Infallibility and Modal Knowledge in Some Early Modern Philosophers

Jonathan Bennett

From: *Mathematics and Necessity: Essays in the History of Philosophy*, ed. T. Smiley (Oxford University Press, for the British Academy (*Proceedings* 103, 2000)), pp. 139–66.

1. Introduction

In the early modern period, the term 'reason' and some of its cognates were associated with several theses, of which I shall discuss two. One is the view that reason informs us about modal truths, showing us what is possible and what is not. In my opinion, the problem of modal epistemology is still unsolved: we have no good account of how we get our modal information. This may help us to look uncondescendingly at some early modern assaults on the problem—which may aid us in our own philosophical thinking and should at least be good for our souls .

First, though, I shall discuss more briefly the attempts by some early modern philosophers to explain why it is that reason when used properly is infallible—absolutely guaranteed not to lead from truth to falsity. Probably none of us think that any of our faculties has that virtue. These days, I should think, most philosophers will agree broadly with Hume's description of reason as 'a kind of cause, of which truth is the natural effect'. For us Humeans, reason's leading someone from a truth to a falsehood would be unnatural or pathological, like a failure of the diaphragm to keep the lungs on the move, but we are not going to say that it absolutely could not happen if reason is used properly—unless we trivialise the whole affair by defining propriety of use in terms of truth of output. Hume was clearly unable to regard such a failure as impossible. He thought of the exercise of reason as a causally explicable process of moving from one idea to another; any absolute guarantee of never moving from truth to falsity would he a guarantee of the reliability of a certain causal process; but Hume is committed to the view that (in his own words) 'To consider the matter *a priori*, anything may produce anything'.

Still, even if the theme of the infallibility of reason is not alive for us today, I hope it is still of some interest to see how philosophers who thought otherwise tried to explain this supposed fact. I shall not try to connect the infallibility theme integrally with the one about modal knowledge. When I announced the topic of this piece, I thought I could usefully link the two, but I have found that I cannot.

2. Descartes on the security of intuition

Let us look at Locke and Descartes—I mean the early Descartes of the *Regulae*. They saw a need to provide some grounding for their belief in the infallibility of reason, and believed they had found it in a pair of thoughts. One is

that any exercise of reason is a chain of small episodes, controlled in such a way that if each episode is sound then so is the whole chain. The traditional antithesis between reason and sense-experience implies that the exercise of reason—whatever it may be—is not a process of looking out at the world, which led Descartes and Locke to their second thought, namely that reasoning is a process of looking inward at one's own mind. So the links of which a reasoning-chain is composed are what both philosophers called 'intuitions', little episodes of self-examination, in which one cannot go wrong. Thus Descartes:

> Everyone can mentally intuit that he exists, that he is thinking, that a triangle is bounded by just three lines, and a sphere by a single surface, and the like. . . Many facts which are not self-evident are known with certainty, provided they are inferred from true and known principles through a continuous and uninterrupted movement of thought in which each individual proposition is clearly intuited.[1]

Why is intuition infallible? Descartes does not answer this in his early work, merely showing his confidence that it *is* infallible in his metaphors about *light*—light of reason, light of nature, natural light, and so on—which occur more than a hundred times in the three volumes of CSM.

In later years, he raised the 'Why?' question in the context of the battle against scepticism in the *Meditations*. 'How do I know that I do not go wrong every time I add two and three or in some even simpler matter, if that is imaginable?', he asked. His answer—one thread in a dense tangle which I shall not try to unravel—relies on this truth rule:

> R: If someone cannot doubt that P while having P perfectly clearly in mind, then P is true.

Descartes announces other truth rules as well, as though they were equivalent to one another, which they are not. R is the one that was attributed to him by his acutest reader, Spinoza, I am sure rightly. The only decent argument Descartes has for any of his truth rules supports R and not the others.

The argument is theological. Descartes thinks he has shown that there exists a God who is maximally real and powerful; he takes it that deception must come from weakness, from which he infers that God does not deceive. Now, God is not convicted of deception by his allowing me sometimes to have false beliefs; for in any such case I have the option of suspending judgement on the proposition and/or of conducting myself as though I were unsure of it by investigating whether my confidence in it results from some muddle on my part. But if God ever allowed me to be wrong about something that I *could* not call into question at a time when it was *perfectly* clear in my mind, that would be deception. Such an error would be unavoidable: I could neither suspend judgement nor actively investigate whether my inability to do so was my fault. God could permit this only if he wanted me to be in error, which would make him a deceiver.

This is a coherent explanation—though in my opinion not a true one—of how it comes to be that an intellectual faculty of ours is infallible. Descartes provides it in a context where he apparently needs also to claim that he knows infallibly that he has this kind or infallibility; but that is no part of my topic, which is what spares me from having to plough the rocky and barren fields of the issue of the 'Cartesian Circle'.

[1] Descartes, *Rules for the Direction of the Mind*, Rule Three; in J. Cottingham, R. Stoothoff, and D. Murdoch (eds), *The Philosophical Writings of Descartes* (Cambridge University Press, 1985), vol. I, pp. 14, 15; henceforth 'CSM'.

3. An aside on Descartes's stability project

Truth rule R fully belongs to Descartes's normative epistemology—his pursuit of reasons, justification, entitlement to believe. One of its concepts, however, points to a quite different project which he engages in along with the normative one, perhaps not clearly distinguishing them. The concept in question is that of a proposition's being indubitable by someone, meaning that he is psychologically unable to call it into question; and the project is that of arriving at a system of beliefs that is stable, durable, secure against changing. The psychological stability project was mostly overlooked by Descartes scholars until Louis Loeb and I independently discovered and reported it.[1] Yet there is more of it than of the normative project, and in many ways it is more interesting and better done than the latter.

My paper on the topic exhibits the texts in detail, which I shall not go into here. For quick evidence that there may be something in what Loeb and I say, look at the title of the First Meditation—'Of Things that Can be Doubted', not things that ought to be doubted or that admit of doubt, but things which it is possible to doubt. Look also at the opening sentences of that Meditation, and see how quickly Descartes moves from truth to stability:

> Some years ago I was struck by the large number of falsehoods that I had accepted as true in my childhood, and by the highly doubtful nature of the whole edifice that I had subsequently based on them. I realized that it was necessary, once in the course of my life, to demolish everything completely and start again from the foundations if I wanted to establish anything at all in the sciences that was stable and likely to last.

Here is a little more evidence:

> As soon as we think that we correctly perceive something, we are spontaneously convinced that it is true. Now if this conviction is so firm that it is impossible for us ever to have any cause for doubt about what we are convinced of, then there are no further questions for us to ask: we have everything that we could reasonably want. What is it to us that someone may make out that the perception whose truth we are so firmly convinced of may appear false to God or an angel, so that it is, absolutely speaking, false? What do we care about this absolute falsity, since we neither believe in it nor have even the smallest suspicion of it?[2]

This resembles Hume's flatly naturalistic treatment of beliefs which are not budged by sound arguments, this being a victory of 'nature' over 'reason'.[3]

The only hint of anything normative in that passage is 'everything we could reasonably want', and I cannot see how to fit that in with the rest.[4] On the next page something similar occurs, this time with no hint of anything normative:

[1] Jonathan Bennett, 'Truth and Stability in Descartes's Treatment of Scepticism', *Canadian Journal of Philosophy*, suppl. vol. 16 (1990), 75–108; Louis Loeb, 'The Priority of Reason in Descartes', *Philosophical Review*, 99 (1990), 3–43, and 'The Cartesian Circle', in J. Cottingham (ed.), *The Cambridge Companion to Descartes* (Cambridge University Press, 1992), pp. 200–235. The central idea in these papers was adumbrated by Jaegwon Kim, 'What is Naturalized Epistemology?' (1988), reprinted in his *Supervenience and Mind* (Cambridge University Press, 1993), at p. 219.

[2] Descartes, Replies to the Second Objections, CSM, vol. 2, p. 103. Descartes means, I take it, 'this alleged absolute falsity'.

[3] Hume, *Treatise* I.iv.1, 'Of Scepticism With Regard to Reason'.

[4] The translation in CSM has two other normative touches, but both are wrong: 'reason for doubting' where 'cause of doubting' is more accurate; and 'why should x bother us?' where the Latin means 'what do we care about x?'

'It is also no objection for someone to make out that such truths might appear false to God or to an angel. For the evident clarity of our perceptions does not allow us to listen to anyone who makes up this kind of story.'

This strand runs strongly through Descartes's thought. Although he did not announce it as doctrine, he often writes as though he would agree with Hume: 'We assent to our faculties and employ our reason only because we cannot help it. Philosophy would render us entirely Pyrrhonian, were not nature too strong for it.'[1]

This section is parenthetical. I now return to my proper topic.

4. Locke on the security of intuition

Like Descartes. Locke thinks of reasoning as composed of little acts of intuition, and as owing its security to that: 'Certainty depends so wholly on this intuition that in the next degree of knowledge, which I call demonstrative, this intuition is necessary in all the connexions of the intermediate ideas, without which we cannot attain knowledge and certainty' (*Essay* IV.ii.1). So it comes down to the credentials of the intuitions. Here is Locke defending those:

> Intuitive knowledge is irresistible, and like bright sunshine forces itself immediately to be perceived, as soon as ever the mind turns its view that way; and leaves no room for hesitation, doubt, or examination. He that demands a greater certainty than this demands he knows not what, and shows only that he has a mind to he a sceptic, without being able to be so. (Ibid., quoted with omissions)

Locke here gives voice to his sturdy British refusal to take seriously an extravagant Frenchman who claims to wonder whether 'three are more than two'; he is rightly sure that everyone is utterly sure of the truth of that. But that does not address the question: what reason is there for such confidence? I cannot find that Locke ever does face up to that.

5. Spinoza on reason's infallibility

The other philosopher I want to report on is Spinoza. He too connects senses/reason with outer/inner, but not as Descartes and Locke do. According to them, one exercises reason by looking inward to get information about one's own mental states, one's own ideas; and there is nothing like that in Spinoza's picture of the human condition. He allows for our being aware of our own mental states and processes, but he does not attribute this to an inner sense, handling it instead through his strange theory about 'ideas of ideas'. So the kind of outer/inner contrast he uses to explain senses/reason has nothing to do with looking inward or, therefore, with reason as consisting in or resting upon self-knowledge.

Here as so often it is instructive to compare and contrast Spinoza with Hume. They are alike in regarding reason as a causal affair; but Spinoza also thought that it is infallible, and, unlike any other philosopher I know, he offered a complete theoretical explanation of how this could be so. No one today would believe it for a moment; but this offering of Spinoza's is too original and daring to pass by without notice.

The crucial causal distinction is that between **(i)** an idea of you that is caused purely from within you and **(ii)** an idea of yours that caused from outside. All the ideas of sense are exogenous, cause from outside. When you hear thunder, a causal chain runs from the thunder to changes in your body, and a parallel chain runs from the mental counterpart of the

[1] Hume, Abstract of the Treatise, §27 ('By all that has been said. . . ').

thunder (whatever that is) to your mind. Both changes, then, are exogenous. In contrast with this, when you withdraw from the world and conduct a train of thought untainted by input from outside, this involves your having ideas that are endogenous, caused from within.

Thus Spinoza is in a good position to say that the ideas (= beliefs) one reaches through reason are endogenous, while those we reach through the senses are not. He gets to his infallibility result—his thesis that reason cannot lead us into error—with help from the premise that all endogenous ideas must be true.

Strictly speaking, Spinoza says, all ideas whatsoever are true: 'There is nothing positive in ideas that makes them false.' There are reasons (of a sort) for this in his official theories, though they are rickety at best. His deepest reason for it, I believe, was his rejection of the idea that a natural object—an item in the real world—could be false. You might think 'Well, beliefs can be false; that is what is special about them'; but Spinoza resembled Wittgenstein in not being willing to appeal to any kind of specialness of the mind to explain anything.[1] Far from regarding the mind as a 'queer kind of medium'—in Wittgenstein's mocking phrase—Spinoza held that the whole truth about a person's mind matches the truth about his body. So the question remains: how can a part of the world be false? Spinoza answered: it cannot.

Still, he has to allow that people sometimes go wrong; he cannot deny that there are errors. He maintains, though, that all error is negative, consisting in the lack of certain ideas. Here is what he says about it:

> There is nothing positive in ideas that makes them false. But error cannot consist merely in lack of knowledge as such, because we do not say that bodies err or are deceived. Nor is it merely ignorance as such (that is, lack of knowledge on the part of things that are capable of knowledge); because ignorance and error are different. So it consists in the lack of knowledge that exogenous ideas involve.[2]

Now, Spinoza is right that if error is ignorance, it must be some species of it. The differentia that he chooses—implying that it is the only option—is being caused from outside oneself. Why? Well, he describes exogenous ideas as 'mutilated, confused, and without order for the intellect', and the terms 'confused' and 'mutilated' come up repeatedly in his discussions of the senses. He has a sober reason for some of this. In any sensory encounter that you have with the world, he thinks, your body is interposed into the causal flow in a quite arbitrary way; the changes in it are truly necessitated by the bodies that impinge on it; but what changes they are depend upon accidents about where your body was placed, how it was oriented, how its most sensitive surfaces were textured, and so on; the upshot of these will typically be a random scatter of parts of the real truth about the situation into which you have intruded. Over the course of time you may be subject to a number of similar effects, and may even become able to predict some of them on the basis of others. But no amount of this will bring you down to bedrock; that is ruled out by the essentially arbitrary and fragmentary nature of every sensory encounter.

This is exaggerated, isn't it? Spinoza writes only about 'random experience', as he calls it. He is silent about the unrandom experience that comes in controlled scientific inquiry, and so does not comment on the fact that it too involves exogenous ideas. Anyway, that is where he stands:

[1] Ludwig Wittgenstein, *The Blue and Brown Books* (Oxford: Basil Blackwell, 1958), pp. 3 f.

[2] Spinoza, *Ethics*, 2p35d, lightly paraphrased for ease of understanding.

in the position of being much impressed by the thought that sensory input is bound to be 'mutilated and without order for the intellect'. (He throws in 'confused' for good measure. He is not entitled to it, but that is too long a story to tell here.)

So Spinoza has a sober, considerable line of thought leading him to conclude that so-called error results purely from the occurrence of mental contents which are mutilated etc., and those have to be caused from outside. From this it follows that reason—the causally self-contained movement of a mind from one state to another—cannot possibly generate error. This is not a believable theory of reason's infallibility, as I said; but I find it interesting and worth pondering.

6. Locke on modal discovery: the relevance problem

I turn now to my second topic: the view of early modern philosophers, and of ourselves, that we can use reason as a source of information about modal truth. The supposed link between necessity and reason was common property. Here it is in Descartes:

> We cannot determine by reason alone how big these pieces of matter are, or how fast they move, or what kinds of circle they describe. Since there are countless different configurations which God might have instituted here, experience alone must teach us which ones he actually selected in preference to the rest.[1]

To know what is actual out of many possibilities, Descartes says, we must have recourse to the senses; reason cannot do the job. But if there were only one possibility, reason might show us what it is. That is my topic—reason as a source of modal knowledge. The question is: how does reason give us such knowledge? What happens in the process that we describe as reason's leading us to modal truths?

Locke had an answer to this. According to him, we learn what is possible and what impossible by attending to relations amongst our ideas—so introspection is the way to modal knowledge. This is confronted by two problems, one famous and the other perhaps less so.

The more fundamental though less famous of the two is what I call *the relevance problem*. It asks how any fact about how my ideas are interrelated can have any bearing on any of the propositions that are ordinarily regarded as necessarily and eternally true. I am not asking how the interrelations of ideas can show the truth of such a proposition; before we come to that, there is the question of how a fact about such interrelations can point to any one modal proposition rather than to any other.

Locke has an answer to this in so far as it concerns geometrical propositions. We get a handle on the truth of these, he held, by seeing them actually instantiated by our ideas, these being images which themselves have geometrical properties:

> Is it true of the idea of a triangle that its three angles are equal to two right ones? It is true also of a triangle, wherever it really exists. (*Essay* IV.iv.6)
>
> He that hath got the idea of a triangle, and found the ways to measure its angles, and their magnitudes, is certain that its three angles are equal to two right ones. (*Essay* IV.xiii.3)

This assumes that mental images have sizes and shapes, which they do not. Also, we now know that the truths of Euclidean geometry are not absolutely necessary, so that this present line of thought does not even partly solve the relevance problem. Locke himself could not claim it as a total

[1] Descartes, *Principles of Philosophy* 3:46; CSM, vol. 1, p. 256.

solution, because he knew that plenty of necessary truths do not belong to geometry.

Furthermore, these geometrical propositions do not involve relations amongst ideas, as can be vividly seen in *Essay* IV.ii.2. Locke there undertakes to discuss 'the agreement or disagreement of. . . ideas', doing so in terms of an example concerning 'the agreement or disagreement in bigness between the three angles of a triangle and two right ones'.

For real modal truths, then, he must appeal to other ways in which ideas can represent. He has not much theory about this, except to say (in connection with secondary qualities) that an idea can represent an external quality by having a 'steady correspondence' with it (*Essay* II.xxx.2). Let us simply credit him with the view that an idea is a particular mental episode which somehow represents a property or quality that may be possessed by something else, with this representation depending on some kind of correlation between the represented property and the intrinsic nature of the idea.

Then his view must be this: the fact that my F-representing idea relates in a certain way to my G-representing one points to the proposition that necessarily whatever is F is also G. Relates in what way? I think it should be inclusion, but that is not what Locke says. Rather, he says repeatedly that we are led to modal knowledge through noticing the identity and diversity amongst our ideas. He might stretch this to cover idea-inclusion, I suppose, by saying that my G-representing idea is identical with a part of my F-representing one, though this puts most of its weight on the part/whole relation—a relation which therefore deserves more attention than Locke gives to it.

His use of the identity and diversity amongst ideas fails to serve his purposes in a much worse way than that. He wants it to be the basis for modal truths about impossibility—it is absolutely impossible that a G thing should be H—as we can see from his favourite examples: when we look inward we find that our idea of black is not our idea of white, that our idea of circle is not our idea of a triangle, and so on. Clearly Locke has his eye on the modal truth that something black cannot be white; but that does not follow from black's not being white, for if it did we could also prove that something black cannot be triangular, for those ideas are distinct also. Locke enables himself to overlook this by always illustrating 'x is not y' with values of x and y that are not merely distinct but incompatible; thus, he gets incompatibility into the reader's thoughts without explicitly mentioning it.

This is not a mere oversight, remediable by adding something to the account. If Locke tried to amplify his theory of modal knowledge by bringing in a relation of logical incompatibility—or something that could serve as the underlay for that—he would be defeated, for no such relation can obtain between particulars such as Lockean ideas are supposed to be.

He might try to steer around this by defining incompatibility through part/whole and negation: triangularity rules out squareness because the idea of triangular contains as a part the idea of not-square. But what account can Locke give of negative ideas?

This is not to inquire about negativeness as such. If we could make Lockean sense of each of a pair of ideas which represented logical complements of one another, we would not need to bother about which member of such a pair is positive and which is negative. Frege conjectured that there is no worthwhile concept of negativeness; but I have found hints of one in writings by Berkeley, Kant, and Ayer, and have developed them into something fairly substantial in work of

my own.[1] In our present context, though, negativeness does not matter: logical complementarity is all we need. Let us suppose that Locke has safely got on board the idea of *human*; how can he also make room for its logical complement, the idea of *not-human*? There are two ways he might go.

(1) He could try to devise an idea that represents non-humans in the way that the idea of human represents humans (or that of pebble represents pebbles, etc.). This requires a natural correlation—not necessarily a similarity—between the intrinsic features of the idea and the represented property. This is the property of non-humanity—whatever it is that is possessed by all coyotes and pebbles and lilies and neutron stars and whirlpools and by no human beings. The only mental property that is suitably correlated with that is the absence of whatever it is that is correlated with humans. That is Locke's opinion, too, it seems, for he would presumably call 'non-human' a negative term, and include it in his general statement: 'Negative or privative words. . . relate to positive ideas, and signify their absence.'

I cannot see that it matters much whether we call the mental state that lacks whatever represents humanity **(i)** the absence of an idea of humanity or **(ii)** the presence of an idea of non-humanity; but it is not surprising that Locke prefers formulation **(i)**. His natural discomfort about **(ii)** comes from a more general difficulty confronting his theory of mental representation—which I now explain. The theory is at its most comfortable when he is discussing ideas of superficially perceptible properties of things such as shapes. or dispositional properties where the disposition is, precisely. to cause a certain kind of idea—I refer here of course to the so-called secondary qualities such as colours. When it comes to ideas associated with terms such as 'human', 'house', and 'dandelion', the account comes under increasing strain, though Locke does not acknowledge it. One source of still further strain is the move towards ever greater generality: human, animal, organism, body. . . ; house, building, artifact, body. . . And it is a famous fact that when he came to the extreme of generality with *thing* or *substance* Locke openly proclaimed the suspect nature of the idea that he was nevertheless forced to postulate. Well, an idea of *non-human* would suffer from this problem of extreme generality, as would any other idea that we intuitively counted as negative; and so Locke's problem with them is just a special case of the generality problem.

Fortunately, we need not dig down into all that. The crucial point is just this. Consider the true proposition that it is absolutely impossible that a cannibal should be a vegetarian. According to the version of Locke's theory that I am now exploring, I can discover this to be true by inspecting my idea of cannibal, and discovering that it includes as a part my idea of non-vegetarian or—if you prefer—includes a part that lacks the representative features that suffice for representing vegetarian. Either way, it is true. But if that established the truth of 'No cannibal can be a vegetarian', we could also show the truth of countless falsehoods. For example, 'No cannibal can be a human', because my idea of cannibal contains as a part something which is not sufficient for representing a human, namely the part which represents eaters, or the part which represents animals. By these standards, no triangle can have three sides, because part of my idea of triangle is not sufficient to represent three-sidedness. And so on through endless other examples. In short, an idea merely *containing* a part which

[1] Jonathan Bennett, *The Act Itself* (Oxford University Press, 1996), chap. 6.

lacks a certain representative feature can never be enough to establish a proposition about impossibility.

(2) That was Locke's first option. The second is to try instead to get at non-humanity not through a suitably general representative idea but rather through an operation upon the idea of humanity. Locke provides for something like that in his doctrine about the meanings of particles:

> The mind in communicating its thoughts to others does not only need signs of the ideas it has then before it, but others also to show or intimate some particular action of its own at that time relating to those ideas. This it does several ways: as *is* and *is not* are the general marks of the mind, affirming or denying. (*Essay* III.vii.1)

Locke explicitly ties this to the meanings of words whose role is to link other words to make sentences, or to link sentences to make arguments and other discourses; but he might have been open to the suggestion that the same general approach could be applied to smaller linguistic units, including those that take one from a given classificatory general word to its logical complement.

That, however, would not combine well with the thesis that we learn modal truths by discovering how our ideas are interrelated. I am to learn that it is absolutely impossible that a cannibal should be a vegetarian by attending to my idea of cannibal and finding that it contains. . . what? The item that you get through a negating operation on the idea of vegetarian? How did it get there? Had I already performed the negation operation and left its upshot sitting there within my idea of cannibal? I can find no way of telling this story without making it seem ludicrous and unbelievable, even in the eyes of someone who is not sceptical about Lockean ideas in a general way, and has no discomfort about such ideas as those of humanity and animality.

I should add that the theory of 'particles' looks apt to be useful for the most general modal truths, for which Locke's system of classificatory ideas is quite useless—for example the proposition that if (if P then not-P) then not-P. But here again the theory that modal truths are learned by introspection seems to be pushed aside. What would we be introspecting?

7. Locke on modal discovery: the contingency problem

Now, forget all that, or suppose the relevance problem to be solved, e.g. by focusing on geometrical propositions and pretending to think that they can be read off from geometrical properties possessed by ideas themselves. That frees you to attend to the contingency problem. Leibniz brought this to the fore in his complaint that Locke's procedure of attending to particular ideas could establish only contingent truths, and that knowledge about absolute necessity cannot be arrived at through such empirical means.

The core of the difficulty has been nicely stated by Michael Ayers: I find in my mind a particular image of a triangle, and perceive that it has (or is an image of something that has) internal angles equal to two right-angles; but to get a general proposition out of this I need to know that the image has that property *purely because* it is an image of a triangle. Locke does not try to explain how I could perceive that.[1]

Locke, when discussing a different problem, says something that could be a response to Leibniz's criticism, namely that the eternity of the truths we learn from inspecting ideas is ensured by the fact that 'The same idea will eternally have the same habitudes and relations' (*Essay* IV.i.9). His

[1] Michael Ayers, *Locke* (London, Routledge, 1991), vol. I, p. 255.

only grounding for this, however, is a remark earlier in the same section about 'the immutability of the same relations between the same immutable things'. But Lockean ideas are not immutable things; they are dated and mentally located psychological particulars. Locke tells us this clearly and often; and, anyway, if ideas were not like that, how could we examine them by looking into ourselves?

Later in Book IV, Locke returns to the problem of eternal truths while holding fast to the status of ideas as psychological particulars.[1] I quote him as briefly as I can:

> Universal and certain... knowledge is the consequence of the ideas... that are in our mind producing there certain general propositions... Whatsoever we can suppose such a creature as man is, endowed with such faculties and thereby furnished with such ideas as we have, we must conclude he must needs, when he applies his thoughts to the consideration of his ideas, know the truth of certain propositions that will arise from the agreement or disagreement he will perceive in his own ideas. Such propositions are therefore called *eternal truths*... because once made about abstract ideas so as to be true, they will, whenever they can be supposed to be made again at any time past or to come, by a mind having those ideas, always actually be true. For names being supposed to stand perpetually for the same ideas, and the same ideas having immutably the same habitudes to one another, propositions concerning any abstract ideas, that are once true, must needs be eternal verities. (*Essay* IV.xi.131.)

The clearest part of this is also the weakest. When Locke says that names are 'supposed to stand perpetually for the same ideas', we can say exactly how much he achieves: namely that a sentence which now expresses a truth will always express that same proposition. This does not secure that the proposition will always be true.

For the rest, Locke says that an event in the mind of one person at one time will be duplicated in the mind of a relevantly similar person at any other time, and we can accept this. But he says it in the language of discovery or even of making-true: if events in my mind teach me that P or make it the case that P, the mind of any similar person can or will be the scene of similar events—ones in which P's truth will also be revealed or created. If the point concerns discovery, Locke needs and does not have an account or what the initial discovery consists in. If it concerns making-true ('being once made about abstract ideas, so as to be true', 'to come, by a mind having those ideas, always actually be true'), he is even further from having explained what he ought to explain. This is, I think a mixture of the relevance and contingency problems.

8. Leibniz on inner and outer

Leibniz was onto something, then. Locke's account of modal knowledge has no sound defence against the accusation that it reduces all necessities to contingencies. For much of the time, however, Leibniz failed to get this criticism properly into focus. Things start to become blurry when he writes that 'necessary truths... are proved by what lies within, and cannot be established by experience as truths of facts are'.[2] This relies on the contrast between 'what lies within' and 'experience'; but Locke has been basing his

[1] *Essay* IV.iii.29 also looks relevant; but Locke there adduces eternity and immutability only as a challenge to Descartes's voluntarism about modal truths; he does not treat them as problematic for himself.

[2] Leibniz, *New Essays on Human Understanding*, ed. and tr. P. Remnant and J. Bennett (Cambridge University Press, 1983), p. 79.

modal epistemology on, precisely, experience of what lies within.

Leibniz becomes uncomfortable about this, but not enough to be driven to get clear about it. He remarks that Locke first rejects innate ideas and later seems to espouse them: 'Perhaps our gifted author will not entirely disagree with my view. . . He admits at the start of his second book. . . that ideas which do not originate in sensation come from reflection. But reflection is nothing but attention to what is within us, and the senses do not give us what we carry with us already' (ibid., p. 51). There is a mistake here. The notion of 'what we carry with us' that Leibniz hopes to profit from does not include the casual psychological episodes that Locke calls 'ideas'. More generally, the necessary/contingent line could not possibly coincide neatly with the line between inner and outer.

The reason why 'the senses are inadequate to show. . . necessity' is not that the senses look outwards, but that they inform us only about particular instances. Leibniz says as much:

> Although the senses are necessary for all our actual knowledge, they are not sufficient to provide it all, since they never give us anything but instances, that is, particular or singular truths. But however many instances confirm a general truth, they do not suffice to establish its universal necessity; for it does not follow that what has happened will always happen in the same way. (Ibid., pp. 49f.)

That is true whether the 'instances' are inner or outer, and so if Leibniz has a good point here it cannot depend on the latter difference.

The emphasis on particulars is Locke's as well as Leibniz's and mine. In Essay IV.vii he scornfully discusses what he calls 'maxims'—general propositions which have been taken to be innate and to be in some way the foundation of all our knowledge. His candidates for the title 'maxim' are all highly general—'The whole is greater than the part', 'It is impossible for the same thing to be and not to be', and so on—and he denies that these are the sources for our knowledge that my body is larger than my finger and that whatever is white is not unwhite. He argues 'that such self-evident truths must first be known which consist of ideas that are first in the mind; and the ideas first in the mind. . . are those of particular things'.

9. Leibniz and the relevance problem

Throughout Book I of the *New Essays* Leibniz seems to propose a modal epistemology in which the relevance problem is solved and the contingency problem does not even arise. It holds that we learn modal truths because they are engraved on our souls. To take an example that would have challenged Locke:

Q: If (if P then not-P) then not-P.

I learn that Q is true, according to this theory of Leibniz's, by finding it written on my soul.

This does not involve a relevance problem. Where Locke speaks of looking in and finding psychological states of affairs which somehow point to the truth of Q—the relevance problem being the question of what the 'pointing' is—Leibniz's theory of soul-writing says that we look in and find Q itself.

Or so one might think, but let us not go too fast. What is it to find a proposition in my soul? It might be to introspect and discover that I have a certain belief: I find Q in there by finding myself believing that Q. That seems not to be Leibniz's principal view, however. As his metaphor about writing or engraving implies, he apparently holds that in many and perhaps most cases what is written on the soul is something which means a modal proposition—a sentence

in soul-script, as it were—and that might seem to re-raise the relevance problem. For Locke it was the question of how a psychological particular can *point to* any one universal proposition; now Leibniz confronts the question of how a psychological particular can *mean* a universal proposition. Isn't that just as bad?

Leibniz has nothing to say about this, and seems not to have noticed it. He does say that some soul-sentences are more 'legible' (p. 76) than others; but his topic there is one's awareness of the sentence, not one's knowledge of what it means. However, it is not really on a par with Locke's relevance problem. It raises a more general question about linguistic meaning, which arises for us all. If it upsets the soul-writing answer to the question of modal epistemology, then it also makes trouble for the question itself; for that is stated in a sentence, which we think we understand.

Still, a question remains. Granted that a given soul-sentence means Q. why should its presence in my soul count us showing me that Q is true? What if I found it inscribed instead on a tree-trunk or in the sand on the beach? We know how Leibniz would answer this: the sentence is written on my soul because God wrote it there, and God can be trusted not to write lies in people. This seems reasonable. If I believed in God at all, I would believe that much about him.

One might object that until Leibniz knows some modal truths he cannot justify his belief in a truthful God, so that in this matter his procedure is circular. Well, so it would be if he were aiming to establish a modal epistemology from a starting-point that assumes nothing about what is possible or necessary. But he may not have been attempting that; indeed it may be that no such attempt could succeed; yet there could still be an epistemology of modality. I have learned this from my colleague William Alston.[1] Consider the question of how we discover how matter is distributed at the actual world. The right answer includes the thesis that material things leave informative traces of their action upon us. That thesis is the core of an excellent account of how we are informed about the material world at which we live; but our evidence for it relies on things we believe about the material world. Such 'epistemic circularity'—Alston's phrase—is sometimes inevitable, and it is not fatal. So the written-by-God theory may be a coherent epistemology of modal truth.

Still even someone who believes in Leibniz's God, knows what he means by 'inscribed in the soul', and accepts this whole story, ought to find it disappointing because it passes on the epistemological problem from humans to another person. I might get my basic modal beliefs from my brother, whom I trust; but obviously that would not be a down-payment on a decent epistemology of modality. Well, when someone tells us that God told him the basic truths about what is possible and impossible, should we not react in the same manner? Perhaps not. A believer might think that God knows everything and that this is a basic fact about Him, not the upshot of any epistemic modes, ways or means that might be the topic of an explanatory theory. But if that is your view, and if you also hold that your best explanation of our modal knowledge is that God handed it to us on a plate, you ought as a philosopher to he disappointed in this state of affairs. Even if this is the entire truth of the matter, that it is so is a matter for regret.

[1] W. P. Alston, *The Reliability of Sense Perception* (Ithaca: Cornell University Press, 1993), chap. 2.

10. Rationalists, empiricists, and silver spoons

There is a real difference between the view that I find a proposition inscribed on my mind and the view that I find in my mind the materials which satisfy me that the proposition is true. It is characteristic of Locke to prefer the latter to the former; he sees all our knowledge as having to be worked for—he holds, in Aaron's words, that 'Knowledge is always discovery'.[1] Thus, his fundamental complaint against innatism is that it would give us epistemic possessions that we have not worked for. On this topic—the work-shy nature of innatism—he is eloquent:

> We may as well think the use of reason necessary to make our eyes discover visible objects, as that there should be need of reason, or the exercise thereof, to make the understanding see what is originally engraven in it. (*Essay* 1.ii.9)
>
> There is a great deal of difference between an innate law and a law of nature; between something imprinted on our minds in their very original, and something that we being ignorant of may attain to the knowledge of by the use. . . of our natural faculties. (*Essay* I.iii.13)

Price elegantly described this aspect of Locke's thought. What he says about the acquiring of ideas holds even more thoroughly for the learning of eternal truths:

> It is, of course, historically false that the Empiricists thought the human mind passive. It would be more just to criticize them for making it more active than it can possibly be. It is the Rationalist Mind, if either, which is the passive one, or at least the lazy one, born, if one may say so, with a silver spoon in its mouth. The Empiricist Mind has to acquire these basic ideas for itself. . . by its own effort and initiative.[2]

Gibson made the same point, warning us against being led by Locke's metaphor of the 'white paper'[3] to think that he sees the mind as generally passive:

> The upholders of the theory [Locke] opposes commonly employed the metaphor of the stamp and its impression in describing the source of innate principles. . . Indeed, so far as the question of mental activity is involved in the controversy at all, one of Locke's objections to the theory he opposes is that it represents certain truths as merely given to the mind, apart from the exercise of that active comparison and examination which he holds to be involved in all human knowledge.[4]

It is true that Leibniz also stresses the need for work, and rails against those who use the doctrine of innateness as an excuse for laziness and dogmatism (*New Essays*, pp. 50, 85). But the work he calls for is proving whatever can be proved, using as premises those basic innate truths 'which can be neither doubted nor proved' (pp. 75, 91, 108). The latter, according to Leibniz, are just given—they are among the 'writings in inner light' which 'sparkle continuously in the understanding' (p. 100).

11. Leibniz and the mind of God

Later in the *New Essays*, Leibniz backs off from the modal epistemology which he has seemed to accept in the work's

[1] R. I. Aaron, *John Locke* (3rd edn, Oxford University Press, 1970), p. 97.

[2] H. H. Price, *Thinking and Experience* (London, 1953), p. 199n.

[3] Did you think that Locke uses *tabula rasa*, or some English equivalent thereof, somewhere in the *Essay*? The belief that he does is one of the persistent little myths of the history of philosophy.

[4] James Gibson, *Locke's Theory of Knowledge and its Historical Relations* (Cambridge University Press, 1917), pp. 32 f.

opening chapters. He does this with help from his metaphysic of modality—his account of what the truth-makers are for modal propositions. If Locke had such a metaphysic, it must have been the view that modal truths are made true by facts about our ideas, which invites the charge that he has made them psychological and contingent.

Leibniz is warier on this topic:

> Eternal truths are fundamentally all conditional. For instance, when I say: *Any figure which has three sides will also have three angles*, I am saying only: given that there is a figure with three sides, that same figure will have three angles. How can a proposition about a subject have a real truth if the subject does not exist? The answer is that its truth is a merely conditional one which says that if the subject ever does exist it will be found to be thus and so. What is the ground for this connection? The reply is that it is grounded in the linking together of ideas. Where would these ideas be if there were no mind? and what would then become of the real foundation of this certainty of eternal truths? This question brings us at last to the ultimate foundation of truth, namely to that Supreme and Universal Mind who cannot fail to exist and whose understanding is indeed the domain of eternal truths.[1]

This account is as psychologistic as Locke's: the ideas in question are in minds—some in God's, others in ours, But for Leibniz this does not revive the contingency problem, because his theology is, he thinks, absolutely necessary. Although the truth-makers for modal propositions are relations amongst mental particulars, the latter absolutely must exist and be interrelated as in fact they are; so they are eternal and necessary, as rock-hard and durable (logically speaking) as relations amongst the items in Frege's 'third realm'.

Although Leibniz thought of God as personal, as caring for us, and as a fit object of reverence and love, when he writes of God's intellect he makes Him sound like an abstract object. In these contexts, his language is—or anyway his metaphors are—notably Fregean: God's understanding is 'the domain [*région*] of eternal truths', 'the divine understanding is, so to speak, the realm [*pays*] of possible realities'.[2] 'These essences and the so-called eternal truths about them,. . . exist in a certain region [*regio*, Latin] of ideas, if I may so call it, namely in God himself'. Compare that with Frege's 'third realm' and indeed with Wittgenstein's 'logical space'.

12. Leibniz's second epistemology of modality

So much for the metaphysic. Now for the epistemology—I mean the one that does not involve soul-writing 'When God displays a truth to us,' Leibniz writes in Book IV, 'we come to possess the truth which is in his understanding, for although his ideas are infinitely more perfect than ours, they still have the same relationships that ours do' (*New Essays*, p. 397), This relies on the metaphysic of modality that I have just presented: the relationships amongst God's ideas make modal truths true; and an isomorphism between our minds and God's enables us to discover which propositions are necessarily true. There is nothing here about truths inscribed on the soul. Leibniz is explicit about that. Just after presenting the divine-psychology metaphysic, he writes that the mind of God 'is where I find the pattern for the

[1] Leibniz, *New Essays*, pp. 446–7. Did you expect him to invoke possible worlds? That Leibniz explained necessity in that way is another contemporary myth; he never did so.

[2] G. W. Leibniz, *Philosophical Papers and Letters*, ed. Leroy E. Loemker (Dordrecht: Reidel, 1969), p. 336; the next quotation is from ibid., p. 488.

ideas and truths which are engraved in our souls', and goes on to explain: 'They are engraved there not in the form of propositions, but rather as sources which, by being employed in particular circumstances, will give rise to actual assertions.' So they are not engraved there as propositions! This position of Leibniz's does not have the disappointing feature of the 'God told me and I believe him' theory which he seemed to advance earlier; but, unlike that, it does re-raise the relevance problem.

13. Leibniz's relevance problem

The question is: what do relations amongst the ideas in a mind have to do with such propositions as that if (if P then not-P) then not-P? Leibniz cannot brush this off with the remark that the ideas in question are in the mind of God, and that we cannot be expected to grasp what they are or how they do what they do. He has said that the relations amongst our ideas are isomorphic with relations amongst God's, so he is obliged to have some account of what these relations are, and of what they have to do with modal truths.

The problem arose for Locke in an especially acute form because his 'ideas' are supposed to be images, and the relevance of those to modal truths is especially hard to see. Leibniz is free of that trouble, at least. At intervals throughout the *New Essays* he separates ideas from images, and rightly accuses Locke of smudging the line between them. In reply to Locke's saying that one does not have a precise idea of a thousand-sided figure that would let one distinguish it from one that has one side fewer, Leibniz writes:

> That example shows that the idea is being confounded with the image. If I am confronted with a regular polygon, my eyesight and my imagination cannot give me a grasp of the thousand which it involves: I have only a confused idea both of the figure and of its number until I distinguish the number by counting. But once I have found the number, I know the given polygon's nature and properties very well, in so far as they are those of a chiliagon. The upshot is that I have this idea of a chiliagon, even though I cannot have the image of one. (*New Essays*, p. 261)

This fits with Leibniz's general practice of crediting a person with having a certain idea if he is relevantly competent in some intellectual matter. That matches a way we have today of talking about the 'concepts' that people have, and I have no complaint with it in itself.

But while that explains what it is to say 'He has an idea of x', it gives us no help in grasping 'idea' standing on its own. Yet that is what we need to make sense of what Leibniz says about the relations amongst our ideas. For ideas to be relata, they must he distinguishable, countable, identifiable items of some kind. Leibnizian 'ideas' are not images; and are durable dispositions rather than episodes as Lockean 'ideas' are. What sorts of inter-relatable items can they be? The best we can do is to say that they are competences; my idea of chiliagon is my competence in thinking about chiliagons. That, however, will not serve in Leibniz's modal metaphysic and epistemology: it is perfectly unclear what the supposed relations amongst competences could be; and Leibniz would blush to say that I know what is necessarily true because my competences relate to one another in the same way that God's do.

14. Ideas: Fregean or psychological?

Leibniz sometimes seems to understand the term 'idea' differently. Responding to Locke's statement (which I don't think accurately expressed Locke's own views) that an idea is an object of an act of thinking, Leibniz comments:

> I agree about that, provided that you add that an idea is an immediate inner object, and that this object expresses the nature or qualities of things. If the idea were the form of the thought, it would come into and go out of existence with the actual thoughts which correspond to it, but since it is the object of thought it can exist before and after the thoughts. (*New Essays*, p. 109)

Perhaps these 'objects' of thoughts are items that could be inter-related suitably. They certainly could if they are what Leibniz was referring to in a dismissive comment on Spinoza's view that an animal's mind is the idea of its body: 'Ideas are purely abstract things, like numbers and shapes, and cannot act. Ideas are abstract and universal: the idea of any animal is a possibility.'[1] Relations amongst possibilities are just what we need as a foundation for modal truth! But when the term 'idea' is understood in this manner, Leibniz's account of how we get modal knowledge is destroyed. That account makes sense only if ideas are psychological and personally owned, as Leibniz usually held them to be. Here, for instance: '[Ideas] are affections or modifications of our mind. . . For certainly there must be some change in our mind when we have some thoughts and then others.[2]

It looks as though the most Leibniz can salvage from this second theory about modal knowledge is this: the truth-makers for modal propositions are eternally and necessarily existing items (in the third realm or the mind of God—it no longer matters which); and we are capable of thoughts which somehow map onto, or at least inform us about, relations amongst those items. That weak offering is about all that we have today, isn't it? I have encountered philosophers who say that they do so too have an epistemology of modality: 'We learn what is absolutely necessary or possible through our modal intuitions.' But they do not offer details about what those intuitions are, or about why they are pointers to the truth. And I am pretty sure that there are no such details to be given. It seems to me that our modal intuitions are not a basis for our modal opinions; they are our modal opinions. so that the epistemic 'theory' which takes them as our basis is empty, is not a theory at all. That is where Leibniz ended up, and is where we are still today.

15. Idealism about modal knowledge

Our lack of a half-way decent account of modal knowledge is one reason for a different metaphysic of modality, specifically one of the sort that Tyler Burge has called 'idealist'.[3] The thought is that we shall be less cut off from the truth-makers of modal truths if these are somehow not about a third realm but about ourselves. One might see Locke as pushing in that direction, but that would be a whitewash, I believe. I can find no evidence of his having considered this matter and come to the reasoned conclusion that modal truths are only a projection or reflection of facts about human language and/or thought.

Of the famous early modern philosophers, the one who most openly and explicitly—though also briefly—did assert that view was Descartes. Sometimes he sounds like Locke, merely confusing or running together logical with psychological propositions, as when he writes that 'Each of us can see

[1] Leibniz, 'Comments on Spinoza's Philosophy' (1707?), in R. Ariew and D. Garber (eds), G. Ẇ Leibniz: *Philosophical Essays* (Indianapolis: Hackett, 1989), at p. 277.

[2] Leibniz, 'Meditations on Knowledge, Truth, and Ideas', in Ariew and Garber's anthology at p. 27.

[3] Tyler Burge, 'Frege on Knowing the Third Realm', *Mind*, 101 (1992), 633–50.

by intuition that he exists, that he thinks, that the triangle is bounded by three lines only' (CSM, vol. 1, p. 14), and so on. But in one place he does something utterly different.

He is confronting critics who have questioned whether the concept of God used in his *a priori* argument for God's existence is a possible one. Here is Descartes's striking response:

> If by *possible* you mean what everyone commonly means, namely whatever does not conflict with our human concepts, then it is manifest that the nature of God, as I have described it, is possible in this sense because. . . [etc., etc.] Alternatively, you may well be inventing some other kind of possibility which relates to the object itself, but unless this matches the first sort of possibility it can never be known by the human intellect, and so it. . . will undermine the whole of human knowledge.[1]

This subjective concept of possibility, which makes it a relation to our concepts, is the common meaning for the term 'possible', Descartes says; whereas the objective concept of a 'possibility which relates to the object itself' is a contrivance, something faked up for purposes of argument rather than part of our natural conceptual repertoire (he uses the Latin verb *fingo*, which is the source of 'feign' and 'fiction'). Of course a technical concept might be better than a natural, informal one, but not in this case. The objective concept cannot have a life of its own, Descartes declares: if it does not keep in step with the subjective one it will be direly subversive, because it will land us with. . . just precisely the problem of modal epistemology that we are still wrestling with.

So we have Descartes here announcing and defending an analysis of modality, a conceptualist analysis—taking 'concepts' to be aspects of the human condition, of course, and not entities belonging to a Fregean third realm.

This aspect of Descartes's thought seems not to have been adequately noticed in the secondary literature, though it has been properly highlighted by Nicholas Jolley.[2] But even he fails to notice that Descartes's subjectivism about modality helps greatly with his voluntarism, his doctrine that God chose which propositions should be necessarily or eternally true. Scholars who are generally friendly to Descartes have described this doctrine as (in alphabetical order) bizarre, curious, incoherent, peculiar, and strange—and the first and last of those adjectives comes from Jolley.[3] But really Descartes's voluntarism falls into place, once his subjectivism about modality is grasped. P's being necessary is a fact about how it relates to the limits of human thought: God made us and gave us our limits; in so doing He determined which propositions would be necessary or eternally true. What looked like madly extravagant theology turns out to be a combination of a sober conceptual analysis and a routine application of the theology of creation. The analysis is what is interesting, of course, not the theology.

That is enough about Descartes's voluntarism. It is indeed enough altogether. I end with a philosophical remark of my own, namely that the epistemological problem as traditionally understood seems to me quite insoluble, from which I conclude—with Descartes—that the problem is a mistake, being based on a wrong view of what modality is. I think we shall have to return to the 'idealist' or subjectivist approach, trying to show how it can be the case that—in a phrase I think I got from Stalnaker—all the possible worlds are at the actual world.

[1] Descartes, Replies to Second Objections, CSM, vol. 2, p. 107.

[2] Nicholas Jolley, *The Light of the Soul: Theories of Ideas in Leibniz, Malebranche, and Descartes* (Oxford University Press, 1990).

[3] Ibid., pp. 32 and 166f; Loeb produced 'peculiar' and 'curious'; 'incoherent' comes from Curley.