don't believe that p; and if I suppose that it is in doubt whether I am warranted in believing that p, then I simply don't believe that p with the kind of firmness that is required for knowledge.)[9]

Consider the collection of my *beliefs* that I hold with the kind of firmness that is required for knowledge—my *belief-net*.[10] By Koons' lights,

[8] To simplify our discussion, we ignore the possibility—for which Koons makes explicit allowance—of cases in which S knows that p in which (i) there is a common cause of the situation that p and S's belief that p, and (ii) there are situations that mediate between that common cause and S's belief that p (even though there is no chain of situations that mediates between the situation that p and S's belief that p in such a way as to guarantee that S knows that p).

[9] Throughout this discussion, I assume that beliefs that are held irrationally are not—and perhaps cannot be—candidates for knowledge.

[10] Note that it is consistent with this consideration to suppose that *any* degree of firmness can suffice for knowledge. There is no commitment in my discussion to the claim

it seems that he should be prepared to say that I have a strong *a priori* entitlement to the claims (i) that most of the beliefs in my belief-net have situations in the Cosmos as their objects, and (ii) that, where beliefs in my belief-net have situations in the Cosmos as their objects, there are situations in the Cosmos that causally mediate between my beliefs and their objects. Moreover, by Koons' lights, it seems that he should be prepared to say that, for any *particular* belief in my belief-net, I have strong *a priori* entitlement to the claims (i) that it is very likely that that belief has a situation in the Cosmos as its object, and (ii) that it is very likely that there are situations in the Cosmos that causally mediate between that belief and the corresponding situation.

If it is true that one has strong *a priori* entitlement to the claims (i) that it is very likely that any particular belief in one's belief-net has a situation in the Cosmos as its object, and (ii) that it is very likely that there are situations in the Cosmos that causally mediate between any particular belief in one's belief-net and the situation that is the object of that belief, *then* it seems plausible to suppose that it is also true that one has strong *a priori* entitlement to the claim that the 'interior' of the Cosmos is (very close to) a causal plenum. (How could it be true, for any particular belief in one's belief net, that it is very likely that there are situations in the Cosmos that causally mediate between that belief and the situation that is the object of that belief if there are many situations in the 'interior' of the Cosmos that fall within the scope of one's belief-net that do not have causes? How could it be true, for any particular belief in the belief net of any rational agent, that that rational agent is strongly *a priori* entitled to the belief that it is very likely that there are situations in the Cosmos that causally mediate between the particular belief held by that agent and the situation that is the object of that belief if there are many situations in the 'interior' of the Cosmos that fall within the scope of one's belief-net that do not have causes?)

Suppose that I believe that I had Weetbix and vegemite on toast for breakfast last Friday, and that I believe this with the kind of firmness that is necessary for knowledge. In particular, suppose that I take myself to remember having had Weetbix and vegemite on toast for breakfast last

---

that only beliefs held with sufficient firmness can be knowledge; rather, the commitment is to refusing to rule out that claim.

Friday: I vividly recall getting the cereal packet and bread from the cupboard and the milk from the fridge, and so on. Then, by the principles introduced in the preceding two paragraphs, I shall be entitled to the belief that it is very likely that I did have Weetbix and vegemite on toast for breakfast last Friday, and to the belief that it is very likely that there is a causal chain leading from my having Weetbix and vegemite on toast for breakfast last Friday to my currently believing that I had Weetbix and vegemite on toast for breakfast last Friday. But, if I'm entitled to believe that I had Weetbix and vegemite on toast for breakfast last Friday, then I'm entitled to reject the suggestion that history begins with the present moment, i.e. I'm entitled to believe that my current belief-state is a non-first situation.

Can Koons object that this justification of the belief that one's current belief state is a non-first situation is circular: is he entitled to say—as, at least *inter alia*, he does—that one can only justifiably believe that one had Weetbix and vegemite on toast for breakfast last Friday if one is strongly *a priori* justified in holding the belief that one's current belief-state is a non-first situation? I don't think so. Imagine the first moment of the existence of a fully rational agent that has not yet had any experiences, and that is not endowed with any misleading apparent memories of previous experiences, but which has the capacity to perceive the world, and to form memories of previous experiences. At the first moment of its existence, this creature will have no view about the extent of the past: it has had no experiences, and it has no apparent memories of earlier experiences, so it has no data that it could use in the framing of any such view. Suppose, however, that, in its first moment of existence, this creature begins to perceive its environment: it receives initial sensory impressions, etc. Suppose, further, that, having received initial sensory impressions in its first moment of existence, the creature begins to process these impressions, and to store the results of that processing in memory. At *all* subsequent moments of its existence, this creature will be justified in believing that those moments are not the first moment of its existence provided only that it is justified in relying on the results of the memory-processing that is activated by its initial experiences. While we may grant to Koons that *part* of the justification for relying on the results of the memory-processing that is activated by initial experience lies in one's strong *a priori* entitlement to the claims (i) that it is very likely that any particular belief in

one's belief-net has a situation in the Cosmos as its object, and (ii) that it is very likely that there are situations in the Cosmos that causally mediate between any particular belief in one's belief-net and the situation that is the object of that belief, we should also insist that *another* part of the justification for relying on the results of the memory-processing that is activated by initial experience lies in one's strong *a priori* entitlement, in the absence of *rebutting* defeaters, to treat one's apparent memories of one's earlier experiential states as veridical.

If Koons doesn't accept some version of the claim that one has strong *a priori* entitlement, in the absence of *rebutting* defeaters, to treat one's apparent memories and one's apparent perceptions of external objects as veridical, then it is hard to see how he can avoid the sceptical conclusion that rational agents have no way of neutralising sceptical hypotheses in which there is deviant causation of apparent memories and apparent perceptions of an external world. Strong *a priori* entitlement to a causal principle may suffice to defeat sceptical hypotheses about the *absence* of causes of belief states or the *absence* of appropriate causal connections between belief states and parts of the external world, but that kind of entitlement alone plainly won't suffice to defeat sceptical hypotheses about *deviant* causes of belief states or *deviant* causal connections between belief states and parts of the external world. Similarly, while strong *a priori* entitlement to a causal principle may suffice to defeat sceptical hypotheses about the absence of causes of belief states or the absence of appropriate causal connections between belief states and parts of the external world, that kind of entitlement alone plainly won't suffice to defeat sceptical hypotheses about *mismatches* between the contents of belief states held with sufficient firmness to be candidates for knowledge and the external world. In short: it is just a mistake to suppose that appeal to a principle of general causation can suffice to defeat all sceptical hypotheses about the correspondence of the Cosmos to one's memories and perceptions.

Suppose that Koons grants that, when one is confronted with a sceptical hypothesis that entails that one's apparent memories are all *merely* apparent, considerations about the presence of causes for those apparent memories will not suffice to enable one to reject that sceptical hypothesis. Since it is clear that the hypothesis that one's current belief-state is a first situation is a sceptical hypothesis that entails that one's apparent

memories are all merely apparent, Koons will then have to grant that the hypothesis that one's current belief state is a first situation cannot be defeated by appeal to a principle of general causation. But, if that's right, then it is no objection to Cause 1 to observe that one cannot defeat the hypothesis that one's current belief state is a first situation by appeal to Cause 1.

VI.

If the argument of the preceding section is correct, then it establishes that Koons has not managed to show that we could not be 'strongly *a priori* justified in believing that nearly all of our knowledge-net falls within the range of application' of Cause 1. Moreover, if the argument of the preceding section is correct, then it can be readily adapted to establish that Koons has not managed to show that we could not be 'strongly *a priori* justified in believing that nearly all of our knowledge-net falls within the range of application' of others amongst the principles of general causation that he deems unacceptable. Thus, I think, if the argument of the preceding section is correct, we can see that Koons' attempt to provide epistemological foundations for the cosmological argument fails on its own terms: even if we grant the many assumptions that Koons needs in order to argue for his Theorem 2, we still have good reason to insist that it is entirely understandable that his favoured principle of general causation 'has failed to win universal acceptance' (p. 105).

Of course, there is much else in Koons' paper that merits comment. For instance, I have said nothing here concerning his criteria for acceptability of proposed principles of general causation. Any such comment will need to wait for some other occasion.[11]

[11] I am grateful to Brian Leftow for very helpful and detailed comments on the initial draft of this paper. The original version of this paper served as the basis for my presentation at the Formal Epistemology Conference in Leuven in June, 2009. I am indebted to Jake Chandler and Victoria Harrison for the invitation to that outstanding event, and to all of the conference participants for conversations about this and other papers.

BIBLIOGRAPHY

Koons, R. (1997). 'A New Look at the Cosmological Argument', *American Philosophical Quarterly* 34, 193-211.

Koons, R. (2000). *Realism Regained: An Exact Theory of Causation, Teleology and the Mind.* New York: Oxford University Press.

Koons, R. (2001). 'Defeasible Reasoning, Special Pleading, and the Cosmological Argument: A Reply to Oppy', *Faith and Philosophy* 18, 192-203.

Koons, R. (2008). 'Epistemological Foundations for the Cosmological Argument', *Oxford Studies in Philosophy of Religion* 1, 105-33.

Oppy, G. (1999). 'Koons' Cosmological Argument', *Faith and Philosophy* 16, 378-89.

Oppy, G. (2004). 'Faulty Reasoning about Default Principles in Cosmological Arguments', *Faith and Philosophy* 21, 242-9.

Oppy, G. (2006). *Arguing about Gods.* Cambridge: Cambridge University Press.

Oppy, G. (forthcoming). 'Uncaused Existence', *Faith and Philosophy*.

COSMOLOGICAL ARGUMENT: A PRAGMATIC DEFENSE

EVAN SANDSMARK & JASON L. MEGILL

University of Colorado ◆ Old Dominion University

Abstract. We formulate a sort of "generic" Cosmological argument, i.e., a Cosmological argument that shares premises (e.g., "contingent, concretely existing entities have a cause") with numerous versions of the argument. We then defend each of the premises by offering pragmatic arguments for them. We show that an endorsement of each premise will lead to an increase in expected utility; so in the absence of strong evidence that the premises are false, it is rational to endorse them. Therefore, it is rational to endorse the Cosmological argument, and so rational to endorse theism. We then consider possible objections.

INTRODUCTION

Cosmological arguments claim that a special being of some sort (e.g., a first cause, a necessary cause, an unmoved mover, etc.) must exist to explain or account for the existence of the universe; it is then claimed that this special being is God. Such arguments have a venerable history. The first cosmological argument occurs in Plato's *Laws*.[1] Aristotle, along with various Islamic theologians, sharpened the argument before it was discussed again by Aquinas.[2] The argument also received much attention during the early modern period; Leibniz and Clarke both defended it

[1] See Plato, *The Laws of Plato*, translated by T.L. Pangle (New York: Basic Books, 1980), Book X.

[2] See Aristotle, *The Complete Works of Aristotle*, volumes I and II, edited by J. Barnes (Princeton: Princeton University Press, 4th C BCE/1984), especially *Physics* VIII, 4-6 and *Metaphysics* XII, 1-6. See also Aquinas, *Summa Theologica*, translated by the Fathers of the English Dominican Republic (New York, NY: Benzinger Bros, 13th C/1948), especially the first part, question 2, article 3.

while Hume and Kant famously criticized it.[3] The argument has contemporary defenders too, including Craig, Gale and Pruss, and Koons.[4] We also defend the cosmological argument. In the following section, we discuss the version of the argument that we defend. As we show, there are four premises that need justification. Then, rather than focus on metaphysical or scientific considerations in favor of the premises, we attempt a different strategy. We offer a series of pragmatic arguments that show that it is at least rational to endorse the premises; if so, it is rational to endorse theism. Finally, we consider possible objections.

THE ARGUMENT

The version of the argument that we defend consists of seven steps, and can be thought of as a sort of generic Cosmological argument. The first premise is,

(1) If *x* is a contingent, concretely existing entity, *x* has a cause.

The second premise claims,

(2) Our universe is a contingent, concretely existing entity.

---

[3] For the history of the argument, see W. L. Craig, *The Cosmological Argument from Plato to Leibniz* (London: The Macmillan Press, 1980). For Leibniz on the Cosmological argument, see *The Monadology*, translated by George MacDonald Ross, 1714/1999. See also S. Clarke, *A Demonstration of the Being and Attributes of God and Other Writings*, edited by E. Vailati (Cambridge: Cambridge University Press, 1705/1998). For Hume's criticism, see *Dialogues Concerning Natural Religion* (Indianapolis: Hackett, 1779/1980), part IX. See also I. Kant, *Critique of Pure Reason* (Cambridge: Cambridge University Press, 1980/1998), especially the Transcendental Dialectic, Book II, Chapter 3, section 5.

[4] See W. L. Craig, *The Kalām Cosmological Argument* (London: The Macmillan Press, 1979) and "In Defense of the *Kalām* Cosmological Argument," *Faith and Philosophy* 14.2 (1997): 236–247. Also see R. Gale and A. R. Pruss, "A New Cosmological Argument," *Religious Studies* 35 (1999): 461–476 and R. Koons, "A New Look at the Cosmological Argument," *American Philosophical Quarterly* 34 (1997): 171-192. Of course, the argument has many contemporary opponents as well; see, e.g., G. Oppy, *Arguing about Gods* (Cambridge: Cambridge University Press, 2006), chapter 3.

By "universe," we have something like the commonsense usage in mind, i.e., the universe is the sum total of all concretely existing matter and energy.[5] The next step follows from the first two with basic logic,

(3) Our universe has a cause.

Step (4) is a disjunction that has the form of excluded middle,

(4) Either the cause of our universe is itself a part of the universe or it is not a part of our universe, i.e., it is outside the universe.

Step (5) claims,

(5) The cause of our universe is not a part of our universe.

We interpret this claim in the mereological sense; the universe can be thought of as a whole, and the cause of the universe is not a part of that whole. Step (6) follows from (4) and (5) with disjunctive syllogism,

(6) The cause of our universe must be outside the universe.

Finally,

(7) This cause of our universe is "God."

The steps that need justification are (1), (2), (5) and (7). Steps (3) and (6) follow from other steps and (4) is a logical truth. We defend (1), (2), (5) and (7) in successive sections.

PREMISE ONE

We now offer a pragmatic argument for premise one. The argument attempts to show that even assuming that one is not certain of the truth of the premise, it is *still* rational to endorse it. As we show, accepting the first premise has more expected utility than rejecting it, so assuming

[5] We do not take "universe" to mean "everything that exists," as some do. Below, we argue that there is at least one thing that exists that is not a part of the universe.

that it is rational to increase expected utility, it is rational to endorse the premise, at least in the absence of strong evidence that it is false.

So, suppose that it is rational to increase expected utility. Also suppose that it is generally better to have knowledge than not have it; that is, knowledge is desirable and an increase in knowledge is also an increase in utility, at least most of time.[6] Given these modest assumptions, it is rational to endorse any claim that leads to an increase in knowledge and so in expected utility, at least if we do not have sufficient evidence to reject the claim. Given that (1) has lead to a vast increase in knowledge and utility insofar as it is a vital methodological assumption of science, for example, it is rational to endorse (1).[7] To rephrase the argument: (1) is either true or false. And we have two options: (a) we can believe that (1) is true or (b) we can reject it. Suppose that we opt for (a). The expected utility for this option is quite high: the assumption that (1) is true has led to the discovery of countless scientific truths as well as a number of more banal though still useful truths (e.g., "eating causes my hunger to go away"), and this is a very positive outcome. Indeed, (1) is a principle that makes reality itself intelligible.[8] Further, the utility will be high whether (1) is in fact true or not (if we endorse (1) but it is false, we seem to lose very little, at least when weighed against the gains of endorsing (1)). Now, suppose that we opt for (b): the assumption that (1) is false would not lead to a positive outcome; indeed, it would lead to disastrous consequences. If we reject (1), reality would become unintelligible. If we did not think in terms of cause and effect, then we might simply stop eating, we wouldn't bother looking for cures for diseases, and so on. In short, the expected utility for rejecting (1) is extremely low; we might become extinct. If it is rational to increase expected utility, then it is rational to endorse (1).[9]

---

[6] We say "most of the time" because there might be cases in which it is better, at least from an expected utility standpoint, to lack knowledge than have it. Quite simply, some truths might make us unhappy, so knowing them might decrease our happiness. Even so, in general, it is better to have knowledge than not have it; who will deny this?

[7] Science seeks to discover the causal relationships between entities; this is the goal of science. But if (1) is false, this goal is undermined; the search for causal relationships presupposes that there are causal relationships to be found.

[8] See R. Taylor, *Metaphysics* (Englewood Cliffs: Prentice-Hall, 1992).

[9] To be clear, the claims that this argument rests on are extremely weak and plausible: (i) the expected utility for reality being intelligible is greater than the expected utility for

PREMISE TWO

Premise (2) claims that "Our universe is a contingent, concretely existing entity." Our universe does not exist necessarily; it might not have existed. This claim is *prima facie* plausible; many involved, including some opponents of theism, will grant that it is true. Nevertheless, the claim requires some justification. First, note that if (2) is false, then *our* universe exists necessarily, and so everything in our universe necessarily exists exactly as it does etc. Consider the following argument: our universe is composed of a number of entities that stand in various relationships to one another; call our universe "u." Now consider a universe that is different from u in some random respect, no matter how minor (e.g., in this universe, Jupiter has an additional moon, but the difference can be anything); call this universe "$u^{\star}$." By hypothesis, universe u is not numerically identical to $u^{\star}$; this obviously follows by Leibniz's law, for instance. Furthermore, $u^{\star}$ is either possible or is not possible. If $u^{\star}$ is not possible, then everything in our universe is necessary; if nothing, no matter how insignificant, could have been different, then the way things are in our universe is the way they must be. But if (2) is false, then $u^{\star}$ is not possible (because our universe is necessary, and it is not identical to $u^{\star}$). But then if (2) is false, everything about our universe is necessary. The denial of (2) leads to an implausible superessentialism that few would find acceptable (an exception would be Spinoza).[10]

So, if (2) is false, everything about our universe is necessary, and anything that happens necessarily happens, therefore the future is "closed," i.e., it is already set in stone. This would be undesirable for a number of reasons. A gloomy fatalism threatens. It's difficult to see why anyone should be praised for an accomplishment that was fated to occur. It's difficult to see how there could be moral responsibility in such a world, in which every act was already predetermined. It's also difficult to see

---

reality being unintelligible and (ii) it is rational to increase expected utility. These are the only claims this argument needs.

[10] See B. Spinoza, *The Collected Works of Spinoza Vol. I*, edited and translated by Edwin Curley (Princeton: Princeton University Press, 1677/1985), Book I. The denial of (2) is so implausible many will be content to accept (2) already, even without a pragmatic argument for it.

how there could be free will.[11] Intuitively, if nothing could have possibly been different than the way it is, then any "choice" that you make could not have turned out differently, so it appears that the "choice" was not really a choice at all, and so was not really free. If there is only one possible way that the universe might be, then the outcomes of all "decisions" are predestined, and so these "decisions" are not really decisions after all etc. Now, either (2) is true or false, and we can either endorse (2) or reject it. Suppose that we reject (2) and (2) happens to be true. In this scenario, we effectively deny that the future is open, even though it is. This is clearly a negative outcome, at least given the plausible assumption that we would be happier if we believe that the future is open and thereby avoid a gloomy fatalism. Suppose that we reject (2) and (2) happens to be false. In this scenario, we are correct in asserting that the future is fixed. Yet this still appears to be a negative outcome because even if the future is fixed, we would be happier believing that it is not. In short, both outcomes for a rejection of (2) have negative expected utility. Now suppose that we assert (2) and (2) is true. This is a positive outcome: the future is not fixed and we assert it. But even if we assert (2) and (2) is false, this is still a positive outcome because again, it is "better" (at least in terms of expected utility) to falsely believe that the future is open even if it is not. Given that both outcomes for the endorsement of (2) are positive and that both outcomes for the rejection of (2) are negative, there is more expected utility for the endorsement of (2), and if it is rational to increase expected utility, then it is rational to endorse (2).

PREMISE FIVE

Premise (5) claims that "the cause of our universe is not a part of the universe," in the sense that the universe can be thought of as a whole that does not contain its cause as a part.

[11] For more on the importance of free will, see, e.g., T. O'Conner, "Free Will," *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, edited by E. N. Zalta, URL = http://plato.stanford.edu/ archives/fall2008 /entries/freewill/, introduction: "Free will also appears to be a condition on desert for one's accomplishments (why sustained effort and creative work are praiseworthy); on the autonomy and dignity of persons; and on the value we accord to love and friendship…" See also R. Kane, *The Significance of Free Will* (New York: Oxford University Press, 1996) and R. Clarke, *Libertarian Accounts of Free Will* (Oxford: Oxford University Press, 2003).

Prima facie, it might appear difficult to formulate a pragmatic argument for (5) – the claim that the cause of the universe is outside the universe – similar to the ones given for (1) and (2). The *endorsement* of (5), at least taken in isolation, does not seem to have positive expected utility; but it does not appear to have negative expected utility either. At first glance, the *denial* of (5) doesn't appear to have a negative (or positive) expected utility either. However, given (1) and (2), (5) is true. *First*, suppose that (1) and (2) are true and therefore the universe has a cause. Also suppose that (5) is false; so, the cause of the universe is a part of the universe. But this cannot be the case, for if the cause of the universe is a part of the universe, we still lack a cause for the universe. That is, (1) and (2) entail that the universe has a cause, but this cause cannot be a part of the universe because anything that is a part of the universe demands a cause itself along with all of the other entities that compose the universe. But then (5) follows. Also note that even if there is an infinite chain of contingent causes, this is not sufficient to defeat (5), for the infinite chain, as a whole, still needs a cause. (1) and (2) jointly entail that the universe must have a cause; an infinite causal chain of contingent causes (assuming that there is one) would just be the universe; therefore, the infinite causal chain as a whole must have cause. Of course, this issue is related to Hume's famous objection that if we can explain the parts of a whole, we can explain the whole, so we need not venture outside of the universe to find an explanation for it.[12] But as Rowe claims, "When the existence of each member of a collection is explained by reference to some other member *of that very same collection* then it does not follow that the collection itself has an explanation."[13]

Second, suppose again that (1) and (2) are true, so the universe has a cause. Posit the set *S* that consists of all of the entities in the universe. If an entity *e* is the cause of *S*, then *e* is the ultimate cause of everything in *S*. But if *e* is a part of *S*, and if *e* causes everything in *S*, then *e* causes itself, which is incoherent.[14] Some theists (e.g., Descartes) have said that God is self-caused, so there is at least one example of an entity that causes itself,

[12] See again, D. Hume, *Dialogues Concerning Natural Religion* (Indianapolis: Hackett, 1779/1980), part IX.

[13] W. Rowe, *The Cosmological Argument* (Princeton: Princeton University Press, 1975), 264.

[14] We are assuming that there is no backwards causation.

but obviously the atheist cannot claim that *e* is God.[15] So, if one rejects (5), one must reject either (1) or (2), and as we just saw, this would have a negative expected utility. Given that a denial of (5) has negative expected utility, and given that the endorsement of (5) does not, then if it is rational to maximize expected utility, it is rational to endorse (5).[16]

STEP SEVEN

A serious problem for the cosmological argument is the presence of a "gap" between the sort of being that the argument establishes (e.g., a first cause, a necessary cause) and the classical theistic conception of God (i.e., a being that is a necessary first cause, but is also omnipotent, omniscient, omnibenevolent and so on). An opponent can claim, for example, that even assuming that a given cosmological argument is sound, all it establishes is that, e.g., a first cause exists, but perhaps this first cause is a being that is indifferent to us, or is evil or incompetent, or is simply the material that exploded in the big bang and so is not even sentient?[17] This problem applies to the version of the argument we defend as well. In this section, we discuss this issue; i.e., we attempt to justify the inference from (6) to (7).

The final step can also be given a pragmatic justification. We discuss two questions, the first of which is the following: is the cause of the universe a personal or impersonal being? In other words, is this cause (a) a sentient being that, e.g., at least takes an interest in us (and so might care about how we behave and might have some sort of a plan for us and so on), or (b) is this being sentient and completely indifferent to us or is

[15] See R. Descartes, *Meditations on First Philosophy* in *The Philosophical Writigs of Descartes*, *Volume 2*, translated by J. Cottingham, R. Stoothoff, and D. Murdoch (Cambridge: Cambridge University Press, 1641/1984).

[16] Even if one could somehow show that the expected utility for a denial of (5), and so a denial of either (1) or (2), is positive, one would still need to show that it is greater than the expected utility of accepting either (1) or (2).

[17] For more on the "gap problem," see A. Pruss, "Leibnizian Cosmological Arguments," *The Blackwell Companion to Natural Theology*, edited by W. L. Craig and J.P. Moreland (Oxford: Blackwell Publishing, 1999). Pointing to such gaps in theistic arguments to refute them is of course an old strategy; see, e.g., Hume's *Dialogues* again.

it simply a non-sentient hunk of matter etc?[18] So, suppose that we assert that (a) is true, i.e., the cause is a sentient being that takes an interest in us of some sort. And suppose that we are correct. In this scenario, we gain, or at least have the potential to gain, quite a bit. For example, if the cause of the universe cares about what we do, or has a plan of some sort for us, then our beliefs or actions might have some higher meaning, and plausibly it is better to know all of this and act accordingly than not (and this is to say nothing of the possible eschatological implications of believing in and following the will of a perfectly good God; i.e., our ultimate destiny – often but not necessarily conceived of as an eternal residence in heaven or hell – might be affected by our belief in God and our acting in accordance with God's will). Indeed, given the plausible assumption that we would be happier if we think our life has a higher meaning than not, an endorsement of (a) has positive expected utility. If we endorse (a) but (b) is actually true, then we will be living a lie, but plausibly, we will still be better off, from an expected utility standpoint, thinking that our beliefs and actions have a higher meaning even though they do not. Whether (a) is true or not, we gain positive expected utility from believing that it is. Now suppose that we assert (b): the cause of the universe does not take an interest in us, either because it simply doesn't care about us or because it is not even sentient. If we assert (b) and (b) is false, then this seems to have a very negative utility. For example, if our lives have some higher meaning, yet we are not aware of this and do not act accordingly, it seems we lose a great deal (again, if a traditional eschatological policy is embraced, we could lose everything). But if we assert (b) and (b) happens to be true, then we still lose out, because again, it is better to believe that one's life has a higher meaning than not, even if the belief is false. But the main point is that one risks so much – too much – in assuming that the first cause is impersonal because the first cause *might* be personal. A belief in option (a) has a higher expected utility than a belief in option (b), so assuming that it is rational to maximize expected utility, we should endorse (a). That is, in the absence of strong evidence that says otherwise, we should believe that the cause of the universe is

[18] The disjuncts in (b) amount to the same thing from a practical standpoint, in our opinion, so we treat them together. If the cause of the universe is sentient yet wholly indifferent to us, then so far as we are concerned, the cause might as well be a non-sentient entity.

a personal being that at least takes an interest in us because it is in our best interest to do so.[19]

At this point, it is clear that at least some of our pragmatic arguments have a strong affinity with another pragmatic argument in philosophy of religion, Pascal's Wager.[20] Of course, Pascal argues that we should believe in God, even in the absence of a proof of God's existence, because doing so is in our best interest. We have little or nothing to gain by not believing in God, but potentially might lose everything. On the other hand, by believing, we might gain everything. The argument above is very similar. We have little or nothing to gain by denying that the cause of the universe is a personal being. On the other hand, by believing, we might gain everything.

The second question we address is: given that the cause of the universe is sentient and takes an interest in us, is this being all good or all evil?[21] Again, we have two options: (a) we can assert that the being is good, and so is just, rational, non-deceitful, full of love, or has whatever properties one might consider to be good, or (b) we can assert that it is evil, and so is unjust, irrational or at least non-rational, deceitful, hateful or has whatever properties one might consider to be evil. Suppose we opt incorrectly for (b), i.e., we assert that the being is evil but it is in fact good. This is clearly a negative outcome. For suppose, as certainly seems to be the case, that we would be happier believing that the being that caused the universe – and so us – is good than believing that it is evil. If so, then we lose happiness while believing a falsehood. Furthermore, if we incorrectly opt for (b), then it will be impossible to correctly determine the being's intentions for us. Again, this being takes an interest in us, so it might want us to live a certain way, or might have a higher purpose in mind for us etc. But if we assume that the being is evil, then we could never accurately

[19] Again, this argument relies on two very weak claims: (i) it is better (in terms of expected utility) to think that one's life has a higher meaning or higher purpose than not and (ii) it is rational to increase expected utility.

[20] B. Pascal, *Pascal's Pensées*, translated by W. F. Trotter, 1910.

[21] This issue arises in a different context in W. Morriston, "The Evidential Argument from Goodness," *The Southern Journal of Philosophy* XLII (2004): 87-101, not to mention in Descartes's *Meditations*. One might think we have posited a false dichotomy; perhaps God is more good than evil on average or is more evil than good on average? Even if one adds these possibilities though, our argument still works (because our argument can still show that it is rational to believe that God is good on average etc).

discern what it is this being wants us to do if it happens to be good. Now suppose that we assert (b), i.e., we assert that the being is evil, and the being is in fact evil. This option also has negative utility, for a number of reasons. First, the thought that there is an evil being that created the universe and so us is ominous, to say the least; we would be much happier believing that the being is good, even if it isn't. Second, again, this being might have some higher plan for us, it might want us to live a certain way, and so on, but the will of an evil being could never be discerned, or at least could never be discerned with any degree of certainty, because of the likely character traits (i.e., irrationality, deceitfulness, etc.) that an evil being would have. So, we could never determine with any confidence what this being expects from us. Third, even if we could correctly ascertain this being's intentions for us, we hopefully would not wish to fulfill them anyhow; since the being is evil, its intentions for us would likely be evil as well. In short, if we assert that the being is evil, then nothing is gained and much will be lost; whether the being is evil or not, there is negative expected utility to claiming that it is evil.[22]

But if we assert (a), and so claim that the being is good, both outcomes are positive. For if we assert that the being is good and it is good, then we might be able to discern its intentions for us, and we might wish to fulfill them, and the thought that a good being caused the universe is much more pleasant than the thought that the being is evil, and so on. Likewise, if we assert that the being is good yet it is evil, this is still better than thinking that the cause of the universe is evil for various reasons given above. So, given that both outcomes for holding that the cause of

[22] Someone might object that if we assume that the being is evil and it is evil, this gives us certain advantages. For example, perhaps we could somehow get on this being's "good side" by performing immoral acts, like kicking puppies etc. But this is problematic for the reasons given above. If God is evil, God might be irrational, or at least deceitful, unjust etc. If so, there is no reliable way to determine what this God wants from creation. Thus, it is not as if a believer in this deceitful being can gain favor by, for example, performing evil acts. One might get on an evil being's good side this way, but one just as easily might not (for all we know, this evil being might prefer those who go against the wishes of their creator, and so do good etc.). Again, since the being is not even truthful, there is simply no way to figure out what it wants. It could, for example, reward everyone, or it might randomly torture some and reward others, or it might simply torture everyone. Nothing is gained, but much is potentially lost, by claiming that the cause of the universe is evil.

the universe is evil are negative, while both outcomes for holding that the cause is good are positive, and given that it is rational to increase expected utility, it is rational to assert that the cause is good.

Here is another way to look at our argument. There are four possible scenarios: (i) we assert that God is evil and God is evil, (ii) we assert that God is evil but God is good, (iii) we assert that God is good but God is evil, and (iv) we correctly assert that God is good. The expected utility for (i) is zero at best because there is no reason to assume that an evil being would prefer that people believe in it; again, this being might torture everyone indiscriminately etc. Option (ii) has very low – if any – expected utility because a good God would probably not appreciate being considered evil. It is important to note that a decision to assert that God is evil might yield *actual* utility; for example, perhaps this being is evil and rewards those who believe in it. However, there is no *expected* utility to such a decision because there is no reason to *expect* that an evil God would be "pleased with" someone believing in it, and there is certainly no reason to assume that a good God would be "pleased with" someone who believed God to be evil (we are assuming that God being "pleased with" us would increase positive utility). This is hardly a profound claim: decisions do have *unexpected* consequences. Even so, it would be absurd to assume that all decisions will have unexpected consequences and decide accordingly. If we were to do this, our decision-making faculties would be worthless. We would not, for example, wager a penny for a million dollars, even if the odds of winning the bet were 99.9%, because of concerns related to possible unexpected consequences. Option (iii), for reasons expressed above (i.e., an evil God's will is inherently undecipherable), has low expected utility as well; but option (iv) has high expected utility, so high that it makes the expected utility for claiming that God is good greater than the expected utility for claiming that God is evil. So, if we are rational, we should assert that God is good.

OBJECTIONS

In this section, we consider some possible objections.

One might object, "The pragmatic arguments can only show that it is in our best interest to endorse your premises, and so theism. But this

does not mean the premises are true." But this objection misconstrues the goal of the pragmatic arguments (and indeed, the goal of pragmatic arguments for theism in general). The pragmatic arguments do not attempt to show that the premises are true; rather, they merely attempt to show that it is *rational* to endorse them. The claim that theism is rational is still a significant claim; many atheists claim that theism is irrational.[23]

One might also object that it is rational to endorse the premises of the Cosmological argument only in the absence of strong refutations of the premises. If there are such refutations, then this overrides the pragmatic benefits we get from believing that the premises are true, so we should conclude they are false. It might be rational to endorse a claim if it increases expected utility, but if it can be shown that the claim is certainly false, or even most likely false, it would then be irrational to endorse it. But the problem with this objection is that none of the arguments against the premises of the Cosmological argument are conclusive.

For example, it has been claimed that the argument is contradictory. The first premise claims that everything must have a cause. If so, there must be a cause for the first cause, but a first cause cannot have a cause, by definition. Pruss refers to this issue as the "Taxicab Problem," after Schopenhauer's remark that in the Cosmological argument, the Principle of Sufficient Reason is similar to a taxicab that is used to get somewhere, then sent away when no longer needed.[24] Note, however, that the first premise claims that every *contingent* entity must have a cause. If one holds that the first cause is necessary, as theists will, there is no contradiction. In other words, since – by design – (1) only applies to contingent beings, one cannot apply it to the existence of a necessary being to ask what its cause is. This is a commonplace response to the Taxicab problem (it is mentioned in Pruss, for example); we bring the issue up only because the objection is so common.

For a different example, some (e.g., Hume) have claimed that premise one is conceivably, and so possibly, false.[25] We can easily conceive of

[23] On the alleged irrationality of theism, see, e.g., R. Dawkins, *The God Delusion* (Transworld Publishers, 2006).

[24] See again A. Pruss, "Leibnizian Cosmological Arguments," *The Blackwell Companion to Natural Theology*, edited by W. L. Craig and J.P. Moreland (Oxford: Blackwell Publishing, 1999).

[25] Again, Hume's most direct attack on the Cosmological argument is in the *Dialogues*, part IX.

a contingent entity simply appearing *ex nihilo*, so the first premise is not necessarily true. The objection is far from conclusive, however; in fact, it seems deeply misguided. The first premise might be true in our world yet false in some logically possible worlds; for instance, perhaps the laws of nature entail (1) in our world, but the laws of nature that hold in our world, or any laws of nature at all, might not hold in some logically possible world, so (1) is false in that world. But (1) need not be necessarily true; it merely needs to be true in our world. Perhaps the argument merely shows that it is rational to think that our universe, at least, must have a first cause?[26] This would still be a theologically interesting result. To rephrase all of this: we can imagine that a claim might be false, yet the claim might still be true in the actual world. We can imagine that George W. Bush was never president, for example; but unfortunately, this doesn't make it so. Even assuming that conceivability entails possibility (and this is of course controversial), conceivability does not entail actuality, but it would have to for the objection to succeed.[27]

For yet another example, one could object that premise (2) might be false. Our universe is certainly a concretely existing entity, but perhaps our universe is not contingent? Perhaps *something* had to exist, for instance? But this too is inconclusive. First, it is not clear that something had to exist. This point has been debated for centuries.[28] But more importantly, *even if* some entity had to exist, this does not mean that *our* universe had to exist. For example, suppose that someone says that no universe can be empty, so a given universe might contain a single grain of sand, for example. Even assuming that this is correct, this universe is not numerically identical to *our* universe (by Leibniz's law) (see above). The claim that there has to be something in any given universe does not entail that any given universe is our universe, and so does not entail that our universe had to exist, and so does not entail that (2) is false.

[26] Of course, this first cause will necessarily exist for the theist, but this is consistent with saying that the argument only establishes a first cause in our world.

[27] Spinoza (Book I of *The Ethics*) claimed that conceivability entails actuality because he collapsed the possible into the actual. Other examples are difficult to find.

[28] See B. Reichenbach, "Cosmological Argument," *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, edited by E. N. Zalta. URL = http://plato.stanford.edu/archives/ fall2008/entries/cosmological-argument/, 2008.

CONCLUSION

We argued that it is rational to accept the premises of the Cosmological argument because endorsing these premises leads to an increase in expected utility; therefore, it is rational to endorse theism. If one thinks that (i) it is rational to increase expected utility, and also (ii) that knowledge is a positive thing, (iii) that we would be happier if the future is open than if it is not, (iv) that if there is a chance that our lives might have some higher purpose, then it is better to know this than not, and (v) that we would be better off assuming that the creator of the universe is good than assuming it is evil, one should conclude that it is rational to endorse theism.[29]

BIBLIOGRAPHY

Aquinas, T. *Summa Theologica*. Translated by the Fathers of the English Dominican Province. New York, NY: Benzinger Bros, 13th C/1948.

Aristotle. *The Complete Works of Aristotle*, volumes I and II. Edited by J. Barnes. Princeton: Princeton University Press, 4th C BCE/1984.

Clarke, R. *Libertarian Accounts of Free Will*. Oxford: Oxford University Press, 2003.

Clarke, S. *A Demonstration of the Being and Attributes of God And Other Writings*. Edited by E. Vailati. Cambridge: Cambridge University Press, 1705/1998.

Craig, W.L. *The Kalām Cosmological Argument*. London: The Macmillan Press, 1979.

Craig, W.L. *The Cosmological Argument from Plato to Leibniz*. London: The Macmillan Press, 1980.

Craig, W.L. "In Defense of the *Kalām* Cosmological Argument." *Faith and Philosophy* 14.2 (1997): 236–247.

Dawkins, R. *The God Delusion.* Transworld Publishers, 2006.

Descartes, R. *Meditations on First Philosophy* in *The Philosophical Writings of Descartes*, *Volume 2*. Translated by J. Cottingham, R. Stoothoff, and D. Murdoch. Cambridge: Cambridge University Press, 1641/1984.

Gale, R. and Pruss, A.R. "A New Cosmological Argument." *Religious Studies* 35 (1999): 461–476.

[29] The authors would like to thank an anonymous referee for extremely helpful comments on an earlier draft.

Hume, D. *Dialogues Concerning Natural Religion*. Indianapolis: Hackett, 1779/1980.

Kane, R. *The Significance of Free Will*. New York: Oxford University Press, 1996.

Kant, I. *Critique of Pure Reason*. Cambridge: Cambridge University Press, 1980/1998.

Koons, R. "A New Look at the Cosmological Argument." *American Philosophical Quarterly* 34 (1997): 171-192.

Leibniz, G. *The Monadology*. Translated by George MacDonald Ross, 1714/1999.

Morriston, W. "The Evidential Argument from Goodness." *The Southern Journal of Philosophy* XLII (2004): 87-101.

O'Conner, T. "Free Will." *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*. Edited by E. N. Zalta, URL = http://plato.stanford.edu/ archives/ fall2008 /entries/freewill/.

Oppy, G. *Arguing about Gods*. Cambridge: Cambridge University Press, 2006.

Pascal, B. *Pascal's Pensées*. Translated by W. F. Trotter, 1910.

Plato. *The Laws of Plato*. Translated by T.L. Pangle. New York : Basic Books, 1980.

Pruss, A. "Leibnizian Cosmological Arguments." *The Blackwell Companion to Natural Theology*. Edited by W. L. Craig and J.P. Moreland. Oxford: Blackwell Publishing, 1999.

Reichenbach, B. "Cosmological Argument." *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*. Edited by E. N. Zalta. URL = http://plato.stanford.edu/archives/ fall2008/entries/cosmological-argument/, 2008.

Rowe, W. *The Cosmological Argument*. Princeton: Princeton University Press, 1975.

Spinoza, B. *The Collected Works of Spinoza Vol. I*. Edited and Translated by Edwin Curley. Princeton: Princeton University Press, 1677/1985.

Taylor, R. *Metaphysics*. Englewood Cliffs: Prentice-Hall, 1992.

TIME, TRUTH, ACTUALITY, AND CAUSATION: ON THE IMPOSSIBILITY OF DIVINE FOREKNOWLEDGE

MICHAEL TOOLEY

University of Colorado at Boulder

Abstract. In this essay, my goal is, first, to describe the most important contemporary philosophical approaches to the nature of time, and then, secondly, to discuss the ways in which those different accounts bear upon the question of the possibility of divine foreknowledge. I shall argue that different accounts of the nature of time give rise to different objections to the idea of divine foreknowledge, but that, in addition, there is a general argument for the impossibility of divine foreknowledge that is independent of one's account of the nature of time.

I. A FUNDAMENTAL DIVIDE: STATIC VERSUS DYNAMIC VIEWS OF TIME

1.1. An Intuitive Characterization of the Divide

The most fundamental divide between different philosophical theories concerning the nature of time is that between what are usually referred to, on the one hand, as *tenseless* theories of time, and, on the other, *tensed* theories – though some philosophers, and this is my own predilection, prefer to talk instead of a distinction between *static* theories of time and *dynamic* theories of time.

One way of describing this divide is in terms of different accounts of tensed concepts – that is, the concept of the present, and of other concepts that involve the concept of the present, such as the concept of the past and the concept of the future. Thus, according to tensed or dynamic approaches to the nature of time, tensed concepts are ontologically significant, either because there are special tensed *properties* that events can have or fail to have – including the property of *presentness*, and, on some

accounts, also the properties of pastness and futurity – or, alternatively, because there are divisions between the past, the present, and the future, with regard to *existence itself*.

By contrast, according to tenseless or static approaches to the nature of time, there are no special properties of presentness, pastness, or futurity, nor is there any ontological divide between past, present, and future: events past, present, and future are all equally real.

How, then, are tensed concepts to be understood, according to tenseless or static approaches to the nature of time? The answer is that the term "now", and similar expressions such as "the present", along with present-tense endings on verbs, are all to be understood as *pure indexical* terms, on a par with the spatial indexicals "here" and "there". Thus, just as to say that something is happening here, rather than there, is neither to assign some special, 'spatially tensed' property to that event, nor to refer to some ontological gulf between things that are here and things that are there, but, rather, is simply to say what could equally be said by saying that the event in question is happening at *this spatial* location, so to say that an event is happening now is simply to say that the event is happening at *this very moment*, at *this temporal* location. Similarly, to say that an event happened in the past, or will happen in the future, is simply to say that the event is, respectively, either earlier than this very moment, or later than this very moment.

1.2. Truth and Actuality: Simpliciter Versus Temporally-Indexed

There is, however, what is to my mind a philosophically more important way of describing the difference between static and dynamic accounts of the nature of time. If there are tensed properties of presentness, pastness, and futurity, then events will change with respect to those properties: an event will initially have the property of futurity, will lose it and acquire the property of presentness, which it will have for only a moment, and it will then lose that property, and acquire the property of pastness. Which tensed propositions are true with regard to a given event will therefore depend upon what time it is.

Similarly, if there is some ontological divide among past events, present events, and future events, then there will be times as of which

a given event is not actual, is not real, and other times when it is actual. Consequently, propositions about a given event will, once again, not always have the same truth-value.

Compare the situation if a static or tenseless view of time is correct. The truth-value of a proposition will, in that case, not depend upon what time it is, since a given event cannot lie on different sides of an ontological divide between existence and nonexistence at different times: all events, past, present, and future, are equally real. Nor, if there are no special tensed properties, is there any possibility for an event to differ with regard to what tensed properties it has at different times.

It is true that what tensed *sentences* are true will depend upon what time it is. At one moment, it will be true to say that event E is now happening, while at a later moment it will be true instead to say that event E has happened, and not true to say that it is now happening. But that doesn't mean that there is some *proposition* that has gone from being true to being false. For if the term "now" is, as defenders of tenseless approaches to time hold, simply a temporal indexical, then the utterance of a given type of tensed sentence at different times must necessarily express *different propositions*, since the temporal indexical term – such as "now" – refers indexically to different times when the sentence is uttered at different times.

The upshot is this. According to static or tenseless views of time, the appropriate concept of truth is what might be called *truth simpliciter*, understood as a relation between propositions and what is *actual simpliciter*. For if all events, past, present, and future are equally real, and there are no special tensed properties, then the totality of what is actual cannot differ from one moment to another, and because that cannot differ, what propositions are true cannot differ from one moment to another. The only relevant notions of truth and actuality, therefore, are the concept of what is true *simpliciter*, and the concept of what is actual *simpliciter*: there is no place for any temporally indexed concept of truth, or for any temporally indexed concept of what is actual.

The situation is very different given a dynamic or tensed approach to the nature of time, since if there is some ontological divide between past, present, and future events with respect to reality, or if there are special tensed properties that objects can acquire and lose, then a given proposition can have different truth values at different times. So, although the

notion of truth *simpliciter* may also be needed – I have argued elsewhere that it is – a temporally indexed notion of truth – the concept of truth at a time – is absolutely indispensible given any tensed or dynamic approach to the nature of time. Similarly, a temporally indexed notion of actuality is needed – the concept of what is actual *as of* a given time – as can be seen in either of two ways. First of all, given a correspondence theory of truth, what propositions are true cannot vary from one time to another unless the totality of what is actual as of one time differs from the totality of what is actual as of another time. Secondly, if there are special tensed properties, then the tensed states of affairs that any given event is involved in must differ between some times: at one time, the state of affairs that is event E's having the property of presentness belongs to the totality of what is actual, whereas, at other times, it does not, having been replaced by the state of affairs that consists of event E's having instead the property of pastness. Alternatively, if the dynamic nature of the world derives instead from some ontological divide between past, present, and future events, then the totality of what is actual as of one time will include event E, whereas the totality of what is actual as of some other time may very well not include event E.

To sum up, then, there are, on the one hand, the absolute concepts of *truth simpliciter* and of what is *actual simpliciter*, and, on the other hand, there are the temporally indexed notions of *truth at a time,* and of the totality of what is *actual as of a given time*. The latter pair of notions is crucial for any tensed or dynamic view of the nature of time, whereas only the former, absolute notions of truth and of actuality play any role in the case of tenseless or static views of the nature of time.

II. DIFFERENT THEORIES OF THE NATURE OF TIME

2.1. Static Theories of the Nature of Time

Static theories of the nature of time do differ on certain matters, of which the most important are perhaps the following:

(1) *Precisely* what account is to be given of *tensed* sentences?
(2) What account is to be given of the *later than* relation, and of the

> direction of time? Is the later than relation a basic, unanalyzable relation? Is the direction of time to be explained in terms of some concept in physics, such as the direction of increase in entropy, or the expansion of the universe? Or is the direction of time to be analyzed in terms of the relation of causation?

These differences, however, philosophically important though they are, are not crucial, I think, in the context of questions concerning the possibility of divine foreknowledge. What is important in that context is the thesis, shared by all static approaches to time, that all events, past, present, and future, are equally real, and that there cannot be any change, over time, with regard to the totality of what is actual.

2.2. Dynamic Theories of the Nature of Time

Tensed or dynamic theories of the nature of time take a number of very different forms, of which the most important, I think, are the following.

2.2.1. The 'Growing Block' View

This is the view that the states of affairs that are actual *as of* any given time are the states of affairs that exist *at* that time plus the states of affairs that exist at earlier times, but not the states of affairs that exist at later times. In short, past and present events are actual as of a given time, but not future events.

On this view there is no special, *intrinsic* property of presentness. Presentness is, instead, a *structural property*: the states of affairs that are present at a given time are the states of affairs that are the *latest* among the totality of states of affairs that are actual as of the time in question.

2.2.2. Storrs McCall's 'Concrete Possible Futures' View of Time

A different view, but one that is in most respects closely related to growing block views, has been advanced by the Canadian philosopher, Storrs McCall (1994). His idea is to treat future possibilities in a way that is similar to that in which David Lewis treated all logical possibilities, so

that just as Lewis held that each logically possible state of affairs exists somewhere in a *concrete* possible world, so McCall holds that possible future states of affairs all exist as concrete states of affairs. What happens, then, is that as time passes, there is a moment when all of the branches that existed at that time drop out of existence, except for a single branch. Before that happens, there were multiple possibilities for branching at the time in question, and because of that, it was *indeterminate* what would be the case at that time. The present is thus the point at which the indeterminacy that existed concerning what would be the case at the time in question is replaced by a single, determinate state of affairs.

As in the case of the growing block view, then, presentness, rather than being an intrinsic property of events, is a structural property: an event lies in the present when it lies at the point where the branching representing future possibilities begins.

2.2.3. Presentism

This is the view, as I would characterize it, that the only states of affairs that are actual as of a given time are the states of affairs that exist at that time. So stated, the question arises as to how any propositions about the past can be true. Advocates of presentism have attempted to answer that question in various ways, but the main one involves the idea that states of affairs that exist at any given time are not restricted to what might be called 'present tense' states of affairs: there are also past tense states of affairs, such as the state of affairs that consists of Caesar's having crossed the Rubicon, and, perhaps also, future tense states of affairs, such as the first unicorn being created five years from now.

Though this position is rather popular, especially in the United States, I believe it is open to decisive objections. In particular, I would argue, first, that the presentist can offer no satisfactory account of the concepts of *the past* and of *the future*; secondly, that on presentism there are *no truthmakers for propositions about the past*; and thirdly, that the presentist cannot offer a satisfactory account of *cross-temporal relations*, including causation.

2.2.4. The 'Moving Now' View

A final tensed view that is interesting, but not very popular today, is very close to tenseless or static views of time in two respects. First of all, it involves the claim that all events, past, present, and future, are actual as of every given moment. Secondly, it claims that events are temporally ordered by a later than relation – a relation that can either be taken as primitive, or analyzed in one of the ways adopted in static accounts of the nature of time.

How, then, does this view of time differ from static views of time? The answer is that it is claimed that there is a *special, intrinsic* property of *presentness* that moves along the series of events ordered by the later than relation, so that every event has this intrinsic property of presentness for a moment. Before it acquires that property, an event lies in the future. Once it loses that property, it lies in the past. The passage of time consists, then, in the movement of this property: it is this movement that transforms what would otherwise be what McTaggart referred to as the 'B-series' into McTaggart's 'A-series'.

III. THE ONTOLOGICAL STATUS OF FUTURE CONTINGENT EVENTS

From a general metaphysical perspective, the crucial divide in the philosophy of time is that between static theories and dynamic theories. It seems to me, however, that with regard to issues raised by the idea of divine foreknowledge, the crucial divide is a different one – namely, to express it in ontological terms, between views on which future states of affairs are either actual s*impliciter*, or else actual as of the present moment, and views on which neither of those things is the case. Or, alternatively, to express it in semantical terms, the crucial divide with regard to issues raised by the idea of divine foreknowledge is that between views of the nature of time on which propositions about future contingent events are either true or false, *simpliciter*, or else true or false at the present moment, and views on which neither of these things is the case, since propositions about future contingents, rather than being true or false,

or true or false at the present time, have, instead, the third truth-value, namely, *indeterminacy*.

If this is right, the crucial division is between, on the one hand, static views of the nature of time and the moving Now view, and, on the other, the growing block view and McCall's concrete-possible-futures view. The main thing that I want to do in the remainder of this talk, then, is to focus on two different problems that arise for divine foreknowledge, one of which arises if future events are real, and the other of which arises if they are not.

IV. HUMAN FREEDOM IN A WORLD WITH TENSELESS TIME

Before doing that, however, I want to mention briefly an argument that is sometimes advanced on the basis of the assumption that a static or tenseless view of time is true. It is an argument that, if correct, would mean that there would be no need to consider static or tenseless approaches to the nature of time in a context where we are interested in the idea of divine foreknowledge of contingent future events.

4.1. An Argument for Logical Fatalism

The argument in question can be expressed as follows:

(1) A static or tenseless view of time is correct.
(2) Let *p* be some proposition about a future event, such as the proposition that John will perform action A, where A is some action that can be performed at time *t* in the year 2012.
(3) According to a static or tenseless view of time, every proposition is either true *simpliciter*, or else false *simpliciter*.
(4) Therefore, either *p* is true, or *p* is false.
(5) If a static or tenseless view of time is correct, the truth-value of a proposition never changes.
(6) If the truth-value of a proposition never changes, then either a given proposition is always true, or it is always false.

(7) So if *p* is true, then *p* is true at all times; while if *p* is false, then *p* is false at all times.
(8) Hence, if *p* is true, then *p* is true now; while if *p* is false, *p* is false now.
(9) Therefore, either it is now true that John will perform action A, or it is now false that John will perform action A.
(10) If it is *now* true that John will perform action A, then it is impossible for John not to perform action A.
(11) If it is impossible for John not to perform action A, then John is not free in a libertarian sense with regard to the choice of whether to perform action A, or to refrain from performing action A.
(12) If it is *now* false that John will perform action A, then it is impossible for John to perform action A.
(13) If it is impossible for John to perform action A, then John is not free in a libertarian sense with regard to the choice of whether to perform action A, or to refrain from performing action A.
(14) Therefore, either way, John is not free in a libertarian sense with regard to the choice of whether to perform action A, or to refrain from performing action A.
(15) In general, no one is ever free in a libertarian sense with regard to the choice of whether to perform a given action A, or to refrain from performing that action.

4.2. The Unsoundness of this Argument

Notice that if this argument were sound, it would show not only that a static or tenseless view of time is incompatible with libertarian free will, but also that such a view of time is incompatible with there being any indeterministic events. For a precisely parallel argument could be used to show, for example, that if an atom undergoes decay at time *t*, then it was logically determined that it would do so.

The argument, however, is unsound. The error occurs at step (6), where the following premise is used:

> If the truth-value of a proposition never changes, then either a given proposition is always true, or it is always false.

The problem with this premise is that it imports into a static or tenseless view of time the notion of truth at a time. Given a static or tenseless approach to the nature of time, however, the notion of truth at a time is either incoherent, or, at least, has no application to the world. For the notion of truth at a time only makes sense if the idea of being actual as of a time makes sense, and given a static or tenseless approach to time, no sense can be assigned to a temporally-indexed concept of what is actual.

Another way of putting the point is this. On a static or tenseless view of time, the *truthmaker* for a proposition such as that John will perform action A, defined as a certain action that can be performed at time *t* in the year 2012 is a state of affairs that exists at the time in question in the year 2012. If one says that the proposition in question is *now true*, that is naturally interpreted as saying that a state of affairs *now exists* that makes the proposition true. But the latter is not the case: there is no such state of affairs. The only truthmaker is the state of affairs that consists of John's performing action A at time *t* in the year 2012, and the fact that John performs that action does nothing at all to make it the case that, before he performed that action, he could not have refrained from performing it.

V. DIVINE FOREKNOWLEDGE AND THE REALITY OR UNREALITY OF FUTURE CONTINGENT EVENTS

Let us now consider what potential problems there are for the possibility of divine foreknowledge, given different philosophical accounts of the nature of time. Basically, there are three questions that we need to answer. First of all, how do things stand if the growing block view of time is correct? According to that view, future contingent events are not real, are not actual, as of earlier times. If that is the case, what are the implications for the possibility of divine foreknowledge of future contingent events?

Secondly, there is Storrs McCall's concrete-possible-futures view of time. Here, rather than there being no future events that are actual as of earlier times, all future events are actual, but so, equally, are all future *possibilities*, and there is nothing in that plethora of concrete future possibilities that picks out the one pathway through all those future

branchings as the unique sequence of events that will actually obtain. So what implications does this picture have for the possibility of divine foreknowledge?

Finally, how do things stand given a static or tenseless view of time, or on a moving Now view? According to a static or tenseless view, future contingent events are actual *simpliciter*, while according to a moving Now view, all events, past, present, and future, except for those that involve the tensed property of presentness, are actual as of every moment. So on either a static view of time, or a moving Now view, one has neither an absence of future contingent states of affairs, as on the growing block view, nor a multitude of concrete, future possibilities, with nothing to mark out the actual future, as on Storrs McCall's concrete-possible-futures view: one has, instead, just the actual future events, and so one does have events that could in principle be related to divine foreknowledge. Is divine foreknowledge possible, then, given either of these views?

VI. DIVINE FOREKNOWLEDGE, COUNTERFACTUALS, AND CAUSATION

What is involved in divine foreknowledge? On the face of it, there would seem to be three ways in which a person could have knowledge of future events. First of all, the person could be an agent who had the power to bring about the future event in question, and could decide to do that. Knowing what he had decided to attempt to do, and knowing that he had the power to do that, he would be in a position to know that the future event in question would occur.

Secondly, if the world contained deterministic laws of nature, and if a person had knowledge of relevant laws of nature, and of the present state of the world, that person would be in a position to draw conclusions about what future events would occur.

But when one is concerned with divine knowledge of future contingent events, such as what some agent with libertarian free will is going to do, neither of the two possibilities just mentioned apply: the actions of a free agent are not determined, either by God's actions, or by the operation of laws of nature. So what other possibility is there?

A third possibility that certainly springs to mind is that future contingent events could cause corresponding beliefs that are present in the mind of God at an earlier time – indeed, at all earlier times.

This third possibility requires *backward causation*, and so one might wonder whether there is some other possibility.

Here is an argument, however, in support of the conclusion that divine foreknowledge of future contingent events does require causation running from those future contingent events to beliefs in the mind of God:

(1) Divine foreknowledge requires that God's beliefs about the future are *counterfactually dependent* upon the future events in question, since if some future event had not occurred, God would not have believed that it was going to occur.

(2) There are two main approaches to counterfactuals. First, there are the *similarity-across-possible-worlds accounts*, advanced by Robert Stalnaker (1968) and David Lewis (1973), according to which, on Lewis's version, the counterfactual proposition that if p were true, q would be true, is true in the actual world if there is a possible world in which both p and q are true that is closer to the actual world than any possible world in which p is true but q is false. Secondly, there are *causal approaches* to counterfactuals advocated by, among many others, Frank Jackson (1977) and Igal Kvart (1986).

(3) Jonathan Bennett (1974) and Kit Fine (1975), in early reviews of David Lewis's book *Counterfactuals*, argued that a similarity-across-possible-worlds approach to counterfactuals leads to incorrect assignments of truth-values to certain counterfactuals. Lewis (1979), a few years later, set out a method of avoiding the specific objections that Bennett and Fine advanced. But it can be shown that their objections can be modified to produce objections that cannot be blocked (Tooley, 2003). The conclusion, in short, is that similarity-across-possible-worlds approaches to counterfactuals are open to decisive objections.

(4) The upshot is that there does not appear to be any viable alternative to some sort of causal account of counterfactuals, so that, given that God's beliefs would have to be counterfactually dependent upon future contingent events, we have the following conclusion:

> Divine foreknowledge of future contingent events requires that those events causally give rise to God's beliefs about them.

VII. DIVINE FOREKNOWLEDGE AND THE GROWING BLOCK VIEW OF TIME

Given that conclusion, how do things stand if a growing block view of time is correct? To answer that question, one has to ask how the actuality of events is related to causation.

In the argument that I offered in *Time, Tense, and Causation* for a growing block view of time, I attempted to show that the postulates that provide a correct analysis of causation can only be satisfied in a world where causes and their effects are related in an *asymmetric* way to what is actual. In particular, I argued that, on the one hand, the effect of a cause cannot be actual as of the time of its cause, while, on the other hand, a cause must be actual as of the time of its effect.

Put intuitively, the former must be the case if a cause is to bring its effect into existence, while the latter is required since it must be true at the time of the effect that it had a certain cause, and this cannot be so at that time unless both the effect and its cause are actual as of the time of the effect.

But if this is right, then, on a growing block view of the nature of time, future contingent events cannot causally give rise to earlier beliefs in the mind of God, since no future contingent event is actual as of any earlier time.

The conclusion, in short, is that the combination of a growing block view of time with what appear to be plausible claims about when a cause and its effect are actual poses a problem for the idea of divine foreknowledge.

VIII. DIVINE FOREKNOWLEDGE AND THE CONCRETE-POSSIBLE-FUTURES VIEW OF TIME

On Storrs McCall's concrete-possible-futures account of the nature of time, all of the future contingent events that will be actual future events are actual as of earlier times. So there could be a causal connection between all of those events and beliefs in the mind of God.

But the problem is this. Let *t* be any future time, and let A and B be possible future events that, if they occur, will occur at time *t*. Suppose

that event A will actually occur at time *t*, and that B will not. It could be the case that the concrete future possibility that is A will give rise to a belief in the mind of God that A will occur, while the concrete future possibility that is B will not give rise to a belief in the mind of God that B will occur. But what could possibly *ensure* that this will be the case? It seems to me that there is no answer to this question. The thing in question *could* occur, but if it did it would be a *pure accident*. For if there were something in the world that could ensure that the future possibilities that will later be the ones that are realized will give rise to corresponding beliefs in the mind of God, there would have to be something in what was actual as of any relevant time that marked out just those concrete future possibilities, something that singled them out as the ones that would be realized, But on McCall's model of the nature of time, there is, by definition, nothing that does this.

Finally, I also think that McCall's concrete-possible-futures view of time is open to strong objections. For one thing, what future possibilities exist at any given time logically supervenes on the totality of the states of affairs that exist *at* that time, together with laws of nature, so there is absolutely no reason to postulate *concrete* future possibilities.

McCall's response to this objection is that, rather than viewing future possibilities as logically supervenient upon laws of nature plus present states of affairs, one can instead view *laws of nature* as logically supervenient upon future possibilities. But the latter idea is open to two very strong objections. The nomological possibilities that exist at any one time *will* suffice to fix what laws of nature there are, but then what makes something a law involves an infinite set of states of affairs, and this in turn makes it extremely improbable that any laws will exist. So the result of McCall's proposal is that there cannot be any solution to the problem of justifying induction.

Secondly, the situation is made worse by the fact that, for any two future times that one picks, the concrete possibilities that exist at those times must define *precisely the same set of laws of nature*. If laws of nature were primary, this would pose no problem at all. But if it is concrete future possibilities that are, instead, primary, then it is an inexplicable cosmic coincidence that the concrete future possibilities that exist at each of an infinite number of future times all logically determine precisely the same laws of nature.

IX. DIVINE FOREKNOWLEDGE AND BACKWARD CAUSATION

Let us now consider how things stand with regard to the possibility of divine foreknowledge if either a static or tenseless view of time is correct, or a moving Now account.

If it is the case either that future events are actual as of the present moment, as the moving Now account of time holds, or, as static approaches to time maintain, future events are actual *simpliciter*, and a temporally-indexed view of actuality is metaphysically mistaken, then there is no *initial* barrier to God's having true beliefs about future contingent events, since there are events that are either actual *simpliciter*, or actual as of the relevant time, that could, barring any problems, causally give rise to God's beliefs.

The causation in question would, of course, be temporally backward causation, and some philosophers, such as Hugh Mellor in his books *Real Time* (1980), *The Facts of Causation* (1995), and *Real Time II* (1998), have argued that backward causation is logically impossible.

There is a central problem that afflicts at least most such arguments, however – namely, that they attempt to show that *backward causation* is logically impossible by first showing that *causal loops* are logically impossible, and then by arguing that if backward causation were logically possible, causal loops would also be logically possible. But the second part of this type of argument is very difficult to defend, since there certainly seem to be logically possible worlds that contain backward causation, but in which causal loops are nomologically impossible.

In *Time, Tense, and Causation*, I attempted to show that backward causation, even when it did not give rise to causal loops, was logically impossible. The argument that I offered, however, also supported a growing block view of time, and so an appeal to that argument cannot be combined with either a tenseless view of time, or a moving Now view.

The upshot is that it is not clear that divine foreknowledge can be ruled out via an appeal to any proof of the general impossibility of backward causation.

X. DIVINE FOREKNOWLEDGE AND THE POSSIBILITY OF UNDERCUTTING CAUSAL LOOPS

Even if such arguments for the logical impossibility of backward causation can be shown to be unsound, however, this does not eliminate a nearby potential problem for divine foreknowledge. The reason is that even if backward causation is logically possible, it may be that causal loops are not.

In his paper "The Paradoxes of Time Travel" (1976, 149), David Lewis describes the following time travel case:

> Consider Tim. He detests his grandfather, whose success in the munitions trade built the family fortune that paid for Tim's time machine. Tim would like nothing so much as to kill Grandfather, but alas he is too late. Grandfather died in his bed in 1957, while Tim was a young boy. But when Tim has built his time machine and traveled to 1920, suddenly he realizes that he is not too late after all. He buys a rifle; he spends long hours in target practice; he shadows Grandfather to learn the route of his daily walk to the munitions works; he rents a room along the route; and there he lurks, one winter day in 1921, rifle loaded, hate in his heart, as Grandfather walks closer, closer, ...

A serious problem is clearly visible. Since Grandfather, in Lewis's story, dies in bed in 1957, evidently Tim, for some reason, did not succeed in killing Grandfather. But might he not have succeeded? Must it not have been possible for him to succeed? But if he had, and if we assume – as is part of Lewis's story – that Grandfather had not yet fathered the son who was Tim's father – then Tim would, in this *counterfactual variant* on Lewis's story, have undercut the causal chain leading to his own existence. So we would have a self-undercutting causal chain: Tim's killing of Grandfather would mean no more Grandfather, and, therefore, no father, and, therefore, no Tim.

The problem, in short, is that even if backward causation in general is not logically impossible, if the world is one where backward causal processes can connect up with forward causal processes, and vice versa, then there is the possibility of *causal loops*. Some of those causal loops may be *self-supporting* ones, which seems puzzling. Much more threatening, however, are causal loops of the *self-undercutting* variety. Indeed, mightn't it be argued that such self-undercutting causal loops entail con-

tradictions: if Tim kills Grandfather, then Tim beings it about that Tim doesn't exist.

Lewis argues that the counterfactual variant does not generate such consequences:

> If you suppose Tim to kill Grandfather and hold all the rest of his story fixed, of course you get a contradiction. But likewise if you suppose Tom to kill Grandfather's partner and hold the rest of his story fixed–including the part that told of his failure–you get a contradiction. If you make *any* counterfactual assumption and hold all else fixed you get a contradiction. The thing to do is rather to make the counterfactual supposition and hold all else as close to fixed as you consistently can. That procedure will yield perfectly consistent answers to the question: what if Tim had killed Grandfather? In that case some of the story that I told would not have been true. Perhaps Tim might have been the time-traveling grandson of someone else. Perhaps he might have been the grandson of a man killed in 1921 and miraculously resurrected. Perhaps he might have been not a time-traveler at all, but rather someone created out of nothing in 1920 and equipped with false memories of a personal past that never was. It is hard to say what is the least revision of Tim's story to make it true that Tim kills Grandfather, but certainly the contradictory story in which the killing both does and doesn't occur is not the least revision. Hence it is false (according to the unrevised story) that if Tim had killed Grandfather then contradictions would have been true. (1976, 152)

Is Lewis's response satisfactory? I think that it is not, and the reason is this. First of all, Lewis's discussion of what would be the case if Tim killed Grandfather makes use of his own similarity-across-possible-worlds approach to counterfactuals: what would happen is a matter of what is the case in the closest possible worlds in which Tim kills Grandfather. But such an approach to counterfactuals is demonstrably unsound. Secondly, if one switches to a causal approach to such counterfactuals, what would have been the case if Tim killed Grandfather is fixed by what follows in virtue of causal laws from the proposition that Tim killed Grandfather combined with propositions that describe the minimally modified world *at the time in question*, or shortly before it, which is such that Tim kills Grandfather. But when this approach to counterfactuals is adopted, the combination of causal laws and the propositions describing that minimally modified world certainly seem to entail Tim's non-existence.

This whole question certainly needs to be looked at in a much more detailed way than I have just done. My point, however, is that this question *is crucial* for the question of the possibility of divine foreknowledge. For if, for example, God at some time has foreknowledge of the fact that a person, John, will perform action A at time *t* in 2012, and has that foreknowledge because the state of affairs that is John's performing action A at time *t* in 2012 brings about, via backward causation, God's belief that John will perform action A at time *t* in 2012, then God, at the earlier time, has the power to make a contradiction true, for being omnipotent, he certainly has the power to prevent John from performing action A at time *t* in 2012.

The argument, in short, is this. Assume the following thesis, which is to be refuted:

(1) Divine foreknowledge is possible because (a) it is possible either that a static or tenseless view of time is right, or that a moving Now view of time is right, (b) on either view, there are future events that are either actual *simpliciter*, or actual as of earlier times, and (c) backward causation is logically possible, so that God can have foreknowledge of a future state of affairs, S, by having beliefs that are caused by S.

The refutation of (1) now proceeds as follows:

(2) God's omnipotence makes it possible for him to intervene at the earlier time to prevent the later state of affairs S from occurring.
(3) Given a correct, causal account of the relevant counterfactuals, it follows that if God were to act to prevent the later state of affairs S, it would be true both that S will occur and that S will not occur, and so a contradiction would be true.
(4) There cannot be true contradictions.
(5) Hence (1) is false. The defense of the possibility of divine foreknowledge proposed there cannot be correct.

XI. THE GENERAL APPLICABILITY OF THIS FINAL OBJECTION

I have directed this last argument against the conjunction of the proposition that divine foreknowledge is logically possible with the proposition that future events are either actual *simpliciter*, or actual as of the present moment. In fact, however, the objection applies regardless of which view of the nature of time one embraces. For on any view of time, divine foreknowledge requires that contingent future events can give rise to beliefs, or belief-like states, in the mind of God, and God, being omnipotent, can then act on the world to bring about a state of the world at some time prior to the contingent, future event in question, where the state is such as to rule out the occurrence of the contingent, future event in question. Regardless of the view that one adopts concerning the nature of time, then, divine foreknowledge would entail God's having the power of making contradictions true. But it is logically impossible for contradictions to be true, and so it is logically impossible for there to be such a power. Accordingly, divine foreknowledge is logically impossible.

SUMMING UP

1. With regard to the possibility of divine foreknowledge, the crucial division with respect to philosophical views on the nature of time is between, on the one hand, views according to which future states of affairs are actual, or else actual as of earlier times, and, on the other hand, views according to which there are no determinate future events that are actual, or actual as of earlier times.

2. There are good reasons for holding that divine foreknowledge of future contingent events that are neither causally determined nor caused by God requires that those future events causally give rise to beliefs in the mind of God at earlier times.

3. Causation is either a relation between states of affairs both of which are actual *simpliciter*, or else a relation where the cause is actual as of the time of its effect.

4. It follows, given a growing block view, that backward causation is not logically possible.

5. On Storrs McCall's concrete-future-possibilities view of time, the problem is that there is nothing that can make it the case that those concrete future possibilities that will ultimately be realized give rise to corresponding beliefs in the mind of God.

6. If either the moving Now view of time, or a static or tenseless view of time is correct, then there are future states of affairs that either are actual, or else actual as of earlier times, so if backward causation is logically possible, it initially looks as if future states of affairs can causally give rise to corresponding beliefs in the mind of God. But now the problem is that, because of God's omnipotence, undercutting causal loops are possible, and given a sound account of the truth conditions of counterfactuals, it appears to follow that God would have the power to make contradictions true. Since nothing can have that power, this account of the possibility of divine foreknowledge also appears untenable.

7. The argument based on the idea of undercutting causal loops is not restricted to the combination of divine foreknowledge with views of the nature of time according to which contingent, future events are either actual *simpliciter*, or actual as of earlier times. It is an argument against the possibility of divine foreknowledge regardless of what view one adopts concerning the nature of time

8. *Overall Conclusion*: The idea of divine foreknowledge gives rise to very serious problems, and there is a very plausible, and completely general argument that appears to show that divine foreknowledge is logically impossible.

BIBLIOGRAPHY

Bennett, Jonathan, 1974, "Counterfactuals and Possible Worlds," *Canadian Journal of Philosophy*, 4, 381-402.

Fine, Kit, 1975, "Critical Notice – Counterfactuals," *Mind*, 84, 451-8.

Jackson, Frank, 1977, "A Causal Theory of Counterfactuals," *Australasian Journal of Philosophy*, 55, 3-21.

Kvart, Igal, 1986, *A Theory of Counterfactuals*, Indianapolis: Hackett Publishing Company.

Lewis, David, 1973, *Counterfactuals*, Cambridge, Massachusetts: Harvard University Press.

Lewis, David, 1976, "The Paradoxes of Time Travel," *American Philosophical Quarterly* 13 (1976): 145-52. Reprinted in David Lewis, 1986, *Philosophical Papers – Volume II*, Oxford, Oxford University Press, pp. 67-80.

Lewis, David, 1979, "Counterfactual Dependence and Time's Arrow," *Noûs*, 13, 455-76.

McCall, Storrs, 1994, *A Model of the Universe*, Oxford: Oxford University Press.

Mellor, D. H., 1981, *Real Time*, Cambridge: Cambridge University Press.

Mellor, D. H., 1995, *The Facts of Causation*, London: Routledge.

Mellor, D. H., 1998, *Real Time II*, London: Routledge.

Stalnaker, Robert, 1968, "A Theory of Counterfactuals", in Nicholas Rescher (ed.), *Studies in Logical Theory*, Oxford: Blackwell, pp. 98-112.

Tooley, Michael, 1997, *Time, Tense, and Causation*, Oxford: Oxford University Press. Reprinted in a revised, paperback edition in 2000.

Tooley, Michael, 2003, "The Stalnaker-Lewis Approach to Counterfactuals," *Journal of Philosophy*, 100/7, 371-77.

THE ETERNITY SOLUTION TO THE PROBLEM OF HUMAN FREEDOM AND DIVINE FOREKNOWLEDGE

MICHAEL ROTA

University of St Thomas

Abstract. In this paper I defend the eternity solution to the problem of human freedom and divine foreknowledge. After motivating the problem, I sketch the basic contours of the eternity solution. I then consider several objections which contend that the eternity solution falsely implies that we have various powers (e.g. to change God's beliefs, or to affect the past) which, according to the objector, we do not in fact have.

I. WHAT'S THE PROBLEM?

Say that a human being H is *free in the libertarian sense* iff at least sometimes, (i) there are multiple possible alternatives for H's action, and (ii) it is up to H which alternative gets realized.[1] There are some good reasons, I think, to believe that human beings are free in the libertarian sense. My main task in this paper, however, is not to defend the view that humans have libertarian freedom, but to discuss a problem for theists who believe that humans have libertarian freedom.

Regardless of your actual beliefs, suppose for a moment that you do believe that humans have libertarian freedom. And imagine that a friend of yours in the archaeology department asks you if you would be willing to help in an archaeological dig near Rome. You agree, and spend several weeks digging up artifacts in what was once a monastery in the seventh century A.D. You find various artifacts that have not seen the light of day for 1300 years. And then you find something exceptionally unusual. You

[1] I take this approach from Peter Unger, "Free Will and Scientiphicalism," *Philosophy and Phenomenological Research* 65 (2002): 1-25.

yourself uncover a box containing a well-preserved Latin manuscript, buried all these long centuries. Being a well-trained Latinist, you begin to read the manuscript, and, to your astonishment, find that the manuscript contains a narrative of your entire life thus far. It names and describes your parents, identifies the day of your birth, and describes the subsequent course of your life – what career you have chosen, your professional accomplishments, and whom you have married.

You would no doubt initially think that this was an elaborate joke concocted by your archaeologist colleague. But suppose that you had convincing evidence that this was not so – suppose that you somehow had convincing evidence that the manuscript had really been there, buried in the dirt for 1300 years. Now, if you were indeed convinced that the manuscript was real, what would you think about your own freedom? How could you have been in control of your life, when the whole course of your life, including your most personal decisions, was laid out long before you were born?

In such a situation, your belief that you had been in control of your life would be shaken, I should think. Such a scenario would seem to provide good evidence that you did not have libertarian freedom.

The traditional Christian theist has a very similar problem. According to traditional Christian theism, God has perfect knowledge of events that are future with respect to us. On the traditional Christian view, before I was even born it was true that God knew the whole course of my life. Although there is no manuscript describing my life buried somewhere in the ground, there is something else which might seem even more problematic. A much more detailed, infallibly accurate account of my future life has been contained, as it were, in the mind of God, since long before I was born. How then, can any of my actions really be up to me?

II. THE ETERNITY SOLUTION TO THE PROBLEM OF FREEDOM AND FOREKNOWLEDGE

In this section I'll attempt to sketch a model according to which God can know all of our future actions, including our actions that are free in the libertarian sense.

Suppose, with Boethius, Anselm, and Aquinas, that God is timelessly eternal. God is not located in the flow of time. He is outside time. Yet the events of all times are equally present to Him. The events of the temporal present are present to Him. But so are the events of last year, and the events of next year. They are all eternally present to Him.

If that's right, then we can tell a coherent story about how God could know what you will do tomorrow without taking away your freedom. Suppose that you will freely choose to answer the phone tomorrow at 9am. Then God knows (eternally) that you will choose to answer the phone tomorrow at 9am simply because He sees you doing it. And God does not see you choosing to answer the phone *before* you actually do so. Rather, for God in eternity your action is a present event, not a future event. In Aquinas's words: "…all things which are in time are present to God from eternity…because His glance is cast over all things from eternity, as those things exist in their own presentness".[2]

So, you freely choose to answer the phone at 9am tomorrow. *Because* you choose this then, God knows that you choose this then. But just as my knowledge in the temporal present of what you are doing does not take away your freedom, so too God's knowledge in the eternal present of what you are doing does not take away your freedom. God has perfect knowledge of all events that are future with respect to us simply because He sees them occur.

The four key elements of the eternity solution, as I conceive it, are as follows:

(1) God is timelessly eternal.

Proposition (1) implies, among other things, that no part of God's life is before or after any other part of God's life. God's entire life is present all at once in eternity.

(2) For any time (including all times that are past with respect to us and all times that are future with respect to us, as well as the

[2] *Summa Theologiae* I.14.13c. Translations of Aquinas are my own. I use the text available at the Corpus Thomisticum website, http://www.corpusthomisticum.org/iopera.html.

temporally present time), and for any event or substance temporally located at such a time, that event or substance is present to God.

(3) For any time (including all times that are past with respect to us and all times that are future with respect to us, as well as the temporally present time), and for any event or substance temporally located at such a time, God has immediate knowledge of that event or substance.

(4) God knows events that occur in time because they occur, and not the other way around. (It is not the case that they occur because God knows that they occur).

Let's call this set of four propositions the Basic Divine Eternity Model, or, for short, the Basic Model.[3]

III. OBJECTIONS RELATING TO POWERS

I now turn to a group of related objections – each makes the contention that the Basic Model implies that we have powers that we do not, according to the objector, in fact have.

3.1. Counterfactual power over God's eternal beliefs

Suppose that an agent S performs an act A (say, makes a choice) at a time T1, but that at an earlier time T0, S had the power not to do A at T1. If God is eternal, then God eternally sees S doing A at T1, and so God eternally believes that S does A at T1. Now focus on that earlier time, T0. At T0, S had the power not to do A at T1. But if S had exercised that power, and not done A at T1, then either God would have had a false belief or God would never have eternally believed that S does A at T1 in the first place. It is impossible that God have a false belief. So, if S had not done A at T1, then God would not have eternally believed something that He

[3] I call this the *Basic* Model because it can be developed in different ways, according to one's view about the nature of time.

in fact does eternally believe. Thus, the eternity model (together with the claim that at T0 S has the power not to do A at T1) implies that, at T0,

(5) S has the power to do something[4] such that if S did it, God would not eternally believe something that He does in fact eternally believe.

Note that proposition (5) does not imply that S has the power to *change* or *alter* God's eternal beliefs, at least not in the normal sense of the word 'change'. Changing God's beliefs would involve making it such that God once had one belief, and then later did not have that belief (or making it such that God once did not have some belief, and then later did have that belief). But there's no question of that sort of thing here. If (first situation) S does do A at T1, then God eternally believes that S does A at T1. And if (second situation) S does not do A at T1, then God eternally believes that S does not do A at T1. There is no change in God's beliefs in the first situation. And there is no change in God's beliefs in the second situation. So in neither situation (actual or counterfactual) is there a change in God's beliefs. So in neither situation (actual or counterfactual) does S do something that changes God's beliefs.

Now there is of course a difference between the two situations, and it is tempting to think that S has the power to move us from being in the first situation to being in the second. We imagine that we are first in the first situation, in which S is going to do A at T1. But then, we imagine, S exercises his power not to do A at T1, and now we are in the second situation. When we imagine things in this way, it is natural to think that S can do something which would change God's beliefs, because by moving us from one situation to the other, S would be moving us from a situation in which God believes one thing, to a situation in which God believes something else.

But this involves a confusion – it is not in fact the case that S can do something which would move reality from being in the first situation (in which S does A at T1) to being in the second situation (in which S does not do A at T1). If S does not in fact do A at T1, then we were in the second situation all along. This is because what situation we are in depends

[4] I'm counting "refraining from doing A at T1" as something S could do.

entirely on what S does at T1. And since S can do at most just one particular thing at T1, we must be in one particular situation all along.

So nothing about the eternity model requires that the power to *change* God's beliefs comes with the power to do otherwise. What does come with the power to do otherwise is the power to do something such that if one did it, God's eternal beliefs would be different than they are in the actual world. Call this counterfactual power over God's eternal beliefs.

3.1.1. An objection to our having counterfactual power over God's eternal beliefs based on the fixed or settled character of eternity

Some have objected that we could not really have such a power over God's eternal beliefs.[5] Some remarks of Linda Zagzebski's suggest the following argument against my claim that we do have counterfactual power over God's eternal beliefs:[6]

(For ease of expression I will develop this argument by talking about you and your power over God's eternal beliefs, but of course what I'll say would apply equally to everyone.)

The past is out of your reach, it is out of your control. No human being now has the power to do anything such that, if he did it, the past would have been different than it in fact is. And this is so because events in the past are fixed or settled. That is,

(6) Past events are out of your control because they are fixed or settled.

But if the reason that past events are out of your control is that they are settled, then any other sort of event that is settled will also be out of your control. So we have the general principle that

(7) If an event E is fixed or settled, then E is out of your control.

---

[5] See William Hasker, "The Absence of a Timeless God," in *God and Time: Essays on the Divine Nature*, ed. Gregory E. Ganssle and David M. Woodruff (Oxford: Oxford University Press, 2002), 198, and Linda Zagzebski, "Foreknowledge and Free Will," in *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, ed. Edward N. Zalta, http://plato.stanford.edu/archives/fall2008/entries/free-will-foreknowledge/, section 2.2.

[6] Linda Zagzebski, *The Dilemma of Freedom and Foreknowledge* (New York and Oxford: Oxford University Press, 1991), 60-63, and "Foreknowledge and Free Will," section 2.2.

And surely any event in eternity (any timeless event) is settled. "[W]hat can be more fixed than eternity?" Zagzebski asks.[7] Suppose, for example, you perform some act A at a future time T1. Then, since God's believing anything is a timeless event (on the doctrine of divine eternity), we have,

(8) The timeless event of God's eternally believing that you do A at T1 is fixed or settled.

From (7) and (8), it follows that,

(9) The timeless event of God's eternally believing that you do A at T1 is out of your control.

But this implies that you do not have counterfactual power over God's eternal beliefs. For if you did have such a power, that is, if there *was* something you could do such that, if you did it, then God would not eternally believe that you do A at T1, then you *would* have the sort of control which (9) says you don't have. So by modus tollens,

(10) You do not have the power to do anything such that if you did it, God would not eternally believe something that He does in fact eternally believe.

Although I grant that this argument is initially appealing, I think it has a serious flaw. We need to ask what "fixed" means. If "fixed" merely meant "not in your control" than (6) would be circular (the past is out of your control because it is out of your control). What's left is to interpret "fixed" as meaning "settled" or "actually a certain way", as I have done. But then (7) is false, as we can see by focusing on the temporal present.[8] Suppose you do A at T1. When T1 is the present time, you are doing A. But events in the temporal present are settled just as much as events in the past. At T1, it is a settled matter that you are doing A at T1. (7) would have us conclude that your doing A is therefore out of your control. But

[7] Linda Zagzebski, *The Dilemma of Freedom and Foreknowledge*, 60.

[8] Katherin Rogers uses this strategy in "The necessity of the present and Anselm's eternalist response to the problem of theological fatalism," *Religious Studies* 43:1 (March 2007), 25-47.

given that you have libertarian freedom with respect to A, this is false. So (7) is false.

Next: if (7) is false, the mere fixedness of an event is not that which explains why that event is out of your control. So the premise that

(6) Past events are out of your control because they are fixed or settled,

is false as well.

In my view, the brief argument just given successfully defeats the objection we're now considering. But we still might wonder, what does explain the common intuition that past events are out of our control, if not the fixedness of the past? I suggest that the real reason we tend to think that past events are out of our control is that we tend not to think that any action we can perform has the sort of causal or explanatory connection to past events that is required for control over those past events.

To explain what I have in mind, I'll begin with a paradigm case of an event that is in my control. Suppose that one minute from now I will freely walk across the room. Then the future event of my walking across the room is something that is in my control. Why do I have control over this event? Simple: I have control over this event because I can cause it to occur. I can will to stand up and will to move my legs in the right way, and thereby cause the event of my walking to take place. More generally, if it is (really) in my power to efficiently cause an event to occur, then that event is not out of my control. I have control over those future events that I can be causally connected to via a chain of efficient causation.[9]

Of course, the future event of my walking across the room would not be in my control if I were not in control of my own act of willing to walk across the room. The source of my control over future events involving my body and future events caused by my body is to be found in my control over my own acts of will. While my control over walking across the room is indirect (it comes via my control over my own will), at some point there must be some basic act I perform over which I have direct control. A choice – or, as we might say, an executive intention – is a plausible candidate for such an act.

[9] More exactly: I have control over those future events such that (a) I can bring them about, and (b) it is up to me whether I bring them about.

In addition to future events that I cause, then, there are present events that I have control over. I have control over my own acts of free choice, which I agent-cause. Suppose that at a time T, I freely make a choice – call this act A. So I do A freely at T. A is a basic action – it is not the case that I do A by doing something else which in turn brings about A. Rather, A occurs, it is caused by me (the substance), and the fact that A occurs is up to me. And it is up to me even as I do it, that is, it is in my control even at T. This is true despite the fact that at T, it is a fixed fact that A occurs. I do not see that a libertarian can plausibly say anything else about acts of choice.

Let's use the phrase "direct control" to refer to the sort of control I have over my acts of choice. We can then say that I have indirect control over certain events (for example, my walking across the room) that are causally consequent to those events over which I have direct control.

Next, I also have indirect control over certain events, or states of affairs,[10] that are explanatorily consequent to events I have direct control over. For example, suppose I do A at T, and I have control over my doing A at T. Let P be the proposition that I do A at T. Now consider the state of affairs consisting in P's being true. The fact that I do A at T makes the proposition P true. That is, my doing A at T is the truthmaker for P. P is true *because* I do A at T, and not the other way around.[11]

A word of explanation: I've stopped talking just in terms of 'events' and started talking in terms of 'states of affairs'. I've done this because I hesitate to call *P's being true* an event. (A proposition's being true isn't something that *happens*, it seems to me.) We need some name for it, though, so I'll use 'state of affairs' in such a way that P's being true counts as a state of affairs. I'll also count anything we can call an event as a state of affairs. So I'll use 'states of affairs' to name a broader category which includes events.[12]

---

[10] I insert "or state of affairs" in case things like its being the case that a given proposition is true do not count as events. I will stipulate that 'state of affairs' includes bona fide events as well as a proposition's being true, if a proposition's being true is not a bona fide event.

[11] I'm helped here, and throughout this section, by Trenton Merricks, "Truth and Freedom," *Philosophical Review* 118:1 (2009): 29-57.

[12] I take the general idea for this strategy from Linda Zagzebski, *The Dilemma of Freedom and Foreknowledge*, 13.

Now, when we say that the proposition P is true *because* I do A at T, the "because" here does not signify the relation of efficient causation. (I don't literally cause P to be true, in the way that I cause a billiard ball to move.) So we can't properly say that P's being true is causally consequent to my doing A at T.[13] But we can say that P's being true is explained by the fact that I do A at T. As I'll put it, the state of affairs consisting in P's being true is explanatorily consequent to my doing A at T. In general, I say that a state of affairs S2 is explanatorily consequent to a state of affairs S1 iff the obtaining of S1 explains (at least in part) why S2 obtains.

If we grant that we have libertarian freedom, then it seems clear that, in addition to certain future events and my present acts of choice, I have control over certain states of affairs that are explanatorily consequent to events over which I have control. For example, I have control over P's being true, simply because I have control over the fact that I do A at T.[14]

Let's return now to

(6) Past events are out of your control because they are fixed or settled.

I gave an argument above that (6) is false. I'm now trying to do something different than just argue that (6) is false: I'm trying to offer an explanation of why the proposition "past events are out of our control" is so attractive and intuitively plausible. And now that we have the language of causal and explanatory consequence on the table, I can state my diagnosis fairly quickly: people typically believe that past events are beyond our control because they typically believe that we do not have direct control over any events that will cause or explain a past event.

An indication that I'm right here is this: if we all suddenly believed in backwards causation (that is, if we all believed that a temporal event at a later time could causally bring about a temporal event at an earlier time), then we would be open to the possibility that a past event could be in someone's control.

[13] Not given that, in contemporary idiom, we usually mean 'efficient cause' when we speak of causal consequence.

[14] See Trenton Merrick's second corollary in "Truth and Freedom," 42.

3.1.2. A positive argument that we have counterfactual power over God's eternal beliefs

I now give a positive argument for the claim that we have counterfactual power over God's eternal beliefs. The fourth proposition of the model I am defending is:

(4) God knows events that occur in time because they occur, and not the other way around.

Suppose, again, that I do A at T. It follows from (4) that God eternally believes that I do A at T *because* I do A at T. But then if my doing A at T is in my control, the state of affairs of God's eternally believing that I do A at T should also be in my control. For if a second state of affairs depends on a first, and I have control over the first, then I have control over the second. And the state of affairs consisting in God's eternally believing that I do A at T *does* depend on the state of affairs consisting of my doing A at T. Since I have control over my doing A at T, it follows that I have control over God's eternally believing I do A at T. And, *mutatis mutandis*, the same goes for you, and everybody else with libertarian free will.

I myself find the language of explanatory consequence helpful for discussing our counterfactual power over God's eternal beliefs. I asserted above that we have control over states of affairs involving the truth of propositions about events over which we have control. So, for example, if my doing A at T is something I have control over, then P's being true is also something I have control over.[15] And I've used the language of explanatory consequence to describe the relation between my doing A at T and P's being true: P's being true is explanatorily, though not causally, consequent to my doing A at T.

Now, on the model I'm defending, God's eternal beliefs about our free actions are also explanatorily consequent to our free actions. For example, the event (E3) of God's eternally believing that I do A at T is explained by the event (E2) of God's eternally seeing that I do A at T, which is in turn explained by the event (E1) of my doing A at T. So God's

[15] P is the proposition that I do A at T.

eternally believing that I do A at T is explanatorily consequent to my doing A at T.[16]

How is it that we have counterfactual power over God's eternal beliefs? In general, we have control over events and states of affairs which are causally and/or explanatorily consequent to events over which we have direct control. In this case, we have control over God's eternal beliefs about our free actions because such beliefs are explanatorily consequent to our free actions.

3.1.3. Do we ever exercise this power? Can we?

It's worth asking whether we ever exercise our counterfactual power over God's eternal beliefs. Suppose I have the power to do A at some future time T1, but I will in fact refrain from doing A at T1. I *now* have the power to do A at T1 (let now be T0, earlier than T1).[17] Because I in fact refrain from doing A at T1, God eternally believes that I do not do A at T1. It follows that

(11) I now (at T0) have the power to do something (A at T1) such that if I were to do it, God would not have eternally believed something that He does eternally believe in the actual world.

Do I ever exercise such a power? In one sense, no; I never actually do anything such that my doing it brings it about that God does not eternally believe something which He does eternally believe in the actual world. But in another sense, yes; for the power referred to in (11) is really just my power of free action, and I exercise that frequently. What's going on here, exactly?

The situation is just the same with the power to do otherwise. For example, still supposing that I in fact refrain from doing A at T1, the libertarian can say:

(11a) I now (at T0) have the power to do something (A at T1) other than what I will do in the actual world.

[16] Given divine simplicity, E3 and E2 are not really distinct.

[17] I myself think that even at T1 I have the power to do A at T1, but here I will just work with the claim that at T0 I have the power to do A at T1.

Do I ever exercise *this* power, the power to do otherwise? In one sense, no; I never actually do anything other than what I do in the actual world. But in another sense, yes; my power to do otherwise is just my power of free action, and I exercise that frequently.

It can seem strange that I would have a power that I never exercise. Upon reflection, however, we can understand the reason why we never exercise such a power as the power to do otherwise, under that description, or the power to make God's beliefs other than what they in fact are, under that description. The reason is that the very description of the power includes a stipulation that entails that the something (which we are said to have the power to do) does not occur in the actual world. So of course we never *actually* exercise such a power by doing such a something.

Still, it would be wrong to think that this fact poses a problem for the claim that we have such a power. Ontologically, the referent of the power mentioned in claims like (11) and (11a) is just my power of free action.[18] So the power referred to in (11) and (11a) does get exercised, it simply does not get exercised to perform an action falling under the description given in (11) or (11a). But that's just to say that it does not get exercised to perform an action that does not occur in the actual world. That can't be a problem.

There is a sense, then, in which we never actually exercise our counterfactual power over God's eternal beliefs. We never exercise our power of free action by doing something that falls under the description "something that accounts for its being the case that God's eternal beliefs are different than what they are in the actual world". We never exercise the power in *that* sense or way. Next question: *could* we exercise it in that sense or way? Yes, so long as we interpret 'the actual world' in statements like (11) as a rigid designator.

To see this, return to the example used in connection with (11). Let α rigidly designate the actual world. In α, I refrain from doing A at T1, and God eternally believes that I do not do A at T1. Let W be a possible world in which I do A at T1. In W, God does not eternally believe something (viz., that I don't do A in T1) that He does eternally believe in α.

[18] Maybe my power of free action is best thought of as a set consisting of my power of free will plus various powers of bodily movement. I'll leave this complication to the side and just talk about my power of free action.

So in W, I do something that accounts for its being the case that God's eternal beliefs are different than what they are in α. Since W is a *possible* world, we can say (in α) that I *could* do something that would account for its being the case that God's eternal beliefs are different than what they are in α. Thus, I could exercise counterfactual power over God's eternal beliefs even in the sense in which I never actually do exercise it.

3.2. Counterfactual power over certain past truths

On the model I'm defending, we also have power over certain past truths. Suppose that Smith does A at T (some time in 2008). And suppose that on the day of Smith's birth, in 1958, someone uttered the following sentence: "God eternally believes that Smith does A at T". Given divine eternity, that sentence expressed a truth, way back then in 1958. So, we have:

(12) In 1958, the sentence "God eternally believes that Smith does A at T" expressed a truth.

Let the letter S denote that sentence uttered back in 1958, "God eternally believes that Smith does A at T." Notice that if Smith had refrained from doing A at T, God wouldn't have eternally believed that he does A at T, so S would not have expressed a truth, not in 1958 or at any time. And let us suppose that Smith has the power (in 2008) to refrain from doing A at T. Then Smith has the power in 2008 to do something (refrain) such that, if he were to do it, the sentence S would not have expressed a truth back in 1958. (Even though S did, in fact, express a truth back then, and 1958 is over and done with.) Call this counterfactual power over the past truth of S. It appears that the view I'm defending leaves us with the conclusion that

(13) Smith has the power (in 2008) to do something such that, if he were to do it, then the state of affairs of *S's expressing a truth* (a state of affairs that obtained in 1958 in the actual world), would not have obtained in 1958.

I concede that the Basic Model has this implication. But I deny that this is a difficulty for the view. Given divine eternity, Smith does indeed

have the power described in (13), as I'll now argue. Let's begin with a question: what *explains* the fact that S expressed a truth in 1958? Well, S expressed a truth back in 1958 because God eternally does believe that Smith does A at T. And why does God eternally believes that Smith does A at T? Because Smith actually does do A at T. So S's expressing a truth in 1958 is explanatorily consequent to God's eternally believing that Smith does A at T, which is in turn explanatorily consequent to Smith's doing A at T. But a state of affairs which is causally and/or explanatorily consequent to an event that is up to Smith is itself up to Smith. Since it is up to Smith that he does A at T, it is up to Smith that S expresses a truth back in 1958.

Or we could use the language of control: if it is in Smith's control that he does A at T, then it is in Smith's control that S expresses a truth in 1958. Or the language of "being in one's reach": if the action of doing A at T is within Smith's reach, then the state of affairs of S's having expressed a truth in 1958 is within Smith's reach,[19] because the second depends entirely on the first, and the first is within Smith's reach.[20]

If you still find yourself skeptical of the conclusion that Smith has counterfactual power over the past truth of propositions about his free actions, consider the following three points. First, the power in question is not a power to change the past. Smith's having the power described in (13) does imply that something about the past is now up to him (the 'something' is the state of affairs, obtaining in 1958, of S's expressing a truth). But it does not imply that he can change the past, for a reason parallel to that given above in the discussion of counterfactual power over God's eternal beliefs. Changing the past would involve bringing it about that the past is now one way, after earlier having been a different way. But (13) does not imply that there is any possible world in which Smith brings that about.

Second, this is a power no one ever actually exercises. (Although it could be exercised.)

Third, so long as it is granted that future contingent propositions have a truth value, the same puzzle is generated simply by the problem of logical

[19] Within his reach to account for, not within his reach to change. See below, next paragraph.

[20] I'm helped here by Trenton Merricks, "Truth and Freedom," 42.

fatalism, quite apart from theological considerations. So if we think there is an answer to logical fatalism, we should not be troubled by (13).

3.3. Counterfactual power over genuine past events

And now for a more difficult objection. . .I have just argued that, given the Basic Model, the past truth of certain sentences is even now up to us. Some might not find this objectionable on the grounds that a state of affairs like S's expressing a truth in 1958 is not something that really happened or occurred in 1958. It's not a genuine past event, like Napoleon's having been defeated at Waterloo.[21] The point is that the past truth of a sentence about the future is a strange sort of "past" state of affairs, so maybe it's not so strange to think that it can even now be up to us. But what about *genuine* past events, can it be that we have counterfactual power over them?

On the Basic Model, God could give a human being this power, and it is epistemically possible that on occasion He has. Consider the following scenario: at a time T3 Smith does A. Because Smith does this, God eternally knows that Smith does A at T3. Because God knows this, God reveals to a prophet at T1, a much earlier time, that Smith will do A at T3, and God instructs the prophet to write this information down on a piece of parchment. The prophet does so, say, at T2 (also much earlier than T3). Then the event of the prophet's inscribing the information at T2 is explanatorily consequent to Smith's doing A at T3 – although it is temporally before Smith's doing A at T3. If Smith were to refrain from doing A at T3, God would not eternally believe that Smith does A at T3, and so God would not have revealed what He did to the prophet, and so the prophet would not have inscribed the parchment. If God were to actually do something like this, I think the thing to say would be that

(14) Smith has the power (long after T2) to do something such that, if he were to do it, then an event that occurred in the actual past (the prophet's inscribing the parchment in the way that he did at T2) would not have occurred.

This would be a counterfactual power over a genuine past event.

[21] See Linda Zagzebski, *The Dilemma of Freedom and Foreknowledge*, 26.

Does God ever do this sort of thing? It's hard to know. But on the Basic Model it is epistemically possible that He does. So it's epistemically possible, according to the Basic Model, that we have counterfactual power over genuine past events. (Although no one ever exercises it, because God gets things right the first time around, so to speak.)

What I think these considerations show is that, given divine eternity, we might have something like a causal connection to events that are past. And this connection is sufficient for giving us control over such events. (Not control in a sense which involves the power to change such events, but control in the sense that what we do now might account for some past event's having occurred.)

To elaborate: given divine eternity, there could be something like backwards-in-time causation. What we might call 'standard' backwards causation would involve a later temporal event having an efficient causal effect on an earlier temporal event. The sort of backward causation relevant to our discussion involves eternity as well as time, and involves explanatory dependence as well as efficient causation. To use our previous example:

(a) Smith does A at T3.
(b) Because Smith does so, God eternally believes that he does so. (The 'because' here signifies a relation of explanatory dependence, not a relation of efficient causation.)
(c) In part because God believes that Smith does A at T3, God wills to reveal that Smith does A at T3 to a prophet at T1. (Again, this is explanatory dependence, not efficient causation.)
(d) Because God so wills, the prophet receives the revelation at T1. (Here we have efficient causation.)
(e) Because the prophet receives the revelation, he records the prophecy at T2.

On this story, an event at an earlier time (the recording of the prophecy) is dependent for its occurrence on an event at a later time (Smith's doing A). It is not correct to say that the later event has an efficient causal effect on the earlier event (because of the routing through God), but it's clear that the earlier event depends on the later. Let's call this backwards-in-time explanation. While the Basic Model does leave us with the coun-

ter-intuitive claim that humans could have counterfactual power over past events, it can also give a coherent account of *how* a person could manage to have such power. On the Basic Model, a person could have such a power because an act in that person's power to do or not to do could backwards-in-time explain a past event, given the requisite divine action.

3.4. An objection relating to backwards-in-time explanation

I now consider an objection to the view I'm defending that has to do with the possibility of circular explanation. Let's return to the thought experiment involving the discovery of an ancient manuscript containing the story of your life. One might object to the Basic Model as follows. If the Basic Model were true, then God could cause such a manuscript to be written, 1300 years before your birth. But if such a manuscript were written, it is possible that you might find it and read it, including the portion about your future. But once you read what your future actions were in the manuscript, you could choose to act differently in various ways. And if you could do that, you could make God wrong. But that is impossible. By *reductio*, the Basic Model must be false.

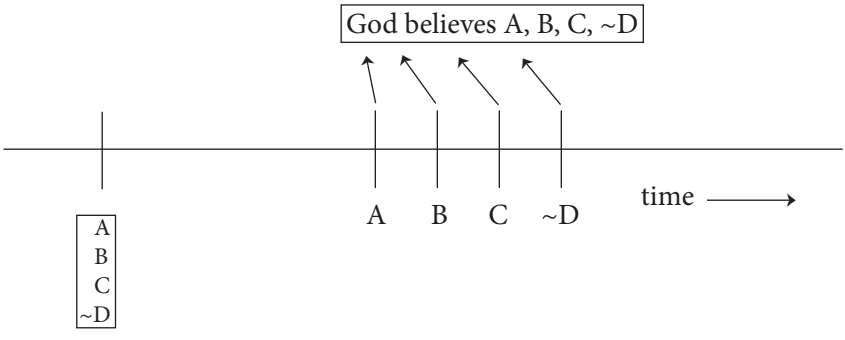
Assessing this objection requires us to take a close look at

(15) If such a manuscript were written, it is possible that you might find it and read it, including the portion about your future.

I'll now argue that (15) is false. For (15) to be true is for there to be a possible world in which God writes the story of your life and then permits you to find and read it (including the part about your future). But when we think carefully about what it would take for God to write such a book, we'll see that there is no such possible world. To put things loosely: God could not write such a story (the story of your life in which you found and read the completed story) because He would not know what to write.

To see this, note that although there are no temporally distinct steps an eternal God would have to go through to write a person's life-story, there is an order of logically distinct steps that would have to be followed. (Or, better: there is an order of logical dependency that would have to be

respected.) For purposes of illustration, let's focus on four choices you'll make, in 2008, 2009, 2010, and 2011. In 2008, you chose between A and ~A, in 2009 between B and ~B, and so forth. In this diagram, the arrows represent relations of explanatory dependence.

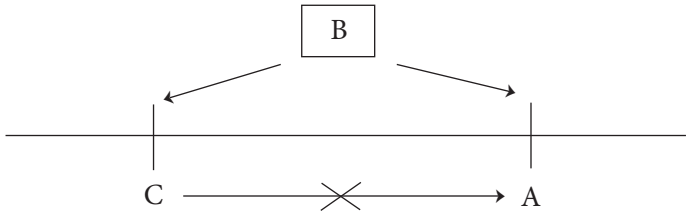


Because you choose A, God eternally knows that you choose A. Because you choose B, God eternally knows that you choose B. And so forth with C and ~D. In part because God knows all this, He has the book be written down at some point in time which is long past relative to the time you choose A. I assert that, so long as you never read a portion of the book describing your future, the situation represented in this diagram is metaphysically possible.

However, no story of your life could be written in which, midway through the story, you read the whole story. For in order for God to write the whole story of such a life (a life in which you find and read the whole story), God would first (not first in a temporal ordering, but first in a logical or explanatory ordering) have to see what you do in the course of your whole life, including the part of your life which follows your reading of the whole story. And in order for God to see what you do in the parts of your life that follow your reading of the whole story, you first (in the explanatory order) have to live the parts of your life that follow your reading of the whole story. And in order for you to live the parts of your life that follow your reading of the whole story, you first (in the explanatory and temporal orders) have to read the whole story. But in order for you to read the whole story, it first has to be written. So in order for God to write the whole story, it first has to be written. Since it is logically im-

possible that such a requirement be fulfilled, it is logically impossible for there to be a possible world in which God writes the whole story of your life and permits you to read the whole thing. So (15) is false.

We can, I think, generalize from this example to formulate a principle about backwards explanation. If a human action A explains (at least in part) why (B) God believes something, and God's believing that something in turn explains (at least in part) why God causes an earlier event C to occur, then C cannot explain (even in part) the occurrence of A.



Otherwise, we would have an explanatory loop of the form: A occurred in part because C occurred, and C occurred in part because B occurred, and B occurred because A occurred. So A occurred in part because A occurred. This, it seems to me, is a metaphysically impossible sort of self-explanation. I think the thing to conclude is that an eternal God can act in such a way that a human action backwards-explains an earlier event, but He cannot act in any way that implies the existence of an explanatory loop.

We can apply this to the case of Christ's prophecy to Peter that Peter would deny him:

> "Simon, Simon, listen! Satan has demanded to sift all of you like wheat, but I have prayed for you that your own faith may not fail; and you, when once you have turned back, strengthen your brothers." And he [Peter] said to him, "Lord, I am ready to go with you to prison and to death!" Jesus said, "I tell you, Peter, the cock will not crow this day, until you have denied three times that you know me."[22]

If Peter's hearing of this prophecy were an explanatory (or causal) factor in his decision to deny Christ, we would have an unacceptable explana-

[22] Luke 22:31-34 (New Revised Standard Version). See also Matthew 26:34, Mark 14:30, and John 13:38.

tory loop: Peter denied Christ at T in part because Christ predicted that he would, but Christ predicted that he would because Peter denied Christ at T. Suppose, however, that God made Peter forget about the prophecy until after the denial. If God isolated the causal effect of the prophecy in the right way, then, it seems, there would be no explanatory loops involved.[23] It's interesting to note that the gospels suggest that Peter had indeed forgotten about the prophecy at the time of his denial (see Luke 22:60-61).

CONCLUSION

In this paper I've presented a solution to the problem of freedom and foreknowledge. God is outside of time and has knowledge of events that are future with respect to us, simply because he "sees" them occur. I've argued that this solution does not falsely saddle us with a power to change the past or to change God's beliefs. On the model of divine eternity I've presented, however, we do have counterfactual power over God's eternal beliefs, and over the past truth of certain propositions. What is more, a person can even have, if God has made prophecies of a certain sort, counterfactual power over genuine past events. But these aren't good reasons to reject the eternity solution – they are simply consequences of the fact that backwards-in-time-explanation is possible, if God is eternal.[24]

[23] Alexander Pruss discusses this same case, and comes to the same conclusion, in an entry on Prosblogion, available at http://prosblogion.ektopos.com/archives/2009/08/prophecy.html.

[24] I'm grateful to Timothy Pawl, Faith Pawl, Matthews Grant, Thomas Sullivan, and the participants in the *God and the Future* conference at Humboldt University in Berlin for their help in thinking through the issues discussed in this paper.

BIBLIOGRAPHY

Hasker, William. "The Absence of a Timeless God." In *God and Time: Essays on the Divine Nature*, edited by Gregory E. Ganssle and David M. Woodruff, 182-206. Oxford: Oxford University Press, 2002.

Merricks, Trenton. "Truth and Freedom." *The Philosophical Review* 118:1 (2009): 29-57.

Rogers, Katherin. "The Necessity of the Present and Anselm's Eternalist Response to the Problem of Theological Fatalism." *Religious Studies* 43:1 (2007): 25-47.

Unger, Peter. "Free Will and Scientiphicalism." *Philosophy and Phenomenological Research* 65 (2002): 1-25.

Zagzebski, Linda. *The Dilemma of Freedom and Foreknowledge*. New York and Oxford: Oxford University Press, 1991.

Zagzebski, Linda. "Foreknowledge and Free Will." In *The Stanford Encyclopedia of Philosophy (Fall 2008 Edition)*, edited by Edward N. Zalta, http://plato.stanford.edu/archives/fall2008/entries/free-will-foreknowledge/.

ON GOD, SUFFERING, AND THEODICAL INDIVIDUALISM

JEROME GELLMAN

Ben Gurion University of the Negev

Recently, Stephen Maitzen has provided an argument for the non-existence of God based on ordinary morality. Here is Maitzen's argument, abbreviated down to the relevant parts for my reply and slightly reformulated for ease of presentation:

(1) Necessarily, God permits undeserved, involuntary, human suffering only if such suffering ultimately produces a net benefit for the sufferer. (Maitzen calls this "TI")
(2) If God exists, then necessarily, all undeserved, involuntary human suffering ultimately produces a net benefit for the sufferer. (From 1)
(3) If**,** necessarily, all undeserved, involuntary human suffering ultimately produces a net benefit for the sufferer, then we never have a moral obligation to prevent undeserved, involuntary human suffering.
(4) We sometimes have a moral obligation to prevent undeserved, involuntary human suffering.
(5) So, it isn't the case that, necessarily, all undeserved, involuntary human suffering ultimately produces a net benefit for the sufferer. (From 3 and 4)
(6) So: God does not exist. (From 2 and 5)

I will assume the truth of (1), and thus of (2) which follows from (1). My problem is with (3). Maitzen must do much more to show that (3) is true, or even very plausible. My argument is that (a) the set of propositions I am about to list is not implausible, granting God's existence and perfect goodness and a proper understanding of ordinary morality, and

(b) since that is so, the consequent of (3) does not plausibly follow from (3)'s antecedent.

Let P be any person that I allow to endure suffering, instead of preventing the suffering. And let S be the particular instance of suffering that P undergoes. And consider the following set of five propositions, the first four of which are about God and the fifth a moral principle:

(G1) God has so created the world that God is able to produce for P a degree, D, of good for P that will make P's existence worthwhile.
(G2) God will bring about D for P. This will not be in the form of compensation but a result produced by P's history.
(G3) D is a maximal degree of good that God can produce for P.
(G4) God will bring about D for P whether P suffers (this particular) S or not. If P suffers S, then God will produce a net good from S for P, to offset the evil of S in order to reach D.
(G5) One can morally allow P to suffer S, if and only if the net benefit to P from allowing S will far outweigh S, and either: (a) the net benefit to P from allowing S will be far greater than the good that will be P's if one prevents S, or (b) the net benefit to P from one's allowing S will not be less than the good that will be P's if one prevents S, and allowing S will significantly increase the net good in the world.

I now explain each of the members of (G1)-(G5):

(G1) and (G2) are consequences of theodical individualism, according to which God must see to it that each person's life be worthwhile, and not just that the existence of the world at large be worthwhile. I include in P's existence P's life after death. I envision the degree of good that is P's in the afterlife as produced by earlier events and not as compensation. I also do not disallow a view like John Hick's wherein we continue to grow and change in the afterlife with new degrees of goodness accruing to our existence.

(G3) follows from God's supreme goodness. There will be constraints upon what the maximal degree of goodness is that God can produce for P. For example, God will want the best distribution of good overall in the world and will want a world that is good overall. So, given such con-

straints, God will produce the best possible degree of good for P. P will have no claims against God that God did not produce a higher degree of good for her, because her existence is the best it could be without fulfilling her selfish desires to get more for herself on other peoples' accounts.

(G4) reflects the fact that God is so damned smart that God has been able to create the world so that whether P suffers S or not, P will receive D. And it reflects God's goodness, in that God would not allow S to occur if it were to detract from the ultimate net result, D, for P.

(G5) makes a claim about our ordinary morality. It says that while consequentialist reckoning is relevant to the allowing of undeserved, involuntary, human suffering, the morality of allowing such suffering is not purely consequentialist. That is because we have a moral *deontological* obligation to prevent undeserved, involuntary, human suffering (when able, when appropriately situated, etc.). And while that obligation is defeasible, it is *not* overridden by a mere net benefit for the sufferer.

What is needed to defeat the deontological obligation not to allow suffering is that either: (a) the net benefit far outweighs the suffering and far outweighs what good would accrue to the sufferer without that suffering; or (b) the net benefit that far outweighs the suffering is at least the same as what good would accrue to the sufferer without that suffering, and produces a significant net increase of good in the world. These are necessary and sufficient conditions for one being allowed not to prevent undeserved, involuntary, human suffering.

On alternative (a), the benefit justifying one's not preventing the suffering is all P's. On alternative (b), however, the benefit that justifies not preventing the suffering is not P's but others'. This can happen in at least two ways. The first is that one might be able to bring a very high benefit to another person by allowing P to suffer. And the second is that in adjudicating one's moral obligations to others as well as to P it turns out that one's obligations to others are stronger than, and so defeat, one's obligation to prevent P from enduring S.

You might protest that (b) violates Maitzen's Kantian stricture, let's call it "K."

(K) Nobody (including God) can treat human beings merely as a means.

The violation appears to occur because (G5) permits us (and God) to allow P to suffer in order to benefit others.

However, (G5) does not violate (K), on Maitzen's interpretation of it, which I accept. For Maitzen, (K) prohibits "*sacrificing* an innocent person who did not ask for it" (p. 116, my emphasis), and (K) prohibits, "exploiting" a person by visiting upon them undeserved, involuntary, human suffering (p. 117). However, (G5) licenses none of these. Remember, whenever (G5) allows the non-prevention of undeserved, involuntary, human suffering the sufferer is either better off or at least as well off as she would have been without having endured that suffering. The only thing that has happened is that the deontological moral obligation that others have not to allow P to endure undeserved, involuntary, suffering has been overridden by either a justified benefit to P or to others. This does not constitute a "sacrifice" or "exploitation" of P in any way, and should not be forbidden by (K). To put it tersely, (G5) does not excuse any instance in which a person treats P *merely* as a means.

It is not implausible to assume that the set (G1)-(G5) is true, granting God's existence and perfect goodness and our understanding of ordinary morality. But, then the antecedent of (3):

(3a) Necessarily, all undeserved, involuntary human suffering ultimately produces a net benefit for the sufferer.

does not plausibly imply (3)'s consequent:

(3c) We never have a moral obligation to prevent undeserved, involuntary human suffering.

That my allowing P's undeserved, involuntary suffering will result in a net benefit for P is, by (G5), not *sufficient reason* for me to not have a moral obligation to prevent P's suffering. More is necessary. So, (3) is not plausible. Even if all involuntary human suffering ultimately were to produce a net benefit for the sufferer, there would still be instances in which I had a moral obligation to prevent it. Such would be an instance, for example, where the benefit to the sufferer is only slightly more than preventing the suffering would yield.

Moreover, according to (G1)-(G4), it will *never* be the case that we may allow an instance of undeserved, involuntary suffering on the grounds

that it will produce a *far* greater benefit for the person than would preventing the suffering. And that is because in *every* instance, according to (G1)-(G4), God will in any case bring about the maximum possible degree of good for the person during his existence – whether he endures that particular suffering or not. So the clause (a) of (G5) will never apply to any of us when faced with the opportunity to prevent suffering. For my money, theodical individualism should entail that we always have an obligation to prevent a case of suffering in P, unless the conditions of (G5) are met. In practice, this means that in the great majority of situations, my theodical individualism will obligate preventing suffering. In any case, since (G1)-(G5) is not implausible, (3) is not yet plausible.

Well, now, you ask, what justifies *God* allowing undeserved, involuntary suffering? After all, (G5) obligates God no less than it does mere mortals. And since God is a perfect being God will faithfully fulfill the dictate of (G5).

On (G1)-(G4), it follows that the suffering God allows a person to endure is such that God allows it either: (i) so as to produce for that person a degree of good that is maximal for that person and also satisfies the conditions of (G5), or: (ii) because allowing it follows from God's juggling of God's obligations to all human persons. Since God is the creator of all persons and sustains them in life at every moment, all human beings stand before God with an equal claim on God to produce for them a maximal possible ultimate benefit in their existence. All human beings are turned to God with an equal claim to a personal maximal net benefit, *everything else being equal*. Thus, God's obligations to others besides P are of vastly greater dimensions than that of any human being to other persons. That is at least one reason why God might be justified in allowing suffering in the world, in vastly many cases, while humans have relatively sparse permission to do so. I have already argued that in light of (G1)-(G5) such thinking does not violate the above Kantian stricture about using people as mere means.

One might be skeptical as to whether the world we live in is one where the God of (G1)-(G4) exists. One might doubt, for example, whether every person receives an ultimate benefit that makes his or her existence worthwhile. Such doubts belong to the problem of evil and must be discussed on their own. They do not relate to the present argument against God's existence from theodical individualism. That argument, I have tried to show, fails to convince.

ON GELLMAN'S ATTEMPTED RESCUE

STEPHEN MAITZEN

Acadia University

I thank Jerome Gellman for his critique of my recent article "Ordinary Morality Implies Atheism."[1] In that article, I argued that traditional theism threatens ordinary morality by relieving us of any moral obligation to prevent horrific suffering by innocent people even when we easily can. Gellman attempts to rescue that moral obligation from my charge that theism destroys it. I believe his attempted rescue fails.

Gellman begins by quoting the main principle on which my argument depends, the principle sometimes called "theodical individualism" that I abbreviated with the initials "TI":

> (TI) Necessarily, God permits undeserved, involuntary human suffering only if such suffering ultimately produces a net benefit for the sufferer.

He then correctly notes that theism—the proposition that God exists—and TI together imply (to use Gellman's numbering)

> (3a) Necessarily, all undeserved, involuntary human suffering ultimately produces a net benefit for the sufferer.

However, Gellman gives reasons to *reject* the claim in my argument that (3a) implies (again, his numbering)

> (3c) We never have a moral obligation to prevent undeserved, involuntary human suffering.

[1] Jerome Gellman, "On God, Suffering, and Theodical Individualism," *European Journal for Philosophy of Religion* 2:1 (2010): 187-191, replying to Stephen Maitzen, "Ordinary Morality Implies Atheism," *European Journal for Philosophy of Religion* 1:2 (2009): 107–26. Further references to these works will cite author and page number only.

That is, he claims to preserve our ordinary moral obligation even if theism and TI are true.

I'll argue that the reasons Gellman gives fail to answer my challenge, because they're implausible reasons on two independent grounds. First, his reasons confuse cases in which God's permission of suffering is *justified* by some benefit to the sufferer (cases that therefore are relevant to TI) with cases in which God confers a benefit in an attempt to *compensate* the sufferer (cases that therefore aren't relevant to TI). Second, his critique assumes that God might face constraints that there's no good reason to think a perfect being could face.

COMPENSATION VS. JUSTIFICATION (AGAIN)

Gellman writes, "Let P be any person that I allow to endure suffering, instead of preventing the suffering. And let S be the particular instance of suffering that P undergoes." His critique of my argument then asserts that the following propositions are "not implausible, granting God's existence and perfect goodness":

(G1) God has so created the world that God is able to produce for P a degree, D, of good for P that will make P's existence worthwhile.
(G2) God will bring about D for P. This will not be in the form of compensation but a result produced by P's history.
(G3) D is a maximal degree of good that God can produce for P.
(G4) God will bring about D for P whether P suffers (this particular) S or not. If P suffers S, then God will produce a net good from S for P, to offset the evil of S in order to reach D.[2]

Whether or not Gellman's propositions are plausible, at least two of them are simply irrelevant to my argument. In my article, I took pains to distinguish justification from mere compensation, and in doing so I quoted Christian philosopher Eleonore Stump's version of the idea that TI tries to capture: "if a good God allows evil," says Stump, "it can only be

[2] Gellman, 188.

because the evil in question produces a benefit for the sufferer and one that God could not produce without the suffering."[3] I then wrote,

> Like Stump's use of it, TI's use of the word "produces" is significant, because otherwise we allow that God's mere compensation of the sufferer—say, in a blissful afterlife—can justify God's permission of suffering even if the suffering bears no necessary connection to the good that compensates for it. Without such a connection, the good may *compensate for* the suffering but can't *morally justify* God's permission of it.[4]

Gellman's proposition G2 gestures at the difference between compensation and justification. However, I find G2's claim that God brings about D "as a result produced by P's history" mysterious. If D is a result *produced by P's history*, how is it that *God brings about* D? Is it because God brings about P's history? If so, then God exercises a degree of "meticulous providence" that many theologians would reject. In the end, however, I can only guess at the meaning of G2.

More important, neither G2 nor G4 ever says or even hints that S is necessary for, or the best means of, God's securing D for P. As Stump seems to recognize, when we ask whether someone's permission of suffering is justified by its relationship to some benefit for the sufferer, we ask (in part) whether the permission is necessary, or if not strictly necessary then optimal, for securing the benefit.[5] Nothing in G2 or G4 requires or even suggests an affirmative answer to that question in the case of God, P, S and D. On the contrary, G4 says that God will secure D for

---

[3] Eleonore Stump, "The Problem of Evil," *Faith and Philosophy* 2 (1985): 392–423; 411–12.

[4] Maitzen, 110, emphasis in original. I continued: "Consider an analogy to our ordinary moral practice. My paying you money after harming you may compensate for my harming you, but it doesn't justify my harming you. Only something like the necessity of my harming you in order to prevent your harming me or an innocent third party has a chance of justifying my behavior: some necessary connection must hold between the harm and the benefit."

[5] See also Maitzen, 115, again quoting Stump; the fuller quotation from Stump reads, "other things being equal, it seems morally permissible to allow someone to suffer involuntarily only in case doing so is a necessary means or the best possible means in the circumstances to keep the sufferer from incurring even greater harm" ("Providence and the Problem of Evil," in *Christian Philosophy*, ed. Thomas P. Flint [Notre Dame, Indiana: University of Notre Dame Press, 1990], 51–91; 66).

P *regardless* of whether God permits S to occur, in which case it's hard to see how D plays any role in justifying God's permission of S in particular. In light of this problem, it won't do simply to *stipulate*, as G2 does, that God's bringing about D for P isn't merely compensation for S, especially when G4 encourages the suspicion that it *is* merely compensation. G2 and G4 are, so far as I can see, simply irrelevant to whether God is justified in permitting S and hence irrelevant to the implications of TI.

A HAMSTRUNG GOD?

The final proposition that Gellman offers as "not implausible" is quite a mouthful even though he describes it as "a claim about our ordinary morality":

> (G5) One can morally allow P to suffer S if and only if the net benefit to P from allowing S will far outweigh S, and either: (a) the net benefit to P from allowing S will be far greater than the good that will be P's if one prevents S, or (b) the net benefit to P from one's allowing S will not be less than the good that will be P's if one prevents S, and allowing S will significantly increase the net good in the world.

While one might doubt that something as complex as G5 does in fact govern our ordinary moral reasoning, let's grant that it does. Gellman says that G5 preserves the moral obligation that I claim theism and TI jointly threaten, because according to G5 we can be obligated to prevent S even if God's permission of S satisfies TI's requirement of a net benefit to P. For, unlike TI, G5 requires that the benefit to P not just outweigh but *far* outweigh S, and G5 imposes a further, disjunctive condition on top of that.

If G5 is so plausible, does it also bind God? Gellman says it does: "G5 obligates God no less than it does mere mortals. And since God is a perfect being God will faithfully fulfill the dictate of G5."[6] In that case, it would certainly appear, God is justified in permitting only the suffering

[6] Gellman, 191.

that G5 justifies *us* in permitting. If so, however, then we don't have the moral obligation that Gellman wants to rescue. Because God is perfect in knowledge, power, and goodness, *any* suffering that occurs is suffering that God permits (else he lacks perfect power or perfect knowledge) and is justified in permitting (else he lacks perfect goodness). So if we're justified in permitting S whenever God is justified in permitting S, we're always justified in permitting S. There goes morality.

Remarkably enough, Gellman tries to avoid this unwanted implication by having God plead other commitments. According to Gellman, God is stuck with

> juggling God's obligations to all human persons.... All human beings are turned to God with an equal claim to a personal maximal net benefit, *everything else being equal*. Thus, God's obligations to others besides P are of vastly greater dimensions than that of any human being to other persons. [Hence] God might be justified in allowing suffering. . .in vastly many cases, while humans have relatively sparse permission to do so.[7]

Thus Gellman retracts his earlier assertion that G5 obligates God and that God "will faithfully fulfill the dictate of G5," for he allows God to opt out of G5 if other commitments prevent God from honoring it.[8]

I can't see how a perfect being could be hamstrung in the way Gellman suggests. In Gellman's story, God isn't obligated to prevent S because God has commitments to others besides P, commitments that force God to allow S (or else somehow justify God's allowing S even though they don't force it). Nevertheless, he says, *we* can be obligated where God isn't, because we don't have God's excuse: unlike God, we don't have an equal obligation toward every human being. So God sometimes needs us to prevent suffering by P that God himself can't prevent without failing in his duty to someone other than P. I can't how see a perfect being could find itself in such a pickle. I often need a little help from my friends: "Can

[7] Gellman, 191, emphasis in original.

[8] Gellman writes that "the suffering God allows a person to endure is such that God allows it either: (i) so as to produce for that person a degree of good that is maximal for that person and also satisfies the conditions of G5, or: (ii) because allowing it follows from God's juggling God's obligations to all human persons" (191). I interpret (ii) as the "out" clause Gellman offers God, because (ii), unlike (i), makes no mention of God's satisfying G5.

you pick up Jack? I promised him a ride, but I also promised to pick up Jill elsewhere at the same time, and I can't be in two places at once." But Gellman claims to be describing the supremely independent God of classical monotheism, not a limited being like me.

Gellman faces a dilemma: Does God honor his obligation to P under G5 or not? If yes, then allowing S confers a benefit to P that far outweighs S (and achieves other good things too), in which case, as I argued in my article, we have no obligation to prevent S. If no, then *why* doesn't God honor his obligation to P? Presumably because God can't honor it and also honor his obligations to others besides P, which is exactly the situation I face in my example. It doesn't matter *how* God ends up in my predicament, since it's implausible enough to think that God *could* end up in my predicament.

Gellman might reply that we can't *know* that a perfect God isn't hamstrung by constraints that force God to permit S when he'd rather prevent it. But surely the burden of proof rests with whoever suggests that such limitations apply to a perfect God. In any case, I hope nobody offers the hoary suggestion that God's constraints include a commitment never to interfere with our libertarian free will, a suggestion I criticized in my article because it contradicts the biblical account of God's conduct, overstates the value we actually attach to free will, and otherwise fails to jibe with ordinary morality.

MORALITY, GOD, AND POSSIBLE WORLDS
A paper inspired by Richard Swinburne's *God and Morality*

JACEK WOJTYSIAK

The John Paul II Catholic University of Lublin

Let us imagine two possible worlds, similar to ours. They are both the same in all non-moral respects with the exception of one component: in the first world God exists and is the creator of all that exists (apart from Him) in this world; in the second world, God does not exist. Do the two worlds – the world-with-God and the world-without-God (the Godless world) – differ as to their moral contents? Richard Swinburne (if I understand him correctly) says yes[1]. According to him, the "moral" differences between the two worlds are considerable, although they do not concern the essence of morality itself. Both worlds are equally "moral" in the sense that there are moral properties in them; these properties are in a sense supervenient on non-moral properties and independent of the existence and activity of God. The world-with-God, however, is morally "richer" than the world-without-God, but only in the sense that:

(I) The existence of God – the creator and benefactor – entails some additional and specific obligations for people: they are obliged to obey and worship Him. These obligations are analogous to our obligations towards our parents, lawful and just rulers, or persons (benefactors) that deserve special respect.

[1] R. Swinburne, "God and Morality", *Think* 20, vol. 7 (Winter 2008), pp. 7-15 (cf. also R. Swinburne, "What Difference Does God Make to Morality?", [in:] R.K. Garcia, N.L. King, *Is Goodness without God Good Enough? A Debate on Faith, Secularism, and Ethics,* Lanham: Rowman and Littlefield Publishers 2008, pp. 151-163). I present Swinburne's theses by means of a conceptual apparatus somewhat different than his. In this text I use the terminology of possible worlds. Swinburne has used this terminology only twice. I distance myself here from the problem of the ontological and methodological status of possible worlds. I treat them only as hypothetical equivalents of certain global possibilities (of all that exists).

(II) God may affect moral "facts" through creating (or not creating) states of affairs on which (by necessity and independently of His will) particular moral properties are supervenient. To use an example other than Swinburne's: if ammunition, shot at people, did not of its nature cause death or wounds, but pleasure, then shooting at innocent or unarmed people would, in principle, be not a bad act (or not as bad as it is in our actual world).

(III) God, as the highest epistemic and deontic authority, may (especially by way of revelation) teach us about our obligations or enhance our motivation to implement them. He may also formulate (within limits) some additional commands that help people to accomplish their (most fundamental) good.

Let us assume that we agree with Swinburne with regard to the above "moral differences" between the worlds under consideration[2]. Contrary to Swinburne, I think that these differences entail differences as to the nature of morality: morality in the world-with-God is indeed something different than morality in the world-without-God. In order to realise this let us have a closer look at the differences (I)-(III).

OBLIGATIONS TOWARDS GOD

If there is a God – a rational creator and benefactor – then we not only have some specific obligations towards Him, but all our obligations are (at least indirectly) obligations towards Him. Why?

In the world-with-God everything that exists (apart from Him) exists thanks to Him, is the result of His conscious, free and gratuitous act of

[2] I think that the most debatable thing in Swinburne's reasoning is that he treats moral properties as logical consequences of non-moral properties. It is not clear how the second (by virtue of logical or quasi-logical necessity) should determine the first. I omit here this question, although some of my remarks may be treated as certain suggestions towards solving this problem. Let me add something which seems obvious, namely that both Swinburne and I presuppose that in both possible worlds – with God and without God – there are people who are free and responsible moral subjects. Thereby we presuppose that neither God's omniscience, and omnipotence, (appropriately understood) nor natural laws (appropriately understood) exclude man's real freedom. The argumentation on behalf of this presupposition, and discussion of attendant problems, go beyond the framework of this text.

creation. In this world all our obligations – both natural and contractual – towards whoever (or whatever) are therefore obligations towards God. Persons or things towards which we have obligations exist and bear obligation-making properties just because of God's creative act. We owe honour to our parents since they are the most proximate source of our existence. However, although we have received our existence by the mediation of certain created beings, ultimately, we have received our existence, and our parents, only from God. He is therefore (as Swinburne writes) 'much more the source of our being than are our parents'[3]. As we can see, in the theistic perspective gratitude, honour and obedience to parents are, ultimately, (though not directly) gratitude, honour, and obedience to God Himself, as the ultimate source of our parents. In like manner, we can say about all beings created by God, towards which we have obligations: these obligations are indirectly obligations towards God. The relationship of being obligated, like the relationship of existential derivation (or of being caused), seems to be a transitive relationship.

It follows from the above that all moral obligations in the world-with-God reach much further than obligations in the world-without-God. In the first world our obligations are always also obligations towards the being on Whom we are ontically dependent and Who is fully (insofar as this is logically possible) rational and good. In the second world our obligations are always obligations exclusively towards beings limited with respect to their ontological status, rationality, and goodness. In this sense, morality in the world-with-God is absolute, and morality in the world-without-God is relative. We may discuss our duties towards beings to whom we owe only a little, and who are only partially rational and good. However, it is difficult to do this in relation to the being to Whom we owe everything and Who is entirely rational and good.

GOD'S INFLUENCE ON MORALITY

According to Swinburne, in the world-with-God 'God brings about the circumstances which (in virtue of some necessary moral truth) make

[3] Swinburne, *God and Morality*, p. 11.

an action of some kind good or bad'[4]. This thesis is illustrated by Swinburne's example of capital punishment and my example of shooting (see above). I think that God's creative prerogatives, in relation to morality in the world-with-God, have much more important consequences. Let us note that:

(a) if God did not create any beings – or non-moral properties – there would not exist any moral properties supervenient on those beings or non-moral properties;
(b) if God created only non-personal beings, the moral properties supervenient on them would exist at most potentially, being made actual only in the moment when persons come into existence;
(c) if God created a world radically different from ours with respect to non-moral contents, it would differ from our world also with respect to moral contents: the lack of certain beings (or non-moral properties) would result in the lack of certain moral properties, and the existence of other beings, or non-moral properties (unknown to us), would entail the existence of other moral properties (somewhat new to us).

I am not going to speculate here how far our world could be different (in its non-moral and moral respects). It suffices to note that once we accept (a)-(c), we are brought to the conclusion that in the world-with-God, God's decisions are (at least) a necessary condition of the existence of morality (see the points a-b) and a necessary condition of the moral contents (see point c). Such being the case, the metaphorical (analogous) statement that God is the "author" or "giver" of morality turns out to be obviously justified (for those inhabitants of the world-with-God who are aware of this fact). One may even say: God, in creating beings that found morality, is (indirectly) the creator of morality; by deciding *what* exists and *how* it exists, God (indirectly) decides *what* is good or bad and *how* it is good or bad. The fact that sense qualities supervene on light or sound waves etc., and aesthetic qualities supervene on sense qualities, does not undermine the theist's belief that God is the creator of sense and

[4] Ibid., p. 10.

aesthetic qualities, and that they were consciously projected or designed by Him. For this same reason the consequent theist should recognise that morality (the set of non-moral properties and their resultant moral properties or normative propositions) is God's conscious work[5]. Now the consequent atheist should recognise that morality is a "brute fact" or a fact that is supervenient on the "brute fact" of the contingent existence of our actual world. Obviously, the atheist may also say that morality supervenes on the non-moral contents of our world, which exists and is such as it is of necessity, or thanks to rational "self-regulation." This position seems, however, fairly peculiar and it is an example of pantheism rather than atheism.

ADDITIONAL COMMANDS

If God can indirectly create morality – (i) as the ultimate correlate of moral obligations and (ii) as the creator of (this and not another) "ontological foundation" of morality – then why could He not make it directly? The absolute sovereignty of God the Creator entails His ontic power and deontic right to directly determine such moral obligations that are not (only) the consequence of His existence or the consequence of the existence of the non-moral contents of the world created by Him.

[5] God is absolutely free, but within the confines of his own – rational and good – nature (the scholastics sought to describe this by means of the first principles of being, knowledge, and morality). Once we assume this thesis we may – following Swinburne's approach – solve the Euthyphro dilemma: part of our duties results from God's decision, and part of them results from the fact that they are good as such. (I am leaving aside here the quantitative and qualitative definition of these "parts"). This does not lead to antitheistic consequences, if we bear in mind that any free decisions on God's part are decisions of a being with a concrete – rational and good – nature, and this nature of God, that determines good (and indirectly evil), is something internal to Him. As we can see, God does not have to be redundant in moral explanation, and the morality established by Him does not have to be arbitrary. I omit here the solutions of the dilemma based directly on the doctrine of God's simplicity. According to E. Stump (*Simplicity*, [in:] P.L. Quinn, Ch. Taliaferro (ed.), *A Companion to Philosophy of Religion*, Oxford: Blackwell Publishers 1999, p. 225), God is 'identical with perfect goodness' and God's simplicity entails that 'there is an essential relationship between God and the standard for moral goodness, and that standard is not external to God.'

Swinburne seems to accept the above thesis, but his examples suggest not that God may add essential new duties (in relation to the "natural" morality that is supervenient on the non-moral contents of the world), but, rather, that He may enhance, make clear, or develop the existing and "natural" duties. For the inhabitants of the world-with-God it is justified to assume that since the recognition of the moral consequences of non-moral properties is difficult and causes conflicts, one may expect that God will directly help people to recognise moral good and bad, and assist in their moral betterment, by means of revelation (confirmed by miraculous signs). I think, however, that for the inhabitants of this world it is justified to assume also that God (by means of revelation) may not only (as Swinburne suggests) interpret, or partially correct, the morality that already exists, but that He may add to it something radically new. Obviously, we can hardly expect that, in doing this, God should be inconsequent and order that which is "naturally" bad or prohibit that which is "naturally" good. However, it is a fact that in the world-with-God some moral obligations may come directly from God's commands, and not from the "natural" contents of the world; and that some of those additional obligations may be known only by virtue of God's revelation, and not for instance through philosophical consideration. Such a circumstance does not occur in the world-without-God: in that world the range of moral obligations is smaller, and their recognition is brought about solely by means of natural human cognitive capacities.

The above point should be supplemented with one reservation. Just as it is not easy to determine the content of our "natural" obligations, it is difficult, and perhaps even more difficult, to determine whether and where God's revelation was put into practice and what it contains. Therefore we have religious debates between believers of various religions and confessions. Such being the case, it is most prudent to base public life on "natural morality," and on its interpretation, which – as Swinburne writes – is 'a result of discussion and experience over many centuries'[6]. Any additional commands which may possibly arise through revelation should, then, be treated as obligatory only for those who have recognised these commands as coming directly from God, and as binding for them, by virtue of their personal or communal relationship with God[7].

[6] Swinburne, *God and Morality*, p. 8.

[7] The Decalogue may be interpreted either as a set of revealed Divine commandments

SWINBURNE AND THE MORAL ARGUMENT

As we can see, in this commentary to Swinburne's article I have tried to prove that the moral differences between the world-with-God and the world-without-God are much greater than Swinburne suggests. I have not undermined, however, his main thesis 'that the existence and actions of God make no difference to the fact that there are moral truths, but that they make a great difference to what those truths are'[8]. Indeed in both worlds there are moral truths. In the world-without-God, however, some moral truths – namely those that refer to God or those that are established by God – are, at most, counterfactual truths ("if God existed, then..."); while all 'general principles of morality' are necessary truths, though their application to the world is based on the "brute fact" of the existence of some contingent and non-moral states of affairs, that, ultimately, are not intended by anyone. Now in the world-with-God the range of factual moral truths is greater, and their obligatoriness and application to the world are based on the following:

- either on the fact that God exists;
- or on the fact that some non-moral states of affairs have on purpose been created by God;
- or on the fact that God issued and revealed special commands or prohibitions.[9]

The above differences between the worlds under consideration lead to 'a great difference' (greater than Swinburne suggests) in understanding

(within the framework of particular religions) or as a cultural inscription of "natural" moral consciousness, whether derived from God or not. This consciousness would, for instance, be a consequence of the fact that an obligation-making value supervenes on human life, or on life in general; thus the principal commandment 'thou shalt not kill!' could be expressed positively as: protect life! The remaining commandments would indicate (incompletely) various ways of protecting life: through the preservation of appropriate relations with the sources of our life, whether more or less proximate, as well as with the means of life, and with those with whom we share our life.

[8] Swinburne, *God and Morality*, p. 7.

[9] The proponents of the so-called Divine Commands Ethics emphasise the latter possibility (see J.M. Idziak, *Divine Commands Ethics*, [in:] P.L. Quinn, Ch. Taliaferro (ed.), *A Companion to Philosophy of Religion*, op. cit., pp. 453 and 458). In the broad sense we find its moderate or indirect representatives among all those who accept all the above three possibilities.

morality: in the world-without-God morality is at most a consequence of a certain contingent complex of non-moral states of affairs; whereas in the world-with-God it is the correlate and intended "work" of a perfect personal being. A difference in how we understand morality entails a difference in how we live it out. One may assume that, in being aware of the fact that God exists in their world, the inhabitants of the world-with-God will live out their moral obligations as more absolute and exceptionless than the inhabitants of the world-without-God, who are aware of the fact that God does not exist in their world: indeed, the foundation and correlate of morality is something greater for the first group than for the second.

The latter remark may be the basis for a certain version of an argument for the existence of God in the style of Swinburne. His arguments for the existence of God (together making up a cumulative argument) are based on a common procedure which, put simply, consists in proving that the theistic explanation of a given phenomenon (or a class of phenomena) is better (especially simpler) than some competing explanations; in other words, it is more probable (or expected) that a given phenomenon (or a class of phenomena) will come about in the light of the theistic hypothesis than in the light of competing hypotheses.[10]

In order to apply the above procedure to the phenomenon of morality let us assume that the majority of people (in the whole history of mankind) experience morality as absolute and exceptionless, as something very serious and objectively binding. Let us ask then in which of the possible worlds – in the world-with-God or in the world-without-God – it is more probable (or expected) that such moral consciousness will occur. I think that the probability of the existence of such consciousness in the second world is small. Moral consciousness in this world could be a result of some coincidence, a result of biological or social processes, or simply an illusion.[11] Such factors, however, can evoke various, and also contrary, effects. We can hardly expect factors of such mutability and accidentalness to evoke moral consciousness, which is permanent and

[10] Swinburne develops this type of argument in his book *The Existence of God*, Oxford: Clarendon Press 1979 (reprinted 1987).

[11] I do not take into consideration the non-naturalistic and at the same time non-theistic explanations of morality, for they seem to be too complicated, therefore they violate the criterion of explanatory simplicity.

absolute. It is different in the case of the first world: if there is a rational and good God-Creator, then we may presume with high probability that (directly or indirectly) He will lead to the existence of moral consciousness in people. He has both sufficient power and sufficient reasons to do this, e.g. He wants to preserve rational order in the life of free creatures created by Him and to show them His absolute and exceptional authority. (In any case the fact of the existence of God makes absolute moral consciousness something understandable: it has sense only if there is a personal source of all beings, towards Whom we are absolutely morally obliged and responsible). If it is more probable (or expected) that the phenomenon under consideration should appear in the world-with-God than in the world-without-God, then it is more probable that the world in which we live, and in which this phenomenon occurs, is identical with the world-with-God, rather than with the world-without-God. I think that the above argument is formally correct.

Let us end with one additional remark. Swinburne (*The Existence of God*, pp. 175-179) criticises the argument 'from man's moral consciousness', for according to him it is not difficult 'to explain [this phenomenon] by normal scientific [esp. evolutionary] processes'. I think, however, that this (scientific) explanation is possible, but not in relation to absolute moral consciousness, that corresponds to absolute, objective claims of moral obligation. According to Swinburne, an argument 'from the fact of morality itself' is faulty, because the fundamental moral truths (principles) are analytically necessary and as such do not need explanation. I think, however, that their necessity (which is, in my opinion, synthetic rather than analytic) is only conditional or relative, i.e. they are indeed related to the world insofar as there are some contingent non-moral states of affairs. Swinburne (though using a conceptual apparatus different than mine) notices this fact and proposes to formulate the following argument from morality: 'actions *a, b, c, d,* are obligatory; they would not be obligatory unless they were *Q, R, S, T*. It is more probable that there are actions which are *Q, R, S, T,* if there is a God than if there is not; therefore the obligatoriness of *a, b, c, d,* confirms the existence of God.' Swinburne, however, rejects this argument because it is difficult to find such actions accepted by standard ethical theories whose probability (of existence and obligation) rises together with the existence of God. In my opinion the probability of the obligatoriness of any actions

rises together with the existence of God because without His existence and creative action there could be no other beings or any obligations. However, the above statement transforms the moral argument into a metaphysical or cosmological argument. No wonder, then, that Swinburne states that: "I find »the moral law within« considerably less good testimony to God than »the starry heavens above«". Swinburne also states that the belief that "the voice of conscience is the voice of God" is only a consequence of our non-ethical knowledge about the existence and action of God. Despite this we may still be surprised by the fact that, in relation to natural phenomena, Swinburne ultimately uses personal explanation and not scientific (nomological) explanation, whereas in relation to moral phenomena he goes the other way round. According to him, morality is not ultimately explained by the intentions and powers of a perfect person, but by certain laws of correlation between natural (non-moral) and moral properties. Perhaps by postulating a personal explanation of morality I would like to be more "Swinburnean" than Swinburne. The problem, however, consists in showing whether its main premise – the premise of the universal existence of absolute moral consciousness – is true and whether the knowledge of its truth is independent of knowledge about the existence of God (or faith in the existence of God). One thing is certain: the debate on the character of moral consciousness and morality itself has been going on for ages, and in a sense it is connected with the debate on the existence of God.

BOOK REVIEWS AND NOTICES

CHRISTIAN HENGSTERMANN
University of Münster

Volker Dieringer. *Kants Lösung des Theodizeeproblems. Eine Rekonstruktion* (Forschungen und Materialien zur deutschen Aufklärung 22). Stuttgart: Frommann-Holzboog, 2009.
[Volker Dieringer. *Kant's Solution to the Problem of Theodicy. A Reconstruction*. Stuttgart: Frommann-Holzboog Publishers, 2009.]

Volker Dieringer's monograph *Kants Lösung des Theodizeeproblems: Eine Rekonstruktion* (FMDA 22), based on his PhD thesis, offers a detailed appreciation of the celebrated, albeit still controversial, argument of Kant's brief essay "On the Failure of all Philosophical Attempts in Theodicy". Dieringer intends this work to serve as a systematic contribution to contemporary debates in philosophy of religion. Referring to a conceptual distinction common in analytic philosophy, the author opts to view Kant's mature approach to the topic as a "defence" of moral theism rather than a "theodicy" proper. As Dieringer shows, Kant revokes his earlier attempts at rational theodicy, denying reason the very capability of seeking grounds for a holy, good and just God permitting evil. However, by so limiting the capacity of human reason, Kant purports to give a principled defence of moral theism from all atheistic refutation, as experience can be invoked neither to prove nor to disprove, in principle, the existence of a benign deity.

After briefly delineating Kant's historic translation of rational theology into practical rationality, in the introduction to his two-part study, Dieringer puts forward his basic claim that there is a break between Kant's own earlier endeavour to give substantive reasons for a benevolent and just God allowing evil to exist and flourish, and his later, more modest, assertion that evil as such poses no logical contradiction to the possibility of rational belief in the existence of a moral deity. In the first part of his monograph, Dieringer demonstrates that Kant initially espouses what Allison terms a "theological eudaimonism" (39): if the moral law, besides being the source of moral judgement, is also to be the "incen-

tive" (*Triebfeder*) of our actions, then practical reason must assume the existence of a moral world order, in which our ethical "worthiness to be happy" (*Glückswürdigkeit*) corresponds to our *de facto* sensual happiness. Consequently, the existence of God, as well as of a future life, must be postulated to bridge the apparent gap between these two principal halves of the "highest good", for which practical rationality strives. However, as the author rightly points out, this "theological eudaimonism" not only effectively undermines Kant's overall concern with moral autonomy, but its underlying moral theism also adds urgency to the question of theodicy, which Kant, well aware of the difficulty, sets out to address, in depth, in his lectures on rational theology, delivered in 1783/84. Closely following the most reliable postscript, the so-called *Religionslehre Pölitz*, Dieringer establishes that Kant, at that time, still advocated an ethical theodicy based upon the necessary hope for the actual realisability of the *summum bonum*. Thus, if it were not for a benevolent God reconciling man's ethical worthiness to be happy with his actual happiness, man would end up torn between an *absurdum pragmaticum* on the one hand and an *absurdum morale* on the other, with the categorical imperative of autonomous morality undermining the pragmatic one of happiness and vice versa. If man chose to follow the categorical imperative, he would be compelled to neglect the pragmatic one stipulating his innate pursuit of happiness. As a consequence, he would have to act like a "dreamer" whose rational hope for eventual justice would be continually frustrated. If, on the other hand, he resolved to obey the dictates of personal fulfilment at the expense of the categorical demands of morality, he would show himself a "villain" (48f.). Dieringer goes to some length to prove that, given Kant's premises, the dilemma presented is, in fact, a genuine one. Without any hope that the highest good will be realized, man would indeed be obliged to implement a moral imperative that he lacks the incentive to put into concrete action. In his lectures on rational theology, moreover, Kant gives a thorough account of his moral theism based upon God's three principal attributes. God, according to Kantian ethical theology, is to be seen as the holy legislator, the good sovereign and the righteous judge. As regards his essential holiness, God is completely governed by the moral law. Hence, his desire for man's happiness, designated by his goodness, is constantly checked by his uncompromising adherence to the moral law. The fact that he metes out happiness, only in ac-

cordance with an individual's worthiness to be happy, marks him as the righteous judge, which last quality, thus, specifies the mediation between the first two. It is with reference to these attributes of the moral deity that Kant defines the three fundamental objections that may be raised against moral theism: the existence of moral evil conflicts with God's holiness, that of physical evil calls his goodness into question, and the disproportion between crime and punishment impugns his justice. As Dieringer shows, Kant, in his early theodicy, generally adopts a threefold strategy, first denying the grievance the atheist puts forth, then tracing it back to necessary metaphysical evil and, lastly, offering a reason for a moral deity allowing evil to occur. Thus, with reference to the first atheistic indictment, he attributes moral evil wholly to man's succumbing to his sensual nature which, as a feature of his essential deficiency as a creature, is strictly inevitable. God, moreover, cannot but permit man's freedom to err, as the latter's free choice is indispensable to God's overall providential aim, the moral perfection of the whole of mankind. Similarly, Kant's approach to the question of physical evil, as well as justifying pain, as an unavoidable complement to sensual happiness, revolves around physical misery being defined as an incentive for man to strive for his ethical worthiness to be happy. The same holds true for the atheistic denial of God's justice: not only do the pangs of conscience that a wicked person is bound to experience prevent his attaining any happiness, but the apparent disproportion between individual morality and outward success also serves as an indispensable means to the divine end of promoting man's disinterested goodness, which, otherwise, might be pursued exclusively for the sake of personal contentment. Moreover, Kant expressly likens his postulate of a world in which the laws of nature and those of morality are eventually revealed to accord with one another, to Leibniz' celebrated tenet that ours is the best of all possible worlds. Kant holds that God, being good, cannot but have created the best possible world, with the latter being exactly the one designated by the universal *summum bonum* of his own theological postulates. Kant substantiates this idea by adopting God's universal perspective from which, Kant contends, the good by far outweighs the evil both physical and moral. Having outlined Kant's basic theodical strategy, Dieringer convincingly points out its overall failure, namely that the idea of God allowing evil, pain and flagrant injustice stands in undeniable contradiction to the basic premises of moral the-

ism. It is not surprising, therefore, that Kant chose to return to the manifest aporias of his moral theism later on.

The second half of Dieringer's study is devoted to Kant's defence of ethical theology in the third Critique and in his essay on theodicy. Dierenger shows that Kant expounds a new distinction between pragmatic and moral maxims, the first no longer belonging to practical rationality proper, but to man's empirical nature instead. Moreover, introducing the concept of "respect" (*Achtung*), he eventually views the moral law as being its own incentive, thus finally jettisoning his earlier aporetic eudaimonism in moral theology. On the basis of this unequivocal separation of the phenomenal and noumenal realms, Kant comes to view the problem of theodicy more clearly as rooted in the apparently irreconcilable tension between two completely distinct laws, i.e. those of morality and those of nature, which, in terms of religious philosophy, translate into God's moral and creative wisdom, respectively. Contrary to his earlier position, Kant, in his theodicy essay, no longer assumes the role of the moral theist, but that of reason itself judging the conflicting claims of the atheistic and theological positions. Moreover, his judicial hearing is characterized by a remarkable scepticism *vis-à-vis* his earlier arguments in favour of rational theodicy. Thus, he rejects any attempt either to deny the evil or grievance in question or, more significantly, to assign to it a larger divine purpose. As reason itself declares, there is in fact no denying the scandal of ubiquitous moral evil, which the atheist, in his accusation, adduces as an obstacle to man's belief in a holy creator. Moreover, the idea that from God's perspective manifest moral evil might ultimately serve a morally superior aim, strikes Kant as downright repulsive. Neither is it plausible that a good deity, bent on our happiness, should create beings subject to considerable suffering, which, on the whole, can generally be seen to outweigh their enjoyment by far. Likewise, the moral theist cannot reasonably reject the atheist's argument that the disproportion between crime and punishment apparently belies divine justice in the world. Injustice and the well-being of the wicked, as Kant now points out, can neither be dismissed as being impeded by the culprit's inevitable remorse, which, considering the latter's lack of ethical zeal, probably fails to materialize, nor explained as a means to promote the good man's moral progress, which would contradict God's goodness. Notwithstanding his more sceptical stance, Kant is still at pains to refine his earlier

arguments in favour of theodicy. Thus, rejecting an earlier view of freedom as the comparative ability to choose between alternative empirical goods, he now views moral autonomy as a capability that no longer bears on the empirical realm at all, but is completely *sui generis*. Hence, man being autonomous in an absolute sense, God is completely exculpated of moral evil: though inextricably intertwined with man's deficiency as a sensual being, evil may eventually be construed to arise solely from his faulty freedom of choice. In the end, even though the atheist seems to have gained the upper hand in the course of the trial, reason, surprisingly, pronounces that both conflicting positions fail in principle, since the dispute at hand cannot be resolved by human rationality at all. Man, reason elaborates its sentence, can never hope to understand the inner relationship between the empirical world and divine wisdom. Thus, the *a posteriori* evidence of the teleological nature of organisms entitles us to attribute to God a creative wisdom. Likewise, we may credit him with moral wisdom on the *a priori* grounds of the inward moral law. However, what eludes human rationality, in principle, is their unity in the Godhead, which both the atheist and the moral theist silently presuppose in their reasoning. Indeed, as regards the question of theodicy, reason faces an insurmountable dilemma which essentially strips the two opponents of their respective arguments: theoretical reason, grasping the phenomena as objects of experience, cannot but assume an outer intelligible substratum to exist behind them, although, according to Kant's mature critical philosophy, it is unable to comprehend it in any way. Practical reason, on the other hand, possesses insight into its own inner intelligible realm, as is warranted by the moral law. However, even if the reflective power of judgement may imbue nature with purpose and design for the sake of rational explanation, reason as such fails to establish in what way the noumenal world within may also be viewed as the principle of the world without. Thus, Kant arrives at a "transcendental argument" (116) that, as reason declares, will end all quarrels with respect to the question of theodicy: the "causal nexus between freedom and nature" (115) and their assumed unity in the Godhead being beyond our comprehension, neither the atheist nor the moral theist can hope to acquire insight into how God's moral wisdom might manifest itself within the empirical world. However, despite there being no possibility for reason either to justify evil or disprove theism on rational grounds, man may hope for eventual

reconciliation by following moral faith alone, which Kant, famously, calls an "authentic theodicy" (119). As is evidenced by the Old Testament figure of Job, man, confronted with the utter inscrutability of God's counsels regarding the nature of things and its connection with the moral order, is to trust that God will at last reconcile his worthiness to be happy with his actual bliss. Hence, in his essay on theodicy, Kant adopts an approach that, as Dieringer points out, is both prior to, and independent of, experience. As a consequence, a critical self-analysis of reason *per se* rules out both a theodicy that claims to give plausible reasons for God allowing evil, and the atheistic denial of a moral deity on the grounds of unbearable moral or physical grievances. However, far from advocating scepticism or a general *epoche* in relation to the question, Kant, in so doing, safeguards the practical hope for final justice, thus offering a defence of moral theism.

It is the chief merit of Dieringer's monograph that it offers a close reading of Kant's trains of thought, thereby rendering intelligible premises that are often tacitly assumed. Although repetitions are thereby not always avoidable, Dieringer's searching analysis, incorporating helpful formalizations, always succeeds in clarifying the single arguments as well as the overall aim of Kant's contribution to theodicy and moral theism. Still, given that the author's intention is a systematic one, more extensive references to topical debates on theodicy could have been added with benefit. It might have proven interesting, for instance, to discuss in more detail whether empirical evidence may indeed be excluded from a discussion of the vexed question of unjust suffering, as Kant proposes, or whether such an approach simply begs the question, as the atheist will probably retort. The same goes for several historical aspects which Dieringer mentions only in passing, such as Kant's debt to Leibniz's celebrated theodicy and Wolff's rational theology. On the whole, however, Dieringer is to be credited with having provided nothing less than an exhaustive appreciation of Kant's defence of moral theism. Moreover, in so doing, he has also shed light on the development of Kant's moral philosophy, revealing a plethora of intriguing details and less well-known aspects that enrich our overall understanding of his practical philosophy.

JASMIN HASSEL
University of Münster

Christian Kummer. *Der Fall Darwin. Evolutionstheorie contra Schöpfungsglaube*. Pattloch: München, 2009.
[Christian Kummer. *The Case of Darwin. Theory of Evolution versus Belief in Creation*. Pattloch: Munich, 2009.]

The past year has marked the 200th birthday of the famous natural scientist Charles Darwin and the 150th anniversary of the release of his major work *The Origin of Species*. This caused quite some activity on the book market and produced a lot of publications on Darwin's person and his teachings, including some works that look at his ideas from a philosophical or theological point of view. Christian Kummer's book, *Der Fall Darwin: Evolutionstheorie contra Schöpfungsglaube,* is an impressive example from the latter category. It is worth noting that the Jesuit Kummer, who has studied theology, philosophy and biology, and is a professor of natural philosophy in Munich and also director of the *Institut für naturwissenschaftliche Grenzfragen zur Philosophie und Theologie,* is the ideal author to combine the disciplines of biology and theology, which normally are seen to be disparate. It is inestimably valuable that there are authors who are willing to deal with the questions surrounding a possible dialogue between biology and theology, and are eager to reconcile the concept of evolution and the concept of creation, since one will find quite a number of prejudices on the biological side when it comes to a confrontation with the religious worldview. Of course, you can also find resentments within the category of religious believers, as there are still people who oppose the methods of the natural sciences, especially evolutionary theory. This is particularly evident in the debates that currently take place in the USA, especially between representatives of so-called "intelligent design" and creationists on the one side, and biologists on the other side. And the unfortunate clash of worldviews is echoed by debates among theologians and philosophers. It is probable that at least some aspects of this furious debate will begin in Europe quite soon. Therefore, it is as wise as it is prudent to be prepared for this, and to deal with it in advance, as Kummer does in his book.

Before we take a closer look at the details, and at the content of Kummer's *Der Fall Darwin,* it is necessary to get acquainted with the author's intention and with the audience Kummer had in mind when he wrote the book. In the preface the author points out that he wanted to address the book to readers he was familiar with from public lectures and other educational settings. So, the book is not meant to meet the interests of scholarly readers, but is dedicated to pretty much everyone who is looking for guidance in an overheated situation (p. 11). It is important to keep that in mind in order to appreciate the sometimes colloquial or common-sense tone of Kummer's writing.

The very first chapter of the book (pp. 15-31) starts with a detailed analysis of the expression "evolution" and seeks to exemplify Darwin's theoretical approach without concealing the conceptually blank positions. Right from the start, it also raises the question whether a biologist has to be an atheist. Kummer illustrates the fact that quite a number of natural scientists are biased against religion, by using some anecdotes, coming from a rich experience as a researcher in both fields. Based on the many forms of prejudice in the natural scientist's camp, Kummer seeks to uncover the motives that may lead to a positive answer to the question at issue. But corresponding to this very concern, i.e. showing the "potential for reconciliation between the Christian faith and the scientific world view" (p. 10), it is Kummer's goal to demonstrate, across the whole argumentative journey of the book, that an alternative answer is possible and even more convincing.

The second chapter (pp. 33-62) takes a closer look at Charles Darwin. It starts with a basic introduction to Darwin's person, provides the reader with some biographical information and gives some hints regarding Darwin's educational and cultural background. This is a noteworthy approach since it reveals the interesting fact that in Darwin's very own educational upbringing, the relationship between theology and science was of utmost importance and that Darwin himself was theologically influenced, if not biased. In other words: A certain form of theology, that was quite influential in Darwin's time, became relevant for the very framework of Darwin's theory; and this very fact is also highly problematic, as Kummer points out, since the type of theology Darwin was acquainted with was neither advanced nor at all willing to adjust to the new findings of the natural sciences. The second chapter of the book

offers also a more narrative approach to the matter: Kummer seeks to familiarize the reader with Darwin's expeditions; furthermore, he explains the main outlines of his concepts and illustrates the theory of evolution utilizing the example of orchids.

The main issue of the third chapter is a very philosophical one since it deals predominantly with the question whether there is purpose in nature (pp. 63-115). In the third part of the book Kummer pleads for a teleological perspective, although he also agrees with certain reservations had by biologists, who make us aware that it is often an unjustified shortcut to attach something like a purpose to a complex event we don't fully grasp or understand. To meet the needs of a sophisticated teleological reflection Kummer introduces three different levels of purpose: He speaks of so-called internal teleology (pp. 73-90), i.e. the "usefulness in the construction of living systems" (p. 73), so-called external teleology and higher development (pp. 90-111) and teleology in human experience (pp. 111-115). In a note explaining the outlines of the first level of teleology Kummer makes us aware that biology, in order to avoid the term "purpose", has introduced the concept of teleonomy which means: something seems to be goal-oriented because of an "(internal) program" (p. 74). In Kummer's opinion this concept is not uncontroversial since it seems like a complicated detour to avoid the word "teleology". Equally debatable, according to the author, is the concept of so-called "higher" evolution, which biologists no longer accept. Kummer, however, tries to give evidence for this idea and seeks to spell out relevant criteria for what is called "external teleology" (p. 106). Ultimately, people's ability to set goals might be, in Kummer's view, an indication for the existence of purpose in nature.

Chapter 4 (pp. 117-155) offers a discussion with representative scholars that are known under the banner of 'intelligent design', namely Michael Behe and others. It is noteworthy that Kummer approaches their rationale coming from a principle of charity. So, Kummer points out that ID-theorists doubt that the complexity of life, illustrated by the human eye considered as a complex piece of biological machinery, can be explained solely in reference to natural selection. In this context the concept of "irreducible complexity" (p. 124) plays a crucial role. As ID-theorists point out, whenever irreducible complexity is attributed to biological features this means that the organs in question do not have any

evolutionary pre-stages to which their current layout can be traced back. But in order to illustrate the tricky parts of the theoretical assumptions of ID, Kummer demonstrates, also using the human eye as a biological example, that Darwin's ideas are perfectly suited to explain the origin of complex organs. However, Kummer admits that ID-theorists still have a point whenever they address certain flaws in evolutionary explanations.

Chapter 5 (pp. 157-195) is dedicated to reflecting on more theological ideas, especially the concept of creation. In an initial step Kummer criticizes an analogy which was brought up by ID-theorists in order to draw parallels between results coming from artificial or natural processes. Kummer underlines that we need to be aware of a difference between "making" (artificial) and "developing, growing, increasing" (natural). While the first category necessary includes speaking of someone who makes something, for the category of development no talk about a maker or creator is necessarily required. In order to prepare a conceptual framework to approach a more sophisticated concept of creation Kummer introduces the theologian, philosopher and palaeontologist Pierre Teilhard de Chardin, whose model of corpuscle and concept of radial and tangential energy the author presents briefly. Using Teilhard de Chardin's ideas as a backbone, Kummer tries to show "that it is quite reasonable to introduce a concept of creation to an evolutionary world view" (p. 185). Thus Kummer argues along the lines of Teilhard de Chardin that there is an "evolutionary creation" (p. 188), in which God does not make things "but rather […] allows things to make themselves" (p. 187).

In a follow-up section (pp. 197-239) Kummer tries to defend Teilhard de Chardin's viewpoint, while dealing with some critical, and especially Catholic voices, directed against Teilhard's concept of radial and tangential energy, and Teilhard's famous concept of the 'the omega point'. Kummer discusses the Catholic Church's attitude towards the theory of evolution and reveals a certain disappointment with the fact that some groups within Christianity still refuse to accept Darwin's ideas and findings. As an excursus Kummer also offers a brief introduction to the philosophical problems of neuro-biology and to the question of religious experiences.

While chapter 5 was dedicated to arguing for the reasonableness of Kummer's own position, i.e. the assumption of a Creator God in the light

of the evolutionary theory, chapter 7 (pp. 241-265) approaches the question whether natural scientists and especially biologists also have a certain metaphysical base they need to be aware of. Ultimately, Kummer wants to prove that not only theologians but also natural scientists cannot do without any idealistic and metaphysical presumptions although their "requirements are more subtle, more hidden but nonetheless still virulent" (p. 243).

The final chapter, chapter 8 (pp. 241-265) once more, and in conclusion, reflects on the relationship between theology and biology. It is quite provocative to see that Kummer does not support a polite side by side existence of the two disciplines; instead he goes much further when he underlines that both disciplines could benefit from each other. In that respect, it is not only Christian doctrine and theology that have learnt and can learn something from the natural sciences, for biology also the dialogue with theology could have its benefits and may be enriching.

After this overview of the main chapters and contents of the book some general comments are in order: As pointed out at the very beginning, this is not meant to be a scholarly book; it is written for a broader audience. This becomes obvious at several points where the line of argument seems to get lost or stuck or isn't continuously kept intact. Moreover the line of argument is sometimes sidetracked by digressions or personal anecdotes. However, these anecdotes lighten up the tone of the book. The quotes and sayings ascribed to Karl Rahner and Richard Schaeffler are especially enjoyable; and one can also uncover interesting thoughts and insights in those more narrative passages.

With the author's decision against a more academic and professional tone the reader has to be prepared for a rather casual style of phrasing and writing. But this also has positive effects, since it catches almost any reader's attention for a prima facie complicated area of research. By dressing up his text in rather low-key means of communication Kummer is able to explain the most difficult issues by using the most vivid illustrations. Nevertheless, there are some things which are even below the standards of non-scholarly writing. And it is stunning that the author didn't even try to avoid them. For example, one can find quite a number of (simply) cut and pasted pages taken from Wikipedia articles. That Wikipedia is quite an unreliable source of information is something every undergraduate student already knows. To find such things in a book

written by a well known author and printed by a well-known publishing house is pretty surprising. Furthermore, Kummer's main source to illustrate and explain the outlines of ID-theories is the website of a confessed atheist. As anybody knows: there are enough alternative sources that are much more unbiased and reliable than the one Kummer is using, one could think of numerous web pages created by Behe and Dembski, as first-hand authorities, which can inform us about the ID-theory.

Beyond these rather general remarks another note on the tone of the book is in order. This tone is sometimes quite apologetic, especially whenever Kummer approaches rival opinions, or natural scientists who clearly belong to the atheistic camp. This attitude may be a result of the origins of the book – in lectures he had offered to a broader audience all over Germany, lectures that were discussed and debated publicly and heatedly.

In addition a more technical error or rather a technical inaccuracy needs to be addressed. Kummer does not distinguish between the terms "creationism" and "intelligent design" and appears to use both phrases synonymously. In this case the low-key tone of the book goes somewhat too far – especially when certain labels induce very specific arguments or counter-arguments.

Another issue also requires a more specific and more detailed handling. As Kummer points out the evolutionary metaphysics of Teilhard de Chardin entails some sort of pantheism. But Kummer does not take into account the necessary difference between *pantheism* and *panentheism* (p. 192) although the latter could be reconciled with the basics of Christian doctrine and could provide a conceptual basis that is beneficial for Kummer's argument, and of genuine interest to him. In addition Kummer hasn't really shown why or rather how his suggested solution, i.e. Teilhard de Chardin's concept of radial and tangential energy, really differs from the assumption of an intelligent designer (p. 181). Hence, we are still left with the task of spelling out systematically, the differences between the notion of creation on the one hand and purely natural evolution on the other hand. This goal is accomplished only partially by Kummer's reference to Teilhard de Chardin. And it would have been fruitful to take a look at contemporary adherents of Teilhard de Chardin's ideas outside the German speaking world – especially at contributions coming from US authors and theologians such as John F. Haught and others.

Nevertheless, Kummer's investigations and deliberations are really of benefit to a broader audience. Especially noteworthy are chapters 5, 7 and 8 in which the author tries to mediate between biology and science on the one hand and theology on the other. Despite the critical remarks on Kummer's methodology and terminology his work can be seen as a very valuable contribution that successfully leaves behind the sometimes narrow framework of purely academic discussions; its main achievement is to demonstrate that the evolutionary theory does not necessarily threaten the belief in a Creator God and to bridge the gap between biology and theology in showing that, ultimately, both disciplines are mutually dependent. Kummer's book can be recommended to those who are seeking an initial but also substantial insight into the subject, since the book is written by an author who is familiar with all the disciplines involved, and is a trustworthy and reliable scholar, who, after all, plays an important part in current debates on the New Atheism.

OLIVER WIERTZ
Philosophisch-Theologische Hochschule, Sankt Georgen

James Beilby. *Naturalism Defeated? Essays on Plantinga´s Evolutionary Argument against Naturalism*. Ithaca and London: Cornell University Press, 2002.

Alvin Plantinga´s "Evolutionary Argument against Naturalism" (EAAN) has created a great stir since its release at the beginning of the 90's. There are several reasons for this: on the one hand, the ontological naturalism that Plantinga opposes in his EAAN is more or less seen as the official doctrine of contemporary analytic ontology; on the other hand, Plantinga argues that the modern synthetic theory of evolution, the sanctum of modern naturalism, has to presume the existence of a theistic God, if it wants to avoid radical skepticism. Plantinga does not attack the theory of evolution in his EAAN, but rather its combination with ontological naturalism, as fostered by critics of theism such as Richard Dawkins. Plantinga tries to constrain the naturalists to a decision between a theory of evolution on the one hand and metaphysical naturalism on the other hand. After all, it is his epistemologically externalist theory of warrant that he refers to in the EAAN. However, epistemological externalism is a vital component of naturalism. Plantinga therefore approves of naturalism in epistemology, in a more or less unrestricted way. From that perspective epistemology has no normative character at all; and so it should become part of an empirical discipline, for instance cognitive science. Since there are, according to metaphysical naturalism, no non-physical or super-natural entities, Plantinga distinguishes rigorously between epistemological and metaphysical naturalism; the latter is what he stoutly denies and tries to disprove in his EAAN.

In order to grasp the core of Plantinga's argument, there is a need to acquaint oneself with the main features of Plantinga's epistemology, in order to understand the basic idea of EAAN. Plantinga presupposes a proper-function-theory of epistemic warrant. Broadly speaking, Plantinga understands `warrant´ as an epistemic feature which transforms true beliefs into knowledge. More specifically: For an epistemic subject S a belief B has warrant, if B is a product of the cognitive faculties of S, these faculties act properly according to their design that is oriented to

creating true beliefs, they work in an adequate environment and are not disturbed by negative influences, and if S knows no good reason against the truth or probability of B, i.e. if S doesn't have a defeater of the warrant of B.

The basic EAAN is made up of three steps: (1) If we assume the truth of naturalism (N) and of the theory of evolution (E), then the (objectively determined) likelihood that we possess a reliable cognitive ability (R) [P(R/N.E)] is rather small, or even not assignable. Plantinga justifies this key premise of EAAN by arguing that a naturalist cannot give reasons for the action-relevance of certain convictions. In other words: true convictions are not necessarily favored by the mechanisms of natural selection. This implies (2) those who accept N and E, possess a rationality – or warrant – defeater for the conviction R that their cognitive capability works reliably. But this implies (3) that every epistemic subject that has the convictions E and N and whose cognitive (defeater-) capability works properly (in order, also, to recognize the negative effects of N and E on R) has a direct defeater for R, and with that an indirect defeater for all of his convictions, including E and N. And so his convictions lose their warrant. The combination of the theory of evolution and of ontological naturalism is self-defeating since, in this way, E and N form the core of a defeater for E and N. Since every possible naturalistic defeater of EAAN has to presuppose convictions and the reliability of convictions within the context of evolutionary naturalism, naturalism as such is defeated by EAAN. And thus, basically, EAAN cannot be naturalistically defeated. Therefore Darwinian naturalism is hopelessly self disproving, and hence not a rational option.

James Beilby, the editor of the present volume, includes in it a brief statement of EAAN stemming from Plantinga himself, eleven articles with different objections and Plantinga's answers to them. The articles and Plantinga's typically precise answers range over a variety of topics. According to their arrangement in the book the articles can be classified into four groups. On the one hand they deal with the relationship between the theory of evolution and the reliability of our cognitive apparatus, i.e. with the quality of P (R/E) (Ramsey, Fodor, Fales). The second group deals with the transition from the second to the third step of EAAN, and thereby with the problem of skepticism, and contains (besides articles by E. Sosa and J.V. Cleve et al.) a reformulation of Thomas Reid's

common sense anti-scepticism (Bergmann). The third group focuses on the question of the nature of conditional probabilities, which question is crucial for the first step of EAAN, i.e. this group examines the relativity of probability adjudication given relevant information and deals with the question of how the relevant amount of information can be assigned (O'Connor, Otte). The final group consists of three articles dealing with the nature of epistemic information and an appropriate interpretation of what it is to be or to have a defeater, as the latter is assumed within the second and third step of EAAN (Talbott, Merricks, Alston). It is impossible to go into the details of each article, and of Plantinga´s answers, due to the plenitude of thoughts and insights.

Plantinga's defense of his EAAN is mostly, though not always, convincing. The cogency of his argument, or lack thereof, shall be outlined with respect to the two arguments he offers in order to introduce the first step of EAAN.

(1) Based on good reasons, Plantinga indicates that naturalistic theories of the mind imply semantic epiphenomenalism, i.e. make convictions for our actions extraneous. His differentiation between the question of the causal relevance of a conviction qua neuronal appearance and the causal relevance of the content, the propositional object, of a conviction is very helpful. Indeed, the content of a conviction qua content has to be causally irrelevant in the naturalistic perspective. However, from the high probability of semantic epiphenomenalism (S) within the constraints of naturalism and the theory of evolution [P(S/E.N)>0.5] it does not necessarily follow that the truth of convictions, and the evolutionary process of selection, have nothing to do with each other. So it does not necessarily follow from a high degree of P (S/E.N) that the probability that our cognitive mechanisms are reliable, within the constraints of naturalism, the theory of evolution and semantic epiphenomenalism [P(R/N.E.S)], is low or not assignable. In other words: Plantinga fails to show this. A naturalist can embrace semantic epiphenomenalism and argue for a causal relationship between the content of convictions and certain occurrences in evolution. What is required as a basis, beyond the basic recognition of the fundamental principles of the theory of evolution, is simply any kind of mind-brain-identity thesis, i.e. the thesis that mental events or patterns are identical with neuronal events or patterns (whether in a token, or type version). Plantinga has to grant this to the naturalist, as he

himself presupposes the identity thesis within his justification of the first step of EAAN. For the identity-theorist a conviction is a neuronal event within which several neurons fire (with inputs and outputs from other neuronal processes and events). Of course, for the naturalist a conviction qua neuronal event is, by causing impulses in the nerves which finally cause the contraction of muscles, causally relevant (which Plantinga admits in part B of his answer). If the resulting movement is within the range of survival-oriented maladaptive behavior with respect to the surrounding environment of the human being, the neuronal pattern behind this movement will be evolutionally withdrawn and, instead, neuronal patterns which cause actions that are better adjusted to the surrounding environment will be preferred. Thus, neuronal patterns get modified via natural selection in order to produce actions adjusted to the surrounding environment. But according to the identity-theory the neuronal pattern determines the content of a conviction (although it may be unclear how this happens in detail). In that case one has to treat convictions, which actually provide the foundation for behavior that is well adapted to the surrounding environment, as if they were causally relevant – like it or not. In other words: in such a case one has to treat those convictions as if they were probably true, because in regard to their causal efficiency they would serve as the foundation of behavior that is well-adapted to the surrounding environment. The convictions as such are not causally relevant, but they can be seen as effective indicators of neuronal patterns that cause behavior that is well-adapted to the surrounding environment and that is, in so far as natural selection is involved, mediated through the modifications of action-relevant neuronal patterns. So, even within a naturalistic theory, a causal influence on convictions and a guided modification of convictions directed by something like a greater reality-accommodation can be spelled out. Plantinga's mistake seems to be that he only takes into consideration the causal relation between conviction and surrounding environment via the causal influence of convictions, qua convictions, on actions. He thereby overlooks the possibility that via natural selection and appropriate modification of neuronal patterns, that determine the content of the conviction within the naturalistic point of view, a selection towards greater truth-likelihood can take place. Such an argument against Plantinga's claim that P(R/N.E.S) is low, is anything but irrefutable, because it presupposes that it is clear that or how

neuronal structures can receive or produce semantic content. But from semantic epiphenomenalism alone, no criterion for Plantinga's thesis that P(R/N.E.S) is low or incalculable can be gained.

(2) Plantinga shows, in his second justification of the first step of EAAN that even within the assumption of the action-relevance of convictions it is impossible to derive their truth from the selection-based advantage of certain convictions. Surely there is no essential relation, which connects especially abstract, philosophical and metaphysical convictions to patterns of behavior. But the vital point is to ask how likely it is, given the truth of a theory of evolution, that the (survival-relevant) convictions of beings that are well adapted to their surrounding, are wrong, or that their cognitive mechanisms concerning certain kinds of convictions work unreliably, i.e. create mainly false convictions. The presumption is that the probability is, by all means, not zero. With the help of the theory of evolution the epistemic reliability of the cognitive mechanisms of well adapted natural kinds of beings cannot be assumed to be certain. But the hypothesis of the reliability of the cognitive mechanisms of well adapted natural kinds of beings, appears to me to be prima facie equally justified, since it is most likely that this is the easiest explanation of what we call 'advantage' in the process of selection. It is indeed possible to explain the well adapted behavior of human beings with a flamboyant combination of false convictions and wishes rather hostile to survival, but in doing so one is just postulating flamboyant combinations, which appear, at least at first glance, less reasonable. For the naturalist a higher figure of P(R/E) will do to defeat EAAN naturalistically.

Anyhow, Plantinga showed two things in his EAAN: 1) although the opposite impression is nourished on a regular basis, ontological naturalism is anything but an unproblematic or invariable position; 2) one cannot readily derive ontological naturalism from methodological-epistemological naturalism. Possibly, methodological naturalism requires more epistemological sophistication or even an ontological "supranaturalism" as its foundation.

Translated by Anna Schneider

CALL FOR PAPERS

International Conference in Philosophy of Religion:
Philosophy of Religion in the 21st Century

27-29 June 2011 – Krakow, Poland

Organizers. The conference is organized under the auspices of the European Journal for Philosophy of Religion in cooperation with the Jagiellonian University of Krakow, Society of Christian Philosophers (USA), Polish Society for Philosophy of Religion, and the Central European Society for Philosophy of Religion.

The focus of the conference will be any issues that from today's perspective seem likely to dominate the debate of philosophers of religion in the decades to come. While the majority of the participants in the conference will be practitioners of analytic philosophy of religion, speakers are invited to consider the meta-philosophical question whether "an exchange between representatives of various traditions within contemporary philosophy of religion" (a phrase taken from the mission statement of the European Journal for Philosophy of Religion) might positively shape the future of the philosophy of religion in the course of the 21st C, and therefore should be encouraged.

Keynote Speakers: Eleonore Stump, Linda Zagzebski, Richard Swinburne, William Wainwright, et al.

Call for papers:
There will be opportunity for 40 concurent papers.
Authors are invited to submit a 400-600 word abstract for a paper of about 30 minutes reading time.
The abstracts are to be submitted by e-mail, as an attachment (pdf or doc).
Please send abstracts and requests for further information to
Dr. Janusz Salamon: journal@philosophy-of-religion.eu

The submission deadline is 31 December 2010.
Prospective presenters will be notified by the end of March 2011.
Selected papers will be considered for publication in the European Journal for Philosophy of Religion.

Further details regarding the event will be posted in due course on the conference website: www.philosophy-of-religion.eu/conference2011.html