

# MML and the Fine Tuning Argument\*

By Steve Gardner, Graham Oppy and David L. Dowe

Our paper aims to apply our favourite Bayesian analysis of inference and prediction to the fine-tuning data. We argue that, on the Minimum Message Length (MML) analysis of inference and prediction, neither the hypothesis of intelligent design nor the hypothesis of many universes leads to any compression of the fine-tuning data; hence, we conclude that neither the hypothesis of intelligent design nor the many-universe hypothesis is supported by that data.

## Introduction

The fine-tuning argument is a special case of the argument for Intelligent Design. Roughly speaking, the argument proceeds from premises about the finely-tuned life-permitting values of the various constants in the laws of physics to the conclusion that the universe is the product of intelligent design. Elliot Sober (2004, p. 119) thinks that the best way to think about the fine-tuning argument is as an argument about likelihoods. He also thinks that the fine-tuning argument fails because of an *observational selection effect*: the likelihood that we observe that the constants are right given that the universe was designed by an intelligent designer is not greater than the likelihood that we observe that the constants are right given that the universe arose by chance. The likelihood is unity in each case, for if the constants were not right, it would not be possible to make observations.

In this paper we will argue against both of Sober's claims. Our first and more modest argument is that even if Sober is right that the fine-tuning argument is best characterised as an argument about likelihoods, the presence of the observational selection effect described by Sober does not defeat the argument. For, as we will show, the force of that argument depends on something that is not subject to an observational selection effect, namely, the *sensitivity* of the constants.

Our second and more ambitious claim is that Sober is wrong that the best way to characterise the fine-tuning argument is as an argument about likelihoods. We have both a specific and a general objection. Our specific objection is that since arguments about likelihoods don't tell you what you ought to believe, no argument couched in terms of likelihoods can be an argument with the conclusion that we ought to believe that there is an intelligent designer of the universe. Therefore, no such argument can be the best way to characterise the fine-tuning argument, since whatever one thinks of that argument, it is surely intended to be an argument whose conclusion is either that there is, or that we have good reasons to believe that there is, an intelligent designer of the universe.

Beyond this specific objection to the application of likelihoodism to the fine-tuning argument, we have a general objection to likelihoodism. We reject Sober's likelihoodism because we are Bayesians, who hold that theories should be assessed according to their

posterior probabilities. Sober's reasons for what he calls "the retreat to likelihoodism" are essentially negative. According to Sober, we must make do with likelihoods because Bayesianism is beset with a host of insoluble problems: the problem of priors, the problem of language variance, the sub-family problem, and the problem of nuisance parameters. Against this we will argue that the Bayesian method of inference by Minimum Message Length (MML) developed by Chris Wallace<sup>1</sup> has given Bayesians new resources with which to counter these objections.

Finally, we return to the fine-tuning argument to show how that argument is assessed from within the MML framework. In that framework, a weak constraint on whether a theory counts as an explanation of some data is that the theory allows the data to be encoded more concisely than would be possible using background knowledge alone. Our primary conclusion is that at present, there are no hypotheses of any kind that explain the fine-tuning data in this sense. It follows from this that the hypothesis of intelligent design also does not explain the fine-tuning data.

## The Fine-Tuning Argument and Observational Selection Effects

As Sober notes, the fine-tuning argument proceeds from the observation that the values of various physical constants are such as to permit life, and if they had been even slightly different, life would have been impossible. Sober abbreviates this observation as the claim that "the constants are right", and goes on to represent the fine-tuning argument as this claim about a likelihood inequality:

$$(1) \Pr(\text{the constants are right} \mid \text{Design}) > \Pr(\text{the constants are right} \mid \text{Chance})$$

Sober goes on to claim that this likelihood inequality does not hold, because of the presence of an observational selection effect:

(OSE) We exist, and if we exist the constants must be right.

Given OSE, what we should claim about the relevant likelihoods is this:

$$(2) \Pr(\text{the constants are right} \ \& \ \text{OSE} \mid \text{Design}) = \Pr(\text{the constants are right} \ \& \ \text{OSE} \mid \text{Chance}) = 1.0$$

Because of the observational selection effect, regardless of whether the universe arose by chance or by design, the likelihood that we observe that the constants are right is the same, namely, unity.

However, this attempt to defeat the fine-tuning argument fails. Recall that "the constants are right" is an abbreviation for a conjunction of two claims: the physical constants are such as to permit life *and if they had been even slightly different, life would have been impossible*. Keeping in mind the second conjunct, it is apparent that (OSE) is false. It is certainly true that if we exist, the physical constants must be such as to permit life. But it is false that if we exist, the physical constants must be such that if they had been even

slightly different, life would have been impossible. Our existence does not require that the constants be sensitive in this way; it is possible that we might have discovered that the physical constants are such that they fall in the middle of a broad range of life-permitting values.<sup>2</sup> Nor is the sensitivity of the constants irrelevant to the fine-tuning argument. On the contrary, the force of that argument is proportional to the sensitivity of the constants. Had we discovered that the values of the constants fall in the middle of a broad range of life-permitting values, the proposal that the universe must be the product of intelligent design would be much less attractive.

## Design Arguments and the Likelihood Principle

Sober's characterisation of the fine-tuning argument as an argument about likelihoods follows from his treatment of design arguments generally. Sober (2003:28) claims that "the best version of the design argument ... uses ... the *likelihood principle* [i.e. the claim that observation O supports hypothesis  $H_1$  more than it supports hypothesis  $H_2$  iff  $\Pr(O/H_1) > \Pr(O/H_2)$ ]. What Sober calls "the likelihood version of the design argument for the existence of God" (29) can be set out as follows: Let O be the claim that the vertebrate eye has features  $F_1, \dots, F_n$ ;  $H_1$  be the claim that the vertebrate eye was produced by an intelligent designer; and  $H_2$  be the claim that the vertebrate eye was produced by a mindless chance process. Since it is plainly the case that  $\Pr(O/H_1) > \Pr(O/H_2)$ , we can conclude, by way of the likelihood principle, that O supports hypothesis  $H_1$  more than it supports hypothesis  $H_2$ .

An obvious objection to this formulation of "the design argument for the existence of God" is that the formulated argument does not have the right kind of conclusion to count as an argument *for the existence of God*. Sober notes that, in general, "[L]ikelihood arguments have rather modest pretensions. They don't tell you which hypotheses to believe; in fact, they don't even tell you which hypotheses are probably true. Rather, they evaluate how the observations at hand discriminate among the hypotheses under consideration." (31) So, in the case at hand, Sober concedes straight off that the conclusion of his design argument is not that God exists, nor even that it is probable that God exists. But, if his argument has neither of these claims as its conclusion, then what right does it have to claim to be an "argument for the existence of God"?<sup>3</sup>

Sober does suggest one way to meet this line of objection. "Since I would like to restrict the design argument as much as possible to matters that are objective, I will not represent it as an argument concerning which hypothesis is more probable. However, those who have prior degrees of belief in  $H_1$  and  $H_2$  may use the likelihood argument to update their subjective probabilities. The likelihood version of the design argument says that the observation O should lead you to increase your degree of belief in  $H_1$  and reduce your degree of belief in  $H_2$ ." (30) But there is a problem with this suggestion. Even if we accept the claim that the likelihood version of the design argument says that the observation O *should* lead you to increase your degree of belief in  $H_1$  and reduce your degree of belief in  $H_2$ , it seems that we still don't have a reason for thinking that what we have been given deserves to be counted as a significant argument for the existence of God.

Let O be the observation that my car won't start and H be the hypothesis that there are powerful and malicious green gremlins in the engine whose purpose it is to prevent my car from starting. It is true that the observation O should lead me to increase my degree of belief in H. What has happened is that the extremely low probability I initially assign to H will have increased a tiny bit. It might even have doubled, from one in a billion, let us say, to two in a billion. But this doubling does not constitute a significant argument for the existence of powerful and malicious green gremlins in my engine.

When Sober claims that the best version of the design argument uses the likelihood principle, it is not entirely clear what he means here by "best". His further gloss that he wishes to find "the soundest formulation that the argument can be given" (28) does not advance matters at all. Indeed, on its strictest philosophical interpretation, this further gloss is retrograde, since it only makes sense if we help ourselves to a theory of degrees of truth. Plainly enough, on the assumption that validity is an all or nothing affair, we can only make sense of the idea that arguments are more or less sound by supposing that we can make sense of the idea that premises are more or less true. Not good. But perhaps we can bypass these worries without trying to establish precisely what it takes for one version of an argument to be better than another.

We will argue in the next section that there are good reasons for thinking that there are *Bayesian* formulations of arguments for design that are more deserving of attention than the likelihood formulation that Sober champions. In particular, we propose to argue that the reasons that Sober gives—both in his paper on the design argument and elsewhere—for dismissing Bayesianism, and hence, in particular, for dismissing Bayesian formulations of arguments for design, are inadequate. Moreover, we propose to investigate the proper formulation of arguments for design in the particular Bayesian setting that is provided by the theory of minimum message length inference.

## **Sober's Anti-Bayesian Arguments**

Sober has for many years and in many different papers, advanced several related objections to Bayesianism.<sup>4</sup> These are: (1) the problem of priors; (2) the problem of language variance; (3) the sub-family problem (a.k.a. the problem of simplicity); and (4) the problem of nuisance parameters. The point of the objections is to show that Bayesianism is beset by insoluble problems at its foundations. In light of these problems, Sober recommends a "retreat to likelihoodism" Sober (2004). It's worth asking: why does Sober characterise the move to likelihoodism as a "retreat"? The answer, we think, is that even a critic of Bayesianism such as Sober can see that the Bayesian program of analysing all genuine epistemic concepts in terms of probabilities would be a very attractive one, if it could be carried out. It's just that, in Sober's view, this program turns out, sadly, to be impossible. So we must retreat to a more modest epistemology.

As we argue in Dowe, Gardner, & Oppy (2007), we think this retreat is premature. In that paper, we show how Wallace's principle of Minimum Message Length makes available to Bayesians new theoretical resources, and we show how these resources can be deployed to solve the problem of language variance, the sub-family problem and, in

passing, the problem of nuisance parameters. But we do not address there in detail the problem of priors. We propose to do so here.

## The Problem of Priors

We think it fair to say that the problem of priors is the most serious problem for Bayesian philosophy of science. We should begin by getting clear about exactly what the problem is. We shall see that it is really two related problems.

Recall that Bayesians assess theories by their posterior probabilities in light of the evidence, where these are calculated according to Bayes's Theorem:

$$\Pr(H | E) = \frac{\Pr(H) \times \Pr(E | H)}{\Pr(E)}$$

How does this work in practice? Imagine you are Isaac Newton, investigating the relationship between force, mass and acceleration. You collect evidence (E) by conducting experiments in which you measure the accelerations of known masses subjected to known forces. You form the hypothesis (H) that  $F = ma$ .

What should you believe about this hypothesis? According to Bayesians, the posterior probability you assign to H depends on three things: the prior probability that you assign to H, the likelihood that you observe E given that H is true, and the unconditional probability of observing E. The last two of these terms are not especially problematic. While it is difficult to know how to calculate the unconditional probability of observing E, the need to do so can be avoided if we restrict ourselves to comparing different hypotheses. For it follows from Bayes's Theorem that

$$\frac{\Pr(H_1 | E)}{\Pr(H_2 | E)} = \frac{\Pr(H_1) \times \Pr(E | H_1)}{\Pr(H_2) \times \Pr(E | H_2)}$$

If this ratio of posteriors is greater than one, then E gives you more reason to believe  $H_1$  than  $H_2$ , and we can see that the calculation of this ratio does not require you to know  $\Pr(E)$ . When comparing theories we need only priors and likelihoods. Nor is there any problem in calculating the likelihood of observing E for some given H.

The problem comes in trying to make sense of the idea of the prior probability we should assign to H. What can it mean to say, for example, that  $\Pr(F=ma) = 0.3$ ? Here Sober argues that Bayesians are caught on the horns of an insoluble dilemma. For, he argues, the probability in question must be understood either as an objective or a subjective probability. If it is an objective probability, then the claim that  $\Pr(F=ma) = 0.3$  must be given a frequentist interpretation, akin to the claim that  $\Pr(\text{coin lands heads}) = 0.5$ . And this seems an implausible account of scientific theories—we do not think that these are the outcomes of chance processes!<sup>5</sup> On the other hand, if the claim that  $\Pr(F=ma) = 0.3$  is interpreted as a claim about subjective probabilities, i.e., about degrees of belief, then two

further objections arise. Firstly, Sober denies that beliefs are the kinds of things that come in degrees. He proposes that for any proposition, either we believe it, we disbelieve it, or we are agnostic about it.<sup>6</sup> Secondly, even allowing that there is a coherent account of degrees of belief, Sober objects that subjective probabilities ought to be irrelevant to the assessment of the truth of a given scientific theory. Subjective probabilities can differ to an arbitrary extent between agents. Therefore scientific disagreements should not be settled by pointing to the fact that different people have different subjective priors. This dilemma illustrates the first problem of priors: **that Bayesians can give no satisfactory account of what prior probabilities are.**

A possible Bayesian response to this dilemma leads to the second problem of priors. For you might have thought that even if you can't assign an objective prior probability to the claim that  $F=ma$ , or come to some agreement about its subjective probability, it might be possible to make use of a Principle of Indifference. The idea of such a Principle is to assign a uniform prior probability to every different possible theory in the domain of theories you are considering. You concede that you don't know what the probability is that  $F=ma$ , but you at least assert that it is no more or less than the probability that  $F=ma^2$ , or  $F = m + a$ , or  $F = \sqrt{m - a}$ , or any of other infinitely many other relationships that are possible between force, mass and acceleration.<sup>7</sup> The problem is that there is no unique or privileged way of describing what all the different possible theories in a domain are, and that different choices lead to different results. Sober gives a different example which makes this point nicely.<sup>8</sup> Say I tell you that my square garden is between 3m and 5m on a side. Applying the Principle of Indifference, you conclude that every length between 3m and 5m has the same probability. This description makes it seem natural to say that the probability is a half that my square garden is between 3m and 4m on a side. But what I told you is equivalent to saying that the area of my garden is between  $9m^2$  and  $25m^2$ , and this description makes it seem natural to say that the probability is a half that the area of my garden is between  $9m^2$  and  $17m^2$ . However, this entails that the probability is a half that my garden is between 3m and  $\sqrt{17} = 4.12m$  on a side. So the prior probability distribution you get from applying a Principle of Indifference to lengths contradicts the distribution you get from applying the Principle to areas. There's no non-arbitrary way to choose between the two parameterisations. We can restate the second problem of priors this way: **that Bayesians can give no satisfactory account of prior ignorance.**

In next section we shall briefly describe the principle of Minimum Message Length and argue that the principle makes available to Bayesians new theoretical resources that enable them to solve both problems of priors. In particular, we shall argue that Sober makes a crucial misstep in his argument when he assumes that objective prior probabilities must be given a frequentist interpretation. In fact there is another possible interpretation of these probabilities, in terms of *algorithmic complexity*. This interpretation relies on objective facts about Turing machines, and does not rely on arbitrary choices of parameterisation. So it is in an important sense objective. Yet the algorithmic complexity interpretation is also connected in important ways with the notion of subjective probabilities. We argue that the algorithmic complexity (AC) interpretation

of probability gives Bayesians satisfactory accounts both of prior probabilities and of prior ignorance.

## The Principle of Minimum Message Length

If the reader will permit us a little self-referential joke, then we'll begin our account of the Principle of Minimum Message Length (MML) by stating it in the briefest possible way, as a slogan:

### **Explanation is compression.**

Unpacking this slogan, what it says is that the best explanation of the facts (i.e., of some data) is the shortest. Given some data that you want to explain, the Principle of Minimum Message Length tells you to infer the theory which can be stated with the data in the shortest two-part message, where the first part of the message states the theory, and the second part of the message encodes the data under the assumption that the theory is true. Granting some connection between the notions of shortness and simplicity, the Principle of Minimum Message Length comes out as a generalization of Occam's Razor. What, then, is new in the Principle? We shall see that the Principle brings in its train two big new ideas of great importance for Bayesians.

The first big idea comes from Shannon (1948): given a proposition  $E$  with probability  $\Pr(E)$ , the Shannon information gained on learning  $E$  is the length of a message optimally encoding  $E$ . This message length is given by:

$$I_s(E) = -\log \Pr(E)$$

This establishes that the Principle of Minimum Message Length is a Bayesian principle. For the higher the probability that  $E$  is true, the shorter will be the length of the message optimally encoding  $E$ . Bayesians believe that the theory that best explains some evidence is the theory with the highest posterior probability. This theory will be the theory with the lowest Shannon information and the shortest message length.

It is important to note that the probabilities involved here are subjective probabilities, and that the Shannon information gained on learning that an event  $E$  has occurred is a subjective measure of information. The optimal encoding of an event is relative to a receiver with a specific set of prior expectations, and will differ from receiver to receiver. An event which is a great surprise to me has high Shannon information and a long optimal encoding for me. Yet the very same event may be just what you were expecting, and if so, it will have low Shannon information and a brief optimal encoding for you. One way of restating the problem of priors is to point out that there is no bound on the ratio of the prior probabilities that different agents assign to the same event. In the worst case, I may deem an event  $E$  to be impossible, thus having infinite Shannon information, while you may deem the same event to be necessary (that is, logically implied by what you already know), and so as having zero Shannon information.

To find a way out of this impasse, we will need the help of the second big idea, which comes from Kolmogorov: given a finite string  $S$  written in some alphabet, and a Universal Turing machine  $T$ , the Kolmogorov complexity (also known as the *algorithmic complexity*) of the string  $S$  is the length  $L$  of the shortest program which, when given to  $T$  as its input, causes  $T$  to output  $S$  and halt. Since the Turing machine is Universal, there is guaranteed to be such a program for any finite string. We can regard the shortest program that produces  $S$  as an optimal encoding of  $S$  relative to  $T$ , thus equating the Shannon and Kolmogorov measures of information. Since the Shannon measure of information is equivalent to the negative log of a probability, it follows that the Kolmogorov measure of information is also equivalent to the negative log of a probability. We can therefore regard every Universal Turing machine (UTM) as implicitly defining a probability distribution over all finite strings: for a given UTM  $T$  and any finite string  $S$ ,  $T$  implicitly asserts that  $S$  occurs with probability  $2^{-L}$ .

All data has to be written down, so there is no problem in regarding the data to be explained as a finite string written in some alphabet. We now observe that a theory is also an assertion, implicit or explicit, of a probability distribution over finite strings, since a theory tells us what data is more or less likely to be observed if it is true. The choice of UTM is therefore equivalent to the choice of theory, or choice of prior probability distribution.

So far, this has not advanced us much. The problem we encountered before was that different agents can have prior probability distributions that differ to an arbitrary extent, and that there seemed to be no principled way to choose between them. Using the concept of Kolmogorov complexity, we have translated the problem of the choice of prior probability distribution into a problem of the choice of UTM. But how does this help?

Universal Turing Machines have an important property: given any two UTMs, there is a program of finite length (called an *interpreter*) to make one precisely imitate, or *emulate*, the other. This means that, in contrast to the situation considered from the perspective of Shannon information, for any pair of UTMs, there is an upper bound on the ratio of the probabilities of a string implicitly asserted by that pair of UTMs. This bound is just two to the power of the length of the interpreter required to make one UTM emulate the other. The existence of this bound overcomes a significant objection to the characterisation of probabilities as subjective probabilities: while different agents may disagree in their subjective prior probability assignments, to the extent that we are prepared to regard agents as analogous to UTMs, algorithmic complexity theory guarantees that the difference between agents is bounded.

Nevertheless, it remains true that agents can disagree sharply in their subjective probabilities. As Bayesians, we think this is an ineliminable feature of scientific investigation. But, as we shall argue in the next section, we also think that the framework of inference by Minimum Message Length provides a solution to the second problem of priors, by giving an account of prior ignorance which is objective in an important sense. If this argument is successful, much of the sting of the first problem of priors is drawn.

## Prior ignorance and simplest possible Universal Turing Machines

A Turing Machine is a logical description of a computer. It has a clock which counts time in discrete units called *clock cycles*. It has a finite number of *internal states* which partly determine what the machine does during each clock cycle. It has a tape of infinite length with a read/write head that can read input from the tape, and write output to the tape. The inputs which are read and the outputs which are written are *symbols* in some finite alphabet. For our purposes we can assume the alphabet consists of the symbols '0' and '1'. The symbols are written on the tape in *cells*, one symbol to a cell. There are no blanks cells on the tape. The TM can perform a small number of *actions*: it can move the read/write head left or right one cell along the tape, and it can write a symbol to the tape. The TM also has a finite number of *instructions*: these describe what the TM does if it is in any of the possible situations it can be in. There are only two different possible situations for each internal state of the TM, corresponding to the possibilities of reading a '0' or a '1'. An instruction associates an action and a next state with each possible situation. At each tick of the clock, the TM reads the symbol under the read/write head, and depending on what state it is in, performs one of the possible actions and goes into the indicated next state.

From this brief description, it is easy to see that every Turing machine has a description of finite length that completely characterises its behaviour. The description is a *state table*: the table has a row for each internal state of the Turing machine, and a column for each symbol in the alphabet the machine uses. The entries in the table are instructions. The table below shows the state table for a TM that adds two numbers together: given a tape with two strings of ones separated by a zero written on it, the TM outputs a tape with a single string of ones on it that is the length of the two strings added together. The instruction in each cell is a triple, e.g. 'R, 1, 2' means 'move right, write a 1, go to state 2'.

State	Read '0'	Read '1'
1	R, 1, 2	R, 1, 1
2	L, 0, 3	R, 1, 2
3	L, 0, 3	L, 0, 4
4	R, 0, HALT	L, 1, 4

**Table 1: Turing Machine state table for a simple adder**

Turing machines vary in complexity. Simple TMs such as the one above can perform simple tasks. Of course this TM is not universal. To create a Universal Turing Machine (UTM), that is, a TM capable of emulating any other TM, a certain minimum level of complexity is required. This invites the question: what would *the simplest possible UTM* be like? Recall the point we made earlier that the choice of any UTM expresses the prior

expectations we have about the data—data (that is, finite strings) with brief encodings can be thought as being expected, or more probable, relative to that UTM. So, what expectations about the data are implied the choice of the simplest possible UTM?<sup>9</sup>

Now we come to the crux of our argument in this section.<sup>10</sup> The simplest possible UTM can be regarded as one that has not been programmed to do anything else other than emulate other TMs. It therefore expresses the most complete ignorance possible about future data; in effect, the only assumption about the data reflected in the choice of the simplest possible UTM is the very weak assumption that the data is the output of a computable function. Recalling that every UTM defines a probability distribution over all finite strings, and that all data are finite strings, we can say that the simplest possible UTM defines a probability distribution over possible data that reflects complete prior ignorance about that data.

## The Characterisation of Prior Knowledge

With this characterization of prior ignorance in hand, we are now also in a position to say how prior knowledge ought to be characterised. In particular, we can answer the question raised earlier: how should a Bayesian interpret claims about the prior probabilities of theories?

The answer is that for a Bayesian the assertion of a theory in the first part of the message *implicitly asserts* a prior probability for that theory. On the algorithmic complexity interpretation of probability, message lengths and (negative logs of) probabilities are interchangeable (take another look at Shannon's Law). More generally, the choice of a language in which to describe the space of possible theories implicitly asserts a prior probability distribution over the theories being considered. Recall our earlier discussion of the possible relationships between force, mass and acceleration. There are an infinite number of different ways in which these three quantities might be mathematically related. But physicists investigating this problem in Newton's time, and physicists today investigating other problems, share a commitment to the idea that the language of mathematical physics is the appropriate language in which to frame hypotheses. In this language, assertions like 'F=ma' or 'F=m+a' are easy to state. The Minimum Message Length Principle attaches a specific significance to this fact: the fact that these theories are briefly assertable in the language chosen as the most appropriate one in which to describe possible theories indicates that these theories are regarded as having higher prior probabilities than those whose assertions in that language are longer.

## Applying MML to a Case of Theory Choice

The arguments in the previous two sections introduced some fairly elaborate theoretical machinery (Universal Turing machines, Kolmogorov complexity) to show how Bayesians can meet Sober's general objections to Bayesianism. Before we return to the fine-tuning argument, and the assessment of that argument from within the MML framework, we provide a brief discussion of the application of MML to one well-known case of theory choice, namely the choice between the Ptolemaic and Copernican theories

of the relative motions of the sun, the moon, and the six innermost planets in the solar system (henceforth ‘planets’). We note in passing that, although the theoretical machinery introduced above was needed to provide a general justification of Bayesianism, it is not needed in assessing the theories that we are about to consider.

It is well-known—see, for example, Hoyle (1973)—that we can think of the Ptolemaic and Copernican theories as expansions of the terms of a series approximation to the correct theory: addition of epicycles and so forth can make each theory approach arbitrarily close to the truth. Furthermore, at any level of approximation, there is nothing to choose between the two theories in terms of their fit to the data concerning relative planetary positions, on the best formulation of each of the theories at the given level of approximation. Thus, under the MML approach, we see that, at any level of approximation, on the best formulations of the two theories, there is nothing to choose between them in terms of the second part of the message, insofar as we suppose that the only relevant data concerns observed planetary positions relative to the fixed stars: planet *x* was in position *y* relative to the fixed stars at time *t*.

Despite this, it does not follow that, on the MML approach, there is nothing to choose between the two theories. For it is also well-known that there are facts concerning the relative observed positions of the planets—i.e. the positions of the planets relative to one another—that are consequences of the geometry of the Copernican system, but that require further *ad hoc* postulates in the Ptolemaic system. For instance, it is simply a consequence of the geometry of the Copernican system that the inferior planets always remain close to the Sun, and that the superior planets only retrogress when they are in opposition to the Sun. On the Ptolemaic theory, however, it is only the further postulate that the centres of the major epicycles of the inferior planets always lie on the line between the Earth and the Sun that ensures that the inferior planets remain close to the Sun (and there is a similar postulate that ensures that the superior planets only retrogress when they are in opposition to the Sun). Consequently, we see that the first part of the message—the statement of the theory—is longer in the case of the Ptolemaic theory than it is in the case of the Copernican theory: there are further parameters that must be fixed in the Ptolemaic theory in order to get it to achieve the same degree of compression of the data as is achieved by the Copernican theory. (Make poor choices of values for the motions of the centres of the major epicycles of the inferior planets, and your error terms for the positions of those planets will go though the roof.) So, the two-part message for the Copernican theory is shorter than the two-part message for the Ptolemaic theory: and hence MML recommends that we prefer the Copernican theory to the Ptolemaic theory.

In the preceding discussion, we supposed that it is appropriate to think of the incorporation of the *ad hoc* hypotheses into the Ptolemaic theory as a matter of fixing the values of further parameters in order to make calculations of the relative observed positions of the planets (i.e. calculations that yield claims of the form planet *x* is in position *y* at time *t* relative to the fixed stars). However, at a more impressionistic level, we might think about matters in the following way. The ‘data’ with which we work includes the fact that the inferior planets are always observed in the same part of the sky

as the sun (and the fact that the superior planets only retrogress when in opposition to the sun). While this ‘data’ is entailed by the geometry of the Copernican theory, it effectively has to be ‘written in by hand’ into the Ptolemaic theory. So, while both theories provide effective compression of the general data involving the relative observed positions of the planets, the two theories differ in their ability to compress the ‘data’ concerning, for example, the relative positions of the Sun and the inferior planets. Since there is no difference between the two theories except for their ability to compress this particular ‘data’, we can think of this ‘data’ as the critical divide which speaks decisively in favour of the Copernican theory. Moreover, we can say that, whereas the Copernican theory provides an explanation of this ‘data’, the Ptolemaic theory fails to do so: ‘data’ is explained only insofar as it is compressed, and the Ptolemaic theory simply fails to compress the ‘data’ that we have highlighted.

## Applying MML to the Fine-Tuning Data

Given the discussion in the preceding section, we have a model for our assessment of the relative merits of the treatment of ‘the fine-tuning data’ on the approaches of intelligent design theories and many universe theories.

We recall from our earlier discussion that the primary ‘data’ that is appealed to in the fine-tuning argument is this: the values of various physical constants are such as to permit life, and if they had been even slightly different, life would have been impossible. Before we can begin to assess how intelligent design theories and many universe theories fare in their treatments of ‘the fine-tuning data’, we need to decide exactly what that ‘data’ is. Since our aim is to transmit our data by transmitting the shortest two-part message, the first part of which is a theory that compresses our data, and the second part of which is our data as compressed by that theory, it is obvious that we cannot make any progress towards our aim until we know exactly what data we are proposing to transmit.

Suppose we think that there are two parts to the data that we aim to transmit. First, we aim to transmit the data about the life-permitting ranges: for each physical constant, we give the variation beyond which life would be impossible. Second, we aim to transmit the actual values of the physical constants. Given that we transmit both of these kinds of data, then we shall certainly have transmitted the information that the constants all lie in the life-permitting regions, and we will have transmitted the information that, had the values differed a little, then they would not have fallen within the life-permitting regions.

If this is how we are thinking about the data that we aim to transmit, then it is pretty clear that we simply aren’t going to come up with *any* theories that can compress this data. In particular, it is clear that neither intelligent design theories nor many universe theories compress this data: in any cases of this kind, the first part of the message—the statement of the theory—is longer than a one-part message in which the data is transmitted neat. For example, if we construct an intelligent design theory according to which an intelligent designer desires that the physical constants  $C_1, C_2, \dots, C_n$  have values  $V_1, V_2, \dots, V_n$ , that lie with the bounds  $B_1, B_2, \dots, B_n$ , then the first part of our message is longer than the neat message that the physical constants  $C_1, C_2, \dots, C_n$  have values  $V_1,$

$V_2, \dots, V_n$ , that lie with the bounds  $B_1, B_2, \dots, B_n$ ; and likewise, if we construct a many universe theory according to which there is one amongst an ensemble of universes in which the physical constants  $C_1, C_2, \dots, C_n$  have values  $V_1, V_2, \dots, V_n$ , that lie with the bounds  $B_1, B_2, \dots, B_n$ , then, again, the first part of our message is longer than the neat message that the physical constants  $C_1, C_2, \dots, C_n$  have values  $V_1, V_2, \dots, V_n$ , that lie with the bounds  $B_1, B_2, \dots, B_n$ .

How else, then, might we think about the data that we aim to transmit, and about the theories that we might hope to use to compress that data? Well, we might be thinking that it could turn out that there are theories that compress a large range of other data as well as the particular data in which we have an interest. Think here about the difference between the Copernican theory of the motions of the planets, and the Newtonian theory of the motions of the planets. Newton's physical theory compresses a vast range of physical data that simply doesn't fall under the Copernican theory: thus, not only does the Newtonian theory of elliptical orbits provide a shorter two-part message of the data concerning relative observed positions of the sun, moon, and six innermost planets in the solar system than does any particular version of the Copernican theory, but the wider theory of Newtonian mechanics provides a shorter two-part message of the data concerning the motion of physical objects in general than does any of the competing theories of the time. Moreover, the Newtonian theory of elliptical orbits is *entailed* by the wider theory of Newtonian mechanics: that is, the wider theory of Newtonian mechanics explains the Newtonian theory of elliptical orbits because (a) the wider theory of Newtonian mechanics compresses a wide range of other data, and (b) the wider theory of Newtonian mechanics entails the Newtonian theory of elliptical orbits.

If we try to apply this model to the case of the fine-tuning data, then we need to suppose that there are wider theories—intelligent design theory and many universe theory—which compress a wide range of other data and which entail that the physical constants  $C_1, C_2, \dots, C_n$  have values  $V_1, V_2, \dots, V_n$ , that lie with the bounds  $B_1, B_2, \dots, B_n$ . But, as things now stand, we don't have any reason to believe that there are wider theories of these kinds that satisfy the first of the constraints: as far as we know, there is no other data that is compressed by either intelligent design theory or many universe theory. If it were to turn out that there is other data that is compressed by an intelligent design theory or a many universe theory that also *entails* that the physical constants  $C_1, C_2, \dots, C_n$  have values  $V_1, V_2, \dots, V_n$ , that lie with the bounds  $B_1, B_2, \dots, B_n$ , then we would have reason to reassess: but, unless that happens, we have good reason to say that neither intelligent design theory nor many universe theory affords an explanation of the fine tuning data.

If we set aside worries about whether the fine-tuning data is merely an artifact of current parameterisations—i.e. if we suppose that the fine-tuning data is something that could have an explanation—then it seems to us that it is pretty clear that there is no currently acceptable explanation of that data. It's not merely that intelligent design theories and many universe theories fail to explain the data: there is *no* theory constructed thus far that has the capacity to explain this data. We think—though admittedly this is speculation, and not required by our argument to this point—that if there is an acceptable explanation

of the fine-tuning data, then it will be in the form of a wider Theory of Everything that explains a great deal about the structure of space, time and matter, and that entails that the constants have the values that they actually have. That is, if there is an acceptable explanation of the fine-tuning data, it will be analogous to the Newtonian explanation of the elliptical orbits of the planets.

## References

- Dowe, D. L., Gardner, S., & Oppy, G. (2007). Bayes not bust! Why simplicity is no problem for Bayesians. *British Journal for the Philosophy of Science*, 58(4), 709-754.
- Forster, M. R., & Sober, E. (1994). How to tell when simpler, more unified or less ad hoc theories will provide more accurate predictions. *British Journal for the Philosophy of Science*, 45(1), 1-35.
- Hoyle, F. (1973). *Nicolaus Copernicus*. London: Heinemann.
- Post, E. L. (1943). Formal reductions of the combinatorial decision problem. *American Journal of Mathematics*, 65(2), 197-215.
- Schmidhuber, J. (1997). A computer scientist's view of life, the universe, and everything. In C. Freska (Ed.), *Foundations of computer science: Potential - theory - cognition* (pp. 201-208). New York: Springer.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379-423, 623-656.
- Sober, E. (2002). Bayesianism - its scope and limits. In R. Swinburne (Ed.), *Bayes's theorem* (pp. 21-38). Oxford: Oxford University Press.
- Sober, E. (2004). The design argument. In W. E. Mann (Ed.), *The Blackwell guide to the philosophy of religion* (pp. 117-147). Oxford: Blackwell.
- Wallace, C. S. (2005). *Statistical and inductive inference by Minimum Message Length*. New York: Springer.
- Wallace, C. S., & Boulton, D. M. (1968). An information measure for classification. *Computer Journal*, 11, 185-194.

---

\* This paper was written with the support of a Monash University Arts/IT Small Grant. The authors would like to thank Vaughan Pratt for helpful discussions about Turing machine complexity. Needless to say, any mistakes are our own.

<sup>1</sup> The first published statement of the Principle of Minimum Message Length occurs in Wallace & Boulton (1968). The best and most complete development of the Principle, its applications and implications, is in Wallace (2005).

<sup>2</sup> Indeed, it is possible that we might yet discover that the physical constants are such that they actually do fall in the middle of a broad range of life-permitting values, that is, that the apparent sensitivity of the constants is only apparent. There are two different ways in which this could happen. What is meant by the claim that some particular physical constant is sensitive is that if one varies the value of that constant *while holding everything else fixed*, then the result is a universe in which life is not possible. The general claim that “the constants are sensitive” is a conjunction of individual claims of this kind made about some set of

---

the physical constants. But now, firstly, it is possible that the apparent sensitivity of the constants is an artefact of our parameterisation, and that in some different parameterisation of the physical constants, the constants are not sensitive in this way. And secondly, regardless of parameterisation, for any small change in the value of one of the constants, it might be that compensatory changes in some or all of the other constants could be made that would again make life permissible in the universe.

<sup>3</sup> As Sober notes (p.31), there are difficulties involved in the supposition that design arguments are arguments for the existence of God rather than merely arguments for the existence of some kind of intelligent designer. However, these difficulties do not bear on the point that is currently at issue: if we focus instead on the claim that there is an intelligent designer, the formulated argument will not have the right kind of conclusion to count as an argument *for the existence of an intelligent designer*.

<sup>4</sup> See for example Forster & Sober (1994), Sober (2002), Sober (2004).

<sup>5</sup> Actually, some people do believe this process model, or something like it, e.g. Schmidhuber (1997) But that's by the by: we agree with Sober that it's bad news for Bayesians if it turns out that Bayesianism alone commits them to anything so metaphysically extravagant.

<sup>6</sup> Sober (2004, p. 121)

<sup>7</sup> It might be thought that some of these relationships should *a priori* be ruled out as impossible on the grounds that the quantity on the LHS of the equation (force, measured in Newtons, or  $\text{kg ms}^{-2}$ ) must have the same dimensionality as whatever is on the RHS. But a simple manipulation can turn any quantity into a dimensionless constant: express the quantity as a fraction or multiple of a physical constant having the same dimension. So, the size of an electrical charge can be expressed as a fraction or multiple of the charge of an electron, length can be expressed as a fraction or multiple of the Planck length, time can be expressed in terms of the Planck time, and so on.

<sup>8</sup> Sober (2004, p. 119). The example is a variant of Bertrand's Paradox.

<sup>9</sup> The use of the phrase "*the simplest possible UTM*", implying the existence of a unique simplest UTM, masks complications that are fascinating in their own right, and which have a bearing on some related questions of interest, but which are not strictly relevant to our purposes here. The main points to note are these: (1) in this paper, both for the sake of clarity and for historical reasons, we use the state table method of characterising TMs, which was Turing's own method. Given this method, the complexity of any TM or UTM can be characterised in terms of the *state-symbol product*, that is, the number of cells in the state table. But there are other ways of characterising UTMs, for example, the Post tag systems of Post (1943), and each of these has associated with it a corresponding characterisation of complexity. Although it is known that every Post tag system is equivalent to a Turing machine and vice versa, the equivalence does not necessarily preserve the complexity ordering; (2) different UTMs can have the same state-symbol product, for example, a 2-state, 3-symbol machine, and a 3-state, 2-symbol machine; (3) for a given machine, you also have to specify the permissible initial states of the tape. The recent controversy between Vaughan Pratt and Alex Smith over whether Smith has or has not proved the existence of a 2-state, 3-symbol Universal Turing Machine (for which Smith was awarded the Wolfram Prize) turns on this question.

<sup>10</sup> This argument is presented in more detail in Wallace (2005, pp. 133-135).