In this paper I am going to concern myself with the concept of consistency, Gödel’s second theorem, and the anti-mechanistic argument of J. Lucas. My approach to the issue involved will be somewhat different from the usual logical/philosophical approach which always seems to provide the mentalist with a convenient loophole. I will not attempt to give a detailed explication of the relevant logical theory, as this is covered adequately elsewhere, but some brief statement of my terms of reference is obviously essential.

Let us begin with a formal logical system S which is adequate for the proofs of all the basic results of elementary number theory. Such a system is said to be simply consistent if and only if no well-formed formulas A of S is such that A and its negation are both theorems of S, and absolutely consistent if and only if at least one formula of S is not a theorem of S. Obviously simple consistency implies absolute consistency and vice versa.

Now Gödel’s second theorem, a corollary of his first, may be expressed approximately as follows: if Con\(_S\) is an arithmetization of the statement that the first-order theory S is consistent, then, if S is sufficiently strong and consistent, Con\(_S\) is not provable in S. This may be expressed symbolically

\[ S \vdash \neg \text{Con}_S \rightarrow G \]

where G is the relevant Gödel sentence. Thus if we were able to prove the consistency of S within S then, by modus ponens, G would follow, which by Gödel’s first theorem would render S inconsistent and leave us with a contradiction.

Lucas’s paraphrase of Gödel’s second theorem is “Gödel has shown that in a consistent system a formula stating the consistency of the system cannot be proved in that system”, and whilst such an interpretation is open to serious criticism I will not pursue the matter in this context. Using this interpretation as his basis, Lucas claims that no consistent machine is capable of producing a computable or computably enumerable expression of its own consistency and whilst such a limitation applies to the formal reasoning of a human computer, the latter, by means of informal reasoning, is able to transcend this limitation and assert his own consistency. Lucas unfortunately does not clearly state what he means by informal reasoning or informal logic. One could take it as a reference to that vast number of inferences drawn in everyday life which are not, and could not be, codified in any formal logic system. Alternatively he could mean by informal logic ‘a theory of valid inference’, but whether this sheds any more light on the matter I do not know. Nevertheless, the message is clear and simple; by means of informal logic, which Lucas knows is consistent, he can informally assert his formal consistency, and since he is formally consistent, so is mathematics.

But how does Lucas know this?

“In saying that a conscious being knows something we are saying not only that he knows it, but that he knows that he knows it, and so on, as long as we care to pose the question: there is, we
recognize, an infinity here, but it is not an infinite regress in the bad sense, for it is the
questions that peter out, as being pointless, rather than the answer.” (3)

There are however one or two questions I wish to pose before I peter out.
I will begin by taking Lucas at his word and define my null hypothesis as “my informal logic
(reasoning) is consistent”; in case my subsequent analysis leads me to reject this I will define
as alternative hypothesis “I do not know whether my informal logic is consistent or not.” Since
it is Lucas’s contention that the consistency of our informal reasoning is the ultimate guarantee
of the consistency of mathematics, my null hypothesis allows me to assume the consistency of
that discipline also. Hence I can use it with confidence.
In order to test my hypothesis it is then essential that I direct my attention to that disorderly
field which Lucas claims, establishes my mental superiority over all possible individual
artefacts. To begin let us assume that our informal mental life-space consists of n propositions,
$A_1, A_2, \ldots, A_n$ and their negations, $-A_1, -A_2, \ldots, -A_n$, where n is a finite or denumerably
infinite, natural number, and that an individual mind, in the course of its life, can acquire any
number of propositions and hold them as axioms. Furthermore, consistency will be taken as the
analogue of simple consistency in a formal system i.e., a mind will be considered inconsistent
if it holds any proposition and its negation as axioms. For the purpose of illustration, consider
the case where n = 2, i.e. the mental life-space consists of four axioms $A_1, A_2, -A_1, -A_2$.
Now a mind in such an environment could hold
(a) 1 axiom giving 4 choices
(b) 2 axioms giving 6 choices
(c) 3 axioms giving 4 choices
(d) 4 axioms giving 1 choice.

Within this model it is also easy to demonstrate that

if a mental system consists of 1 axiom it can be inconsistent 0 ways
if a mental system consists of 2 axioms it can be inconsistent 2 ways
if a mental system consists of 3 axioms it can be inconsistent 4 ways
if a mental system consists of 4 axioms it can be inconsistent 1 way.

Table 1 summarizes the relevant outcomes for n = 1, 2, ..., 7 and in general, the generating
formulae are as follows: –

<table>
<thead>
<tr>
<th>Total number of sets of axioms (consistent and inconsistent)</th>
<th>$2^{2n} - 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sets of consistent axioms</td>
<td>$3^n - 1$, where n = 1, 2, 3, ...</td>
</tr>
</tbody>
</table>

If we now make the assumption that all sets of axioms have equal probability of acceptance by
a mind then the probability of a mind being inconsistent is given by the function

\[
\frac{2^{2n} - 3^n}{2^{2n} - 1}
\]

it is quite clear from the data of Table 1 that according to such a design the probability of accepting an inconsistent set of axioms is extremely high for comparatively small n; thus for n
greater than 16 the probability exceeds 0.9900. One might initially make the rather naïve point
that if mathematics is consistent and this were a real-life design, then the chance of anyone
being informally consistent is very low indeed. There would, statistically, be a tiny minority of
exceptions and no doubt Lucas would be numbered amongst them. However, the assumption
that all sets of axioms have equal probability of acceptance is patently false and a conclusion
such as that suggested above is not valid, though I feel it is a little sobering.
Nevertheless, the exercise has not been futile. Consider once again a mind operating in a space consisting of \( n \) axioms and their negations. Its informal mental calculus (and it is Lucas who says that exists, not I) must consists of 1 axiom or 2 axioms … or \( 2^{2^n} \) axioms together with the necessary inferential rules which I am not questioning at this stage. Now the \( (2^{2^n} - 1) \) sets of axioms are subdivided into consistent sets numbering \( (3^n - 1) \) and inconsistent sets numbering \( (2^{2^n} - 3^n) \). Let us assign probabilities of acceptance \( p_1, p_2, \ldots, p_{2^{2^n} - 3^n} \) to the individual inconsistent sets and probabilities of acceptance \( p_{2^{2^n} - 3^n + 1}, \ldots, p_{2^{2^n} - 1} \) to the consistent sets, where \( p_1 + p_2 + \ldots + p_{2^{2^n} - 3^n} + p_{2^{2^n} - 3^n + 1} + \ldots + p_{2^{2^n} - 1} = 1 \). The mind must at any point in time hold some set of axioms as being true and the null set is necessarily excluded since if a mind is completely agnostic this in itself would constitute an axiom. The probability of a mind being consistent, \( \Pr(\text{con}) \), is

\[
\sum_{i = 2^{2^n} - 3^n + 1}^{(2^{2^n} - 1)} p_i
\]

and the probability of being inconsistent

\[
\Pr(\text{Incon}) = 1 - \Pr(\text{Con}) = 1 - \sum_i p_i
\]

And at this point we run serious difficulties. Despite the fact that we know that for \( n \geq 3 \) the number of inconsistent sets is greater than the number of consistent sets we cannot assert that \( \Pr(\text{Incon}) > \Pr(\text{Con}) \) since we have no means of evaluating \( \sum_i p_i \). The ascription of values depends on what I will loosely describe as ‘another source’ and this we could not possibly hope to analyse mathematically. It is important to stress here that we are not talking of the probability of an axiom being true but only of the probability of an individual mind accepting a particular group composed of independent axioms. It is also to be noted that in the case of Lucas, \( \Pr(\text{Con}) = 1 \) i.e. his ‘other source or sources’ must confer a probability of unity on his acceptance of a particular axiom set and this again cannot be the null set since Lucas does believe certain things to be true.
### Table I

<table>
<thead>
<tr>
<th>Number of consistent sets</th>
<th>( \text{NUMBER OF AXIOMS} (2^n \text{ i.e. } A_1, A_2, \ldots, A_n - A_1, \ldots, - A_n) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Number of sets</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Number of ways of holding n axioms</td>
<td>0</td>
</tr>
<tr>
<td>TOTAL</td>
<td>3</td>
</tr>
</tbody>
</table>

### Table II

| Number of Inconsistent Sets | 1 | 7 | 37 | 175 | 781 | 3,367 | 14,197 | \( \ldots \) | \(2^{n-1}-1\)-(3^{n-1}) |
|-----------------------------|---|---|----|-----|-----|-------|--------|\( \ldots \) |               |
| Probability of Inconsistent Set | 0.33 | 0.47 | 0.59 | 0.69 | 0.76 | 0.82  | 0.87   |               | \( \rightarrow 1 \text{ as } n \rightarrow \infty \) |
I will now consider an obvious criticism that can be levied against this model. It could be argued that no one who would be counted as rational, in the general sense of the word, would accept any proposition and its negation of axioms; for example, no one would believe that God exists and God does not exist, at one and the same time. But, on reflection, one has to admit that in real life, propositions are not always presented to us in this clear-cut form; indeed one holds such a bewilderingly complex array of axioms that the possibility of detecting all inconsistencies must be infinitesimally small. In order to be certain of consistency it would be necessary to ascertain the implications of each and every belief and this, surely, would be a task beyond human endeavour. Nevertheless, one may press the point and ask whether a mind exists which is, conclusively, inconsistent. Some may well regard this question as a joke (with perhaps the exception of those who have been on jury service and witnessed informal logic applied by respectable citizens of sound mind), but, since most people take exception to being called "inconsistent", it must be considered. I will discuss the question initially in a general sense before returning to a statistical analysis, the latter of course being a sub-domain of consistent mathematics.

Let us consider, in the first instance, a subject who holds axioms which are patently false. Our model is a person with a high level of creativity, who is also extremely competent in the field of mathematical logic, but is a confirmed gambler to the extent that he leaves himself penniless. That such a person could exist cannot be denied and we must remember that, according to Lucas’s schema, this subject can assert the consistency of his formal work by reference to his informal reasoning. Now our subject’s ‘gambling axioms’ must belong to the latter classification and these would probably consists of one more of the following: –

G1  I am certain I am going to win, eventually.

(Like Lucas and consistency he ‘just knows’ this and ‘just knows that he knows it’ and so on for ever backwards until the questions or the money peter out).

G2  I am much more intelligent in the gambling field than most other (or all) gamblers.
G3  Life itself is but a gamble and no one accumulates great fortunes who is not a gambler.

Now try as we might with arguments from the arsenals of logic and statistics, we cannot get him to recognize that these axioms are false. But this does not mean that he is inconsistent; in a society of gamblers he would be perfectly consistent. That is my view but I do not think Lucas would share it. He has argued that anyone who uses a rule of inference in one situation and refuses to use it in another one in order to avoid an inconsistency is inconsistent, tout a fait.

“We do not lay it to a man’s credit that he avoids contradiction merely by refusing to accept those arguments which would lead him to it, for no other reason that otherwise he would be led to it. Special pleading rather than sound argument is the name for that type of reasoning. No credit accrues to a man who, clever enough to see a few moves ahead, avoids being brought to acknowledge his own inconsistency by stone-walling as soon as he sees where the argument will end. Rather, we account him inconsistent too, not, in his case, because he affirmed and denied the same proposition, but because he used and refused to use the same rule of inference. A stop-rule on actually enunciating an inconsistency is not enough to save an inconsistent system from being called inconsistent.” (4)

This leaves us with a rather pretty problem. Acting on my null hypothesis that mathematics is consistent (via informal reasoning) I have come across a person who could well assure me (again via his informal reasoning) that mathematics is consistent but whom Lucas would have to classify as informally inconsistent. However there is a loophole. Perhaps our gambling
logician needs informal axioms $A_1$, ..., $A_k$ plus some set of inferential rules to convince himself of the consistency of mathematics and this is a different schema from that applied in his gambling world. The one could be consistent and the other inconsistent. Or, again, perhaps it is an analogous situation to my being able to show that a small portion of the formal system representing Principia is consistent whilst being unable to show that the entire system is consistent. But if we do assert that the informal logic necessary for knowing mathematics to be consistent, is consistent, what grounds have we for stopping there? Could not another philosopher make an equal claim regarding his philosophy of life / religious axioms? Nevertheless all these arguments are irrelevant to Lucas’s thesis as I will demonstrate. In fact I will go further and prove that anyone who believes in his own consistency, and thereby in the consistency of mathematics, and has also given serious consideration to any of the arguments expressed in this paper is, by that very act, inconsistent.

I began this investigation by defining as null hypothesis that ‘my informal reasoning is consistent’ and by implication mathematics, and considered an informal axiom space consisting of $2^n$ independent elements. This space was divided into consistent and inconsistent subsets of axioms and on ascribing probability $p_i$ to the event that a particular mind accepts subset $S_i$ as its informal axioms or beliefs I found that the probability of such a mind being consistent was, $Pr(\text{Con}) = \sum_{i=2^n-3^n+1}^{2^{2n}-1} p_i$ (n = 1, 2, ...). The evaluation of $Pr(\text{Con})$ depends on what I loosely termed as ‘another source’ unknown to me or anyone else, with the exception of Lucas who knows that $Pr(\text{Con})$ in his case is unity. Now Lucas knows this because he thinks he knows what truth is although, unfortunately, he

“…cannot tell anybody else exactly what it is.” (5)

He believes he is consistent because

“it is, in the language of mathematical theology, an act of faith.” (6)

Lucas, in his paper, seems to interchange the words ‘believe’ and ‘know’ which is reasonable since if one knew something to be true than, I presume, one could believe it to be true, although one could believe something to be true without knowing it to be true.

I am now going to make the assumption that I know I am informally consistent, for the same reasons as Lucas does. In this paper I have considered some arguments which make me doubt this belief and, being a person who thinks he is striving very hard to apply logic in a consistent fashion, I must accept these arguments as counter evidence. Representing the probability that I am informally consistent (C) because of my Lucasian reasons (L) by $Pr(C/L)$ it follows that $Pr(C/L, F) = Pr(C/L)$ where $F$ represents my arguments for not –C. Now I know ‘that C’, hence $Pr(C/L)$ must be unity. I know also that mathematics is consistent and since Bayeslan Statistics is a branch of mathematics I can (using Bayes’ Inversion Theorem) (7) confidently write down the equation

$$Pr(C/L, F) = Pr(C, L) \times \frac{Pr(F/L, C)}{Pr(F/L)} = 1 \times \frac{Pr(F/L, C)}{Pr(F/L)}$$
Again, relying on the consistency of mathematics, it follows that the term \( \frac{\Pr(F/L, C)}{\Pr(F/L)} \) must be less than or equal to unity i.e. the denominator must not be less than the numerator. Furthermore, since the probability of my being informally consistent, given Lucas’s other sources’ \( L \), is unity, the probability of my anti-arguments (\( F \)) being true given \( L \) and \( C \) can be no smaller than the probability of \( F \) given \( L \) only. This is so simply because \( C \) adds no information independent of \( L \). Therefore, since the numerator of \( \frac{\Pr(F/L, C)}{\Pr(F/L)} \) cannot be smaller or larger than the denominator it follows that \( \Pr(C/L, F) = 1 \times 1 = 1 \).

The immediate implication is of course that no evidence of any kind can refute the claim that I am (know myself to be) consistent. But as a person who thinks of himself as being rational, one of my informal axioms must be that I make a proper use of reason: and this implies, among other things, that I endeavour to correctly estimate the strength of evidence. My \emph{knowing} that I am consistent however obviates this latter estimation. If \( L \) is to be a conclusive reason for \( C \) then \( L \) must confer upon \( C \) the probability of 1, and this means that no fact \( F \) can serve as an argument against \( C \), and it seems to me that Lucas or I could not \emph{know} ourselves to be informally consistent unless the evidence \( L \) could be expressed in the form of a deductive proof. This, as we know, is impossible.

Let us finally review the situation. I began by accepting as null hypothesis that ‘I am informally consistent’, considered a number of arguments against this belief which were not and could not be conclusive, but which were strong enough to make me feel happier believing the alternative hypothesis that ‘I do not know whether or not I am informally consistent’. I then made the assumption that I \emph{knew} I was consistent (and by implication that I \emph{knew} mathematics to be consistent), expressed my argument in mathematical terms and arrived at a conclusion which could only be considered as a travesty of rational thought. The case may be expressed symbolically as follows:

\begin{align*}
1. & A \supset (B \land C) \\
2. & A \quad \text{This is Lucas’s axiom from (1), (2)} \\
3. & B \quad \text{by Modus Ponens and tautology.} \\
4. & C \\
5. & C \supset \neg A \quad \text{This is the expression of my Bayesian argument.} \\
6. & A \supset \neg C \quad \text{from (5)} \\
7. & \neg C \quad \text{from (2) and (6).} \\
8. & C \land \neg C \quad (4) \text{ and (7).}
\end{align*}

The axiom to be rejected is \( A \), that is, for consistency I must conclude that I do not know whether or not I am consistent.

And that concludes my case. The result, it seems to me, is much more intuitively plausible that either of the extremes adopted by Lucas or Putnam. \( \text{(8)} \) It means of course that Gödel’s theorems have no implications for the mechanistic thesis; if we ever do produce a machine which, it is claimed, replicates a rational being’s mental functioning, then there is one unaskable question.

\textbf{Notes}
(*) Philosophy, 1976, 1, pp. 135-144. © Philosophy. Republished by permission. back
(**) I am indebted to John Watling, U.C.L., for criticism of an original version of this paper. back
(2) Ibid., p. 267. back
(3) Ibid., p. 268. back
(6) Ibid., p. 158. back