

**DAVID J. BARTHOLOMEW**

## **Victor Stenger's Scientific Critique of Christian Belief**

---

*In two recent books, Victor Stenger claims to show that, using the scientific method, it is possible to show that the 'God Hypothesis' must be rejected. To a large extent his refutation is based on the use of ideas of statistical inference. The purpose of this paper is to show that the scientific method is incapable of achieving the goals set for it by Stenger and that, in particular cases, his use of it is fallacious. We deal first with intercessory prayer experiments and then with his understanding of statistical significance, meta-analysis and scientific sampling. In conclusion it is pointed out that a rigorous use of scientific method must include all the evidence which, in the case of Christianity, involves a serious examination of the evidence relating to the incarnation.*

**Key words:** design of experiments, incarnation, meta-analysis, observational data, prayer, scientific sampling, significance testing, Stenger

---

### **Introduction**

The Science and Religion debate is usually about reconciling the world-views presented by science and religion. But there is a related way of using science to attack religion of which Victor Stenger is a principal exponent. He opposes the idea that science and religion belong in separate compartments and that neither can challenge the other. On the contrary, Stenger believes that the God Hypothesis, as he calls it, makes empirical predictions about the world which can be tested by scientific methods. Further, he argues, unless God's existence makes some difference to the world he might just as well not exist! In his own words:

However, independent of the intentions of a hypothetical being of infinite power and wisdom, objective evidence for an entity with godlike attributes should be readily available. After all, God is supposed to play a decisive role in every happening in the world. Surely we should see some sign of that in objective observations made by our eyes and ears, and especially by our most sensitive scientific instruments.

The founders and leaders of major religions have always claimed that God can be seen in the world around us. In Romans 1:20, St Paul says: 'Ever since the creation of the world his invisible nature, namely his eternal power and deity, have been clearly perceived in the things that have been made.' We will look for evidence of God in the things that have been made.<sup>1</sup>

---

<sup>1</sup> Stenger, V.L. *God, the failed hypothesis: how science shows that God does not exist*, New York: Prometheus Books (2008), p. 22.

In other words, according to Stenger, there are many things which Christians believe which can be subjected to empirical testing; and he claims that in every case the result is negative. By this means, the whole edifice of belief in God collapses before the onslaught of scientific method. In short, the God hypothesis must be rejected.

The purpose of this paper is to show that Stenger's method fails, not because it is too scientific, but because it is not scientific enough. I shall meet Stenger on his own ground and show that his methods are incapable of answering the questions he poses, because the requirements of the rigorous scientific approach that he advocates are not met. In essence this is because one cannot fully control the experimental situation in the manner required. Stenger is operating in an area for which his training in experimental science – at the 'hard' end of the science spectrum – is hardly relevant. He is attempting to answer questions which are more properly regarded as within the ambit of social science where observational as opposed to experimental data are the norm.

Stenger is not alone in misunderstanding the nature of scientific method as applied to observational data. In a recent book,<sup>2</sup> Robert L. Park, another physicist, makes the same errors though in a somewhat more circumspect manner.

I write, not as a physicist, but as a statistical scientist who has spent most of his working life getting to grips with the subtleties of the kind of observational data which abound in the social sciences of which Stenger and his kind seem largely unaware.

This paper is necessarily selective and I have chosen four themes which are sufficient to undermine the confidence that pervades Stenger's treatment. The first is the effect of intercessory prayer where the double-blind randomised trial seems exactly what is needed to establish whether prayer really has any effect. I shall show that Stenger's error here has far-reaching implications for the possibility of detecting divine action. Secondly, I shall show that Stenger is simply incorrect, in a fundamental way, in his exposition of statistical inference which plays a central part in many of his conclusions. Thirdly, I shall expose Stenger's misunderstanding of meta-analysis. In conclusion I shall point out that the lack of rigorous sampling procedures places serious question marks against many of the conclusions which Stenger draws from social surveys.

### **Does intercessory prayer work?**

The insertion of the adjective *intercessory* is important if only to dispel the notion, common among atheists, that this is the only kind of prayer in which Christians engage. However, it is certainly relevant for the attempted test of

---

<sup>2</sup> Park, R.L. *Superstition: belief in the age of science*, Princeton University Press (2008).

the hypothesis that God acts in the world. At its simplest it involves asking God to do something and then observing whether or not it happens.

There is, of course, a vast amount of evidence on the matter, most of which would be classed by Stenger as anecdotal and therefore irrelevant. For the moment we accept this stricture and start where Stenger does when he says: 'Surely, with the millions of prayers being submitted daily, totalling billions in recorded history, some objectively verifiable (not just anecdotal) positive evidence should have been found by now!'<sup>3</sup> He justifies subjecting prayer to scientific testing by continuing: 'However, anything with observable consequences is testable by scientific means, and prayer is widely believed to have observable consequences.'<sup>4</sup>

Although there have been a number of prayer experiments that have produced positive results, some of which Stenger quotes, there are only two that appear to measure up to the full requirements of rigorous scientific analysis. Stenger regards the position as essentially the same as it is when testing the efficacy of a new drug for human application, using a double-blind randomised experiment. This type of experiment is fundamental to Stenger's approach so it will be examined in more detail in the following section.

The two experiments that Stenger judges to have passed these basic tests are the Duke study and the STEP project. He gave full details in 2008<sup>5</sup> but he had considered them in a paper written in 2003.<sup>6</sup> The first experiment was conducted by Christian physicians who hoped, Stenger presumes, that the experiment would show a positive result for prayer. Initially, before all the results were in, it did appear that this would be the case but by the end of the trial no effect was observed. This trial involved 748 patients at 9 hospitals in the USA and included other non-physical methods of healing such as music and touch therapy.

The STEP project was on a larger scale, involving 1802 patients receiving a coronary bypass graft and it extended over a decade. Two Catholic groups and one Protestant group were involved and in some cases the patients knew that they were being prayed for, while in others only that they might or might not be being prayed for. The details are not important because we are about to argue that both studies were misconceived if seen as tests of the efficacy of prayer. Suffice it to say that Stenger was satisfied with the protocols and that, in neither case, were any significant results obtained.

The first point to notice is that even if significant results had been obtained, it still could not have been concluded that God was acting. Both trials could

---

3 Stenger *op. cit.*, (1), p. 94.

4 *ibid.*, p. 94.

5 Stenger *op. cit.*, (1), pp. 99-102.

6 Stenger, V.L. *Has science found God?: the latest results in the search for purpose in the universe*, Amherst, New York: Prometheus Books (2003).

equally have been regarded as tests of the effect of ‘mind over matter’ or of some healing effect through extrasensory means. This would probably not have greatly concerned Stenger, who is equally sceptical of all kinds of non-physical influences, but it should have given believing supporters pause for thought before they committed their resources.

However, this is not the main scientific objection to studies of this kind. The model of a clinical trial is inadequate because, with praying, there are three parties, not two, involved. In a clinical trial the only two parties are the patient/subject and the experimenter/doctor and neither party knows whether the drug is being administered or not. In the case of prayer, God is a third party. For if he were not involved it would not be prayer that was being offered! The set-up of all trials of this kind seems to presume that God is completely passive in that he merely acts as an intermediary doing the supplicant’s bidding. In effect, God is being used as a tool to be manipulated by those who are praying so that he is, therefore, no more than a puppet in their hands. This picture of God does not measure up to any orthodox view of God. Stenger himself seems to be half aware of this when, tongue in cheek perhaps, he remarks ‘No one suggested that God was deliberately thwarting the expectations of the researchers.’<sup>7</sup> Taken seriously, this remark should at least raise the question of whether God might be playing an active, rather than a passive role. If that was the case, all the experiment could show was that God, if he exists, could not be relied on to do exactly what his petitioners requested. Stenger should have noted that number 3 in his list of the attributes in the scientific model of God, beginning on page 41 of *God the failed hypothesis*, includes the statement ‘God steps in *whenever he wishes* to change the course of events... in response to human entreaties’ (my emphasis). Christians would also have in mind the text ‘Jesus said to him, “Again it is written, Do not put the Lord your God to the test”’ (Matt. 4:7 NRSV).

These points have been made before – see Bartholomew *God of Chance*<sup>8</sup> and *Uncertain Belief*<sup>9</sup> – but both of these contributions were written before the studies referred to above were carried out. They now have added point in view of the continued funding of such studies by those who expect positive results to confirm Christian expectations and of the criticisms of sceptics like Stenger.

Many of Stenger’s views on these studies are echoed by Park<sup>10</sup> though in a less aggressive fashion. Having drawn attention to the superstition of prayer as he sees it, Park concludes by saying that such studies fail because there is no *metric* for prayer. This appears to mean that there is no way we can measure the effect of prayer. In a broad sense this is no doubt true, but in the con-

---

7 Stenger *op. cit.* (1), p. 102.

8 Bartholomew, D.J. *God of chance*, London: SCM Press (1984), pp. 159ff. (e-version available free of charge at [www.godofchance.com](http://www.godofchance.com)).

9 Bartholomew, D.J. *Uncertain belief*, Oxford: Oxford University Press (1996), pp. 103-104.

10 Park *op. cit.*, (2).

text of these studies there was a very specific metric in that the outcome was recorded as producing an indisputable quantum change. This might only be a binary response, which is the crudest form of metric, but it is a metric nevertheless.

Stenger concludes his discussion in a form that echoes his other conclusions and underlines the main thesis of his book that 'It does not appear – based on scientific evidence – that a God exists who answers prayer in any significant, observable way.'<sup>11</sup> Contrary to this I have concluded that there is no truly scientific evidence, nor can there be, that justifies such a conclusion.

Believer and unbeliever alike might well pause to reflect at this point on where this leaves us. Regarding God anthropomorphically for the moment, it appears that he has been put in a no-win situation. If he acts in answer to experimental prayer, he is denying his nature by becoming the tool of humans; if he fails to act he is judged to be non-existent!

### **A digression on experimentation**

The original and basic idea behind experimentation is that if you vary one factor at a time, holding all others constant, then any effect you observe must be due to that factor. To test the efficacy of a drug we might therefore administer it to a sample of people and compare the result with the effect of giving nothing to a control group. This is the idea behind the rigorous scientific testing advocated by Stenger and strongly approved by the likes of Dawkins. But though it might serve tolerably well when experimenting on inanimate objects, it needs elaborating before it can work in the case of something like drug testing on human subjects. The placebo effect is well known, for example. The mere act of being given an inert tablet may, of itself, produce a change in what is observed. So the subjects must not know whether what they are taking is the real thing, or not. In addition, those administering the drug may unconsciously convey relevant information to the patient so it is important that they shall not know whether what they are administering is the drug or not. But all these precautions are not enough. The patients will almost certainly differ among themselves in their response to any treatment. Some people are particularly sensitive and a given dose may have a much larger effect on them than on others. Suppose it so happened that those allocated to one group were, purely by accident, more responsive. The experiment would then be misleading in that the result would be far from typical. Is there any way that this can be avoided? There is no absolute guarantee in this matter but there is a way in which we can ensure with high probability that we are not misled.

The answer is provided by randomisation. In the experiment we are considering this means allocating subjects between the experimental and control

---

11 Stenger *op. cit.* (1), p. 102.

groups at random. In practice this could be done by the toss of a coin or some equivalent process. Although there is still a small chance of getting a biased allocation, the chance of doing so is very small. Furthermore it is virtually inconceivable that the same bias would be repeated if the experiment were to be replicated several times. When all these precautions are in place we have what is called a double-blind randomised experiment. This is the norm in drug testing and this is what Stenger rightly demanded when testing the effects of prayer.

We can actually go much farther than this and test for the effect of several treatments simultaneously. Testing one treatment at a time can be very costly in time and money and it was one of Sir Ronald Fisher's<sup>12</sup> contributions to experimentation that he showed how to do this. Instead of administering a single treatment to each subject we now administer a combination of treatments. Which combinations, of course, have to be carefully worked out and there must also be appropriate randomisation. There is a double benefit from testing several factors simultaneously. Not only do we test the main effects, as they are called, but also the interactions. The idea of interaction is very important and frequently ignored. The effect of one drug, for example, may depend on what other drugs are being tested at the same time and this is something we should never discover unless we were using a multi-factor experiment. Several factors were, in fact, included in one of the prayer experiments where some subjects knew they were being prayed for and others only knew that it was a possibility.

Unfortunately it is hard to find any situations of interest in the science-religion debate where experimental design can be validly used. It is the failure of Stenger and his like-minded followers to recognise this that undermines their case. We have already pointed out that God is an uncontrolled factor in the prayer experiments of which Stenger approves, and hence, that no validity can be attached to their conclusions. This situation is the norm once we move outside the experimental sciences into the world where we cannot control, or average out, all the factors. When we bring people into the equation, matters become even more complicated, as we shall see.

Before we move on it may be well to make one thing clear. There is no magic wand to bring truth out of the complexity we face. It is not that there is another set of tools which will extract the truth from the mass of uncontrolled factors which are the norm in life. It is not the case that Stenger merely has the wrong toolkit and that if he used another the certainty which he seeks would be delivered. The position, for him, is much worse. There is no way in which the near-certainties of the laboratory can be had in the real world outside. The scientific method can, and should, be used but the results which it yields will be more tentative. God cannot be so easily dismissed as Stenger supposes.

---

12 Fisher, R.A. *The design of experiments*, Edinburgh: Oliver and Boyd (1935), (7 editions, the last published in 1966).

This point is so important that we pursue it to show it limits so much of scientific activity in the real world. Economists and other social scientists typically deal with immensely complicated systems in which many of the relevant factors are unknown and most of those that are known are beyond their control. They have to be content with observing what seems to be the most relevant, or accessible, and doing the best they can. They are certainly aware of this, but many of their interpreters are not, and this often leads to the disillusionment of those whose expectations have been raised too high. The distinction between association and causation is at the root of the problem. Almost daily the news media report some new link that has been discovered between diet, and let us say, cancer. There are so many reports of this kind that the public become cynical, and their opprobrium often falls on science and scientists rather than on the ill-informed middlemen who misinterpret research findings.

If two things are associated it does not follow that there is a direct causal link between them, and even if there is, the direction of the causal link may be unclear. In the early days of studying the link between smoking and lung cancer, it was not clear whether smoking caused cancer, or whether incipient cancer created a disposition to smoke, or whether both were the result of some third factor which no one had then identified. That particular conundrum has now been resolved by identifying the chemical reactions in the lung that lead to cancer. But the problem is always with us, as shown in a recent report that the use of marijuana is linked to testicular cancer. Such reports appear almost daily; this particular one, based on a paper in the journal *Cancer*, was widely reported in the press in early February 2009. These, and others like them, are examples of the multifactorial nature of the phenomena which we struggle to understand.

Even the direction of the apparent causal link can be misleading, as economic pundits readily discover. Decreasing the price of something may be expected to result in increasing demand, but it does not necessarily follow that this is what we shall observe. If, at the same time, a new and better product appears at a competitive price, sales of the original product may fall, even though it may cost less. This is an extremely simple example of how factors may, through their interconnections, produce surprising results. Political debates often begin with one side asserting that certain benefits will surely follow if their policy is implemented: the opposition claims that they will not. If interest is maintained long enough for the debate to be resumed when the outcome is known, the losing side will usually be able to claim that 'other things were not equal' as the winners, if pressed, would have claimed they would have to be! This is tacit recognition that other factors were involved which were uncontrolled. It is this feature of the social world that makes its quantitative study so difficult. Social scientists are frequently upbraided for their 'physics envy' because, by comparison with physicists, they find it so difficult to reach definite conclusions – the more so because they work on topics which have more immediate public impact. Stenger does not appear to view science in this broader context.

## Significance tests

Here we move on to specific aspects of Stenger's approach. He says: 'The process I will follow is the scientific method of hypothesis testing.'<sup>13</sup> But once we get outside the narrow confines of experimental science, the position is more complicated. His understanding is a substratum of much of the book but it outcrops in a conspicuous fashion in a short section entitled 'The Significance of Experiments'<sup>14</sup> and it is on that section that I shall focus. (It should be noted that Stenger also discusses these matters in a section headed 'Statistical Significance'<sup>15</sup> where he correctly points out several common misunderstandings.) The 2008 version is set in the context of a discussion of extrasensory perception in which the ideas can be tested to destruction. The use of the word *significance*, is important because it is being used in a technical sense and it is that that we must examine.

It will be useful to set out the gist of this brief section before exposing the errors that it contains. At the core, according to Stenger, is the significance level, or *P*-value. This is a probability, often expressed as a percentage. It is calculated on the hypothesis that the treatment has no effect. The smaller this probability, the stronger the evidence against this hypothesis. Stenger points out that, in medical science, where treatments are being evaluated, 5% is a commonly used value of *P*. Stenger contrasts this with the much more stringent standard of one hundredth of 1% which, he says, is the norm in physics. He understands that more may be at stake in medicine, where failure to adopt an effective treatment, by adopting a more stringent standard, might cost lives. This might justify a lower standard but, even there, he supports moves to raise the standard to, say, 1%. In something like extrasensory perception (ESP) on the other hand, where little is lost by demanding higher standards, he pours scorn on the claims of ESP researchers that the same standards should apply in their work as are used in other branches of science. As it is, he claims there is so little evidence that he would have expected the field to have been abandoned long ago were it not for the tenacity of those who work in it.

It is, perhaps, pertinent to remark at the outset that statisticians are distinctly uneasy about the way that *P*-values are bandied about by scientists, and especially by editors of journals who insist on their use and impose particular levels of significance that must be achieved. The fundamental reason for their disquiet is that the *P*-value does not tell the whole story and it can have a distorting effect as we shall see below. This is part of a wider issue which has important implications throughout scientific research and it may be followed up by consulting Ioannidis<sup>16</sup> but it is not immediately relevant for our present purpose.

---

13 Stenger *op. cit.*, (1), p. 17.

14 Stenger *op. cit.*, (1), p. 91.

15 Stenger *op. cit.*, (6), p. 226.

16 Ioannidis, J. P. A. 'Why most published research findings are false' *PLoS Medicine* (2005) 2, 0696-0701.

A full discussion of statistical inference would be out of place here and there is some disagreement among statisticians about the best way to approach the whole matter but, at the level at which Stenger is operating, there would be general agreement about the errors he is making. We deal with these first.

The *P*-value is the probability of rejecting the hypothesis under test when, in fact it is true. Stenger makes the elementary mistake of claiming that if a level of 5% is habitually used in medicine, then one in twenty claims that are reported in medical journals are no more than a statistical artifact – in other words they are wrong! His actual words are; ‘But think of what that means. In every twenty claims that are reported in medical journals, on the average one such report is false – a statistical artifact.’<sup>17</sup> This is not true in general. It *would* be true to say that in every case *where there was actually no difference*, an experiment will be falsely reported as showing a difference one time in every twenty. But there is no reason whatsoever to suppose that there is no difference *every* time we make a test. We simply do not know whether there is a difference or, if there is, how big it is. It is quite plausible to suppose that there are often small effects and that these may well not be detected. The proportion of wrong conclusions by using this procedure is therefore unknown. We could only determine it if we knew whether or not there was no difference; and that is something which, in the nature of the case, we cannot know.

Why then do we use the *P*-value at all if it is so readily misinterpreted? This is a good question and leads some to suppose that it should be abandoned in favour of something corresponding more closely to reality. But it continues in use because it is a useful indicator when interpreted in conjunction with other indicators. The position will become a little clearer if we introduce the notion of types of error. The first kind of error, or Type I error, is to reject the hypothesis under test when it is true. But there is a second type of error, called Type II, which is to fail to reject the hypothesis when we should have done so. Unfortunately we cannot, simultaneously reduce the chance of both kinds of error. If we reduce one, the other will increase. Any choice therefore will be a compromise. That is why it is often rational to adopt larger *P*-values in medical research. Although doing so increases the risk of declaring that there is an effect when there is not, it also increases the chance that we will find an effect if there really is one. Making the compromise involves an implicit valuation of the two kinds of error.

## Meta-analysis

A second error that Stenger makes is in relation to meta-analysis (sometimes called metanalysis) which depends on the idea of a significance level. He remarks that meta-analysis is sometimes used to argue that a collection of tests, none significant when taken by itself, become significant when taken

---

17 Stenger *op. cit.*, (1), p. 92.

together. Meta-analysis refers to the combination of results from many separate analyses. The crux of his difficulty is summed up in the following quotation 'If several independent experiments do not find significant evidence for a phenomenon, we surely cannot expect a purely mathematical manipulation of the combined data to suddenly produce a major discovery.'<sup>18</sup> We surely can, though it is not clear what force should be given to the words 'mathematical' and 'suddenly' in this statement. The manipulations are no more mathematical than those involved in the constituent tests and there is nothing sudden about their emergence but we will let this dramatisation of the situation pass.

Before getting down to technicalities let us note that science is built on the accumulation of knowledge and it is surprising to meet an argument that appears to deny that evidence does, in some sense, build up. What is special about the accumulation of evidence from a sequence of tests of a particular hypothesis?

Stenger argues that in real science no conclusion is accepted until it has been corroborated by several independent experiments. This overlooks the fact that the conclusions of which we are speaking are, in the nature of the case, not established with certainty. The real issue is how to take account of these uncertainties when pooling the results. To clarify the matter we can look at the question in two ways. First, suppose that we have three independent experiments, each of which attains significance at the 5% level. The probability of wrongly rejecting the hypothesis of no difference on the combined result is then  $0.05 \times 0.05 \times 0.05$  or 0.000125 which is much smaller, and thus a more stringent level, than the original 0.05. If we are going to follow Stenger's rule we then have to decide what we want the overall significance level to be. Looking at it from the other side, imagine for a moment that we are contemplating using a sample size of 100; we could think of this being made up of ten samples each of size 10. It is quite possible that none of the samples of size 10 would yield a significant result but that the combined sample of 100 would. The ten samples of size ten, none significant in themselves, would, in combination, yield significance. At this most rudimentary level the evidence is adding up. There is nothing surprising or 'sudden' about this. In essence we are asking whether it would make any difference if we took all the samples together, or first looked at them separately. Stenger says that it would make a difference but this is not true.

Either way of looking at the argument shows that the individual significance levels and the overall levels are related and that we must be consistent in the way we regard them – as Stenger is not.

Stenger might respond that there is a difference between the cases he has in mind and the examples just used. This is, that most meta-analysis is based upon experiments conducted by different people, in different places, at different times. It is not simply a matter of the combination or division of independ-

---

<sup>18</sup> Stenger *op. cit.*, (1), p. 93.

ent samples. We took the above very simple cases to make the point as clearly as possible, but the logic is by no means restricted to cases of that kind. It is a technical matter to discover how to combine the results of different experiments to determine the correct significance level but the logic is essentially unchanged.

There is, however, one common situation where meta-analysis is often misused. This is in relation to what is known as publication bias and Stenger recognises this and refers to it as 'the file-drawer effect'.<sup>19</sup> If researchers only publish results which are 'significant' in some sense, the combination of those results will be biased. This may be because journals are only interested in publishing results which seem to say something new or it may be because scientific reputations are built on discovering new things, and not on reporting unsuccessful experiments. Whatever the reason, the set of results which one may wish to combine is biased, in that those least favourable to the outcome desired are excluded. The temptation to do this is present in cases where the results have no religious significance no less than in those that do.

As we remarked earlier, these issues come to the fore in work on ESP and what Stenger has to say about this has some relevance to the present discussion.<sup>20</sup> He is evidently unaware of the section on ESP in my *Uncertain Belief*.<sup>21</sup> It will be clear from that section that I hold no brief for ESP and regard its relevance for theology as marginal at best. But it will also be clear that there is some serious statistical work on the subject, of which Stenger appears to be unaware; this cannot be so easily dismissed, as Stenger implies, when he says 'the claim of scientific evidence for ESP does not stand up under the same scrutiny scientists apply when considering any extraordinary claim'.<sup>22</sup> The journal *Statistical Science*, in which the material was published, is a peer reviewed journal, and though the work is now somewhat dated, reading might well begin with the paper therein by Utts,<sup>23</sup> including the published discussion.

## Scientific sampling

Stenger and others, following a similar line, frequently quote the results of social research in support of their arguments. We shall meet examples later in this section. First, we lead up to the main problem by starting with the one area where social research can be truly scientific. This is the sample survey.

It is remarkable that we can accurately learn things about very large populations from quite small samples. People sometimes question this when it is

---

19 Stenger *op. cit.*, (6), pp. 285-290.

20 Stenger *op. cit.*, (1), pp. 89-91; Stenger, V.L. *Physics and psychics: the search for a world beyond the senses*, Buffalo, New York: Prometheus Books (1990).

21 Bartholomew *op. cit.*, (9), p. 133.

22 Stenger *op. cit.*, (1), p. 91.

23 Utts, J. 'Replication and meta-analysis in parapsychology', *Statistical Science* (1991) 6(4), 363-403.

put in those terms. They question whether a sample of a few hundred say, tells us anything about a population of hundreds of thousands? This is not true of *any* sample, of course, but only for rather special samples known as probability samples. A probability sample is one drawn in such a way that we can state the probability that any particular sample will be chosen. There is a lot more to it than that, of course, but let it suffice to say that we can design a sampling scheme to achieve a specified accuracy. Readers will be familiar with public opinion polls, which are the staple diet of public debate, but few of those carried out by commercial organisations are based on probability samples. This is because such samples are too costly and so other, non-scientific, methods are used instead. In academic research, on the other hand, probability samples are the norm and such research would be seriously suspect if they were not used.

All of that being said, many of the results quoted in such literature are suspect, and all are much more limited in scope than most of their users recognise. For example, with human populations, it may be difficult to obtain information on all of the sample members initially selected. This is liable to introduce bias because those excluded may not be representative of the sample as a whole. They may refuse to supply information or they may give false information. The results cannot be extended beyond the particular population sampled, and even then, populations change over time so that the things observed a year ago may not be the same today.

These are hazards enough and will beset any survey we attempt to carry out. If we merely observe a particular thing about each subject, such as their place of residence, it may be possible to proceed with the analysis with a fair degree of confidence. But often, subjects are questioned about beliefs and attitudes where no direct information is available. Subjects may interpret the question in different ways according to their individual perceptions. For example, the answer to the simple question, 'Do you believe in God?' depends on what you mean by the word 'God'. Questions on complex issues such as attitude to abortion can be graded according to the strength of feeling, but it is notoriously difficult to calibrate the answers. Social scientists are aware of these complexities and have devised ways of coping with them – using batteries of questions to probe the matter from different angles, for example. It is unusual to find signs that those from other disciplinary fields are aware of, or allow for, such subtleties.

In order to illustrate the difficulties of interpretation we shall take one of Stenger's examples but it is typical of others which can be found in the literature propagating atheism. I have deliberately taken a fairly extreme example which illustrates most of the failings commonly encountered. 'The empirical fact is that people with no religion are not noticeably less charitable than those with religion.'<sup>24</sup> To begin with, no source is given for the data which would con-

---

24 Stenger *op. cit.*, (1), p.249.

firm the alleged 'empirical fact'. This evidence therefore belongs to the class of assertions which elsewhere Stenger dismisses as anecdotal, but it will be instructive to proceed further. Who are the 'people' referred to – in other words what is the population in question? Was the information derived from a sample and if so was it a probability sample or not? Was the population American only, and if so, did it consist only of adults and within what age range? Was 100% response obtained, and if not what proportion did not respond and is anything known about the characteristics of any non-responders? What was the date of the survey, and have other surveys on similar topics been carried out with which the results could be compared? Finally, for our present discussion at any rate, how was 'charitability' measured? Was it based on the amount of money or time given to charitable activity? If so, was this restricted to a particular period? What counts as a charitable activity? Does giving blood or contributing to the weekly collection at the local church count? We could go on by asking how 'being religious' was defined as this is clearly as important as knowing how charitability was defined.

It might be protested that this is pedantry gone mad and that life would come to a standstill if such questions were asked about every assertion we make. Politics would certainly be paralysed if every claim made in political debate were to be put under the microscope in this fashion! But Stenger's assertion was made in a book claiming that religious belief dissolves once it is exposed to the corrosive action of scientific method. If religious belief must be subject to the full rigours of scientific enquiry then the case against must also be subject to the same rigour. It is extraordinary that those who are so insistent that religious claims can be subjected to scientific testing should be so cavalier in their use of evidence in fields outside their own speciality.

At this point we might note the warm endorsement by Richard Dawkins when he says: 'It needed a good physicist to show us the fallacy...' He goes on to mention 'the enormous amount' he learnt from 'this splendid book'.<sup>25</sup> That he still has more to learn, though not from this book, is evident from Dawkins' own use of a home-made sample survey to arrive at his widely reported conclusion that only 3.3% of Fellows of the Royal Society believe in God. The many flaws in the naive interpretations which may be put on this figure have been neatly exposed by Alexander,<sup>26</sup> not least that of assuming that the Fellows responding to his e-mail enquiry were a random (representative) sample of all Fellows. Stenger is thus not alone in the cavalier use of supposedly scientific methods outside the field of his own expertise.

### **Taking all the information into account**

The critics claim the high ground of rationality by contrasting their position

---

<sup>25</sup> Dawkins, R. on the back cover of Stenger, *op. cit.*, (1).

<sup>26</sup> Alexander, D. 'The Divine is in the detail', *Times Higher Education* (2008) 26, p. 16.

with believers whom, they suppose, depend on indoctrination and authority. In this section, as in the foregoing, we turn the tables by agreeing that the rationality advocated by the critics is highly desirable, but that it does not go anywhere near far enough because it does not take all of the evidence into account. Even if all of the criticisms of method made so far could be satisfactorily answered, the charge of irrationality would remain.

Talk of the 'elephant in the room' has become a common way of referring to the fact that many arguments take place with both sides ignoring what to the outsider is the most important factor in the situation. In the case of Christianity, at least, this is the person from whom the name Christian is derived.

Much of the current anti-religious polemic is targeted at the more fanatical branches of some non-Christian religions and at the most vociferous wing of conservative Christianity whose main strength is located in the southern states of the United States of America. This tactic is understandable on grounds of economy, if for no other reason. If one can identify and extract the core elements of most religions and demolish them at one fell swoop, then there is no need to waste time picking off the peculiarities of particular religions. This is obviously the way that Stenger sees it because his main target is a God whose basic characteristics are common to most religions and which he conveniently sets out.<sup>27</sup> Stenger anticipates the criticism that he is not a theologian by saying

Another common criticism has been that I am not a theologian and so should not be writing on the subject. While it is true that I am not a trained theologian, I certainly know the core theology for the God I am addressing. Furthermore, I am fully aware that theologians and apologists can come up with rational explanations for many of the inconsistencies and downright errors found in traditional beliefs. I wrote about some of these in my 2003 book, *Has Science Found God?* I admit there are logically possible gods.<sup>28</sup>

In passing, it is not clear how theologians and apologists can have 'rational' explanations for any 'downright errors' unless Stenger is conceding that what he sees as downright errors may not be errors after all.

If his general attack on theism were to be successful little remains to be said because if there is no God, the person of Jesus, the Christ, can have little significance. Indeed the only reference I can find on this is the dismissive remark

In other words, God does not wish to spend eternity with all human souls, but only the chosen few who, by blind faith in the absence of all evidence, accept a Jewish carpenter who may or may not have lived two thousand years ago as their personal savior. Of course, this is hardly a new idea but

---

<sup>27</sup> Stenger *op. cit.*, (1), pp. 41-42.

<sup>28</sup> Stenger *op. cit.*, (1), p. 240.

was essentially the teaching of John Calvin (d. 1564).<sup>29</sup>

This statement makes it abundantly clear that Stenger does *not* understand the core theology of the Christian God which centres on the 'Jewish carpenter' who Stenger so airily dismisses. For a scientist, taking all the evidence into account is essential and one cannot have confidence in a critic who so casually dismisses what, for Christians, is definitive. If Stenger had grasped that point, he would have had a much bigger task on his hands.

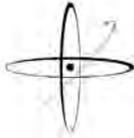
---

**David J. Bartholomew is Emeritus Professor of Statistics, London School of Economics.**

---

---

29 Stenger *op. cit.*, (1), p. 264.



CHRISTIANS IN SCIENCE

### **Co-sponsors of Science and Christian Belief**

CiS is an international network of those concerned with the relationship between science and Christian faith, open to scientists, teachers, students and all those with an interest in this dialogue.

- Developing and promoting Biblical Christian views on the nature, scope and limitations of science, and on the changing interactions between science and faith
- Bringing Biblical Christian thought on scientific issues into the public arena
- Encouraging Christians who are engaged in scientific work to maintain an active faith and to apply it in their professional lives
- Communicating the Christian gospel within the scientific community
- Stimulating responsible Christian attitudes and action towards care for the environment
- Helping Christians who are science students to integrate their religious beliefs and their scientific studies

[www.cis.org.uk](http://www.cis.org.uk)

Contact: CiS General Secretary, Dr Diana Briggs, email: [secretary@cis.org.uk](mailto:secretary@cis.org.uk)

Registered Charity No. 1121422