THE UNIVERSITY OF ALBERTA

CONSISTENCY, MECHANICALNESS AND INCOMPLETENESS

BY

QIUEN YU

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR DEGREE

OF MASTER OF ARTS

DEPARTMENT OF PHILOSOPHY

EDMONTON, ALBERTA

(SPRING, 1989)

ISBN 0-315-52902-4

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: QIUEN YU

TITLE OF THESIS: CONSISTENCY, MECHANICALNESS AND INCOMPLETENESS

DEGREE: MASTER OF ARTS

YEAR THIS DEGREE GRANTED: 1989

Qiuen Yu
....
(Student's signature)

c/o Sixing Yu, Kun Lung Road #17
....
(student's permanent address)

Nanjing, Jiang Su. P.R.China
....

Date: Jan, 13, 1989

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Consistency, Mechanicalness and Incompleteness" submitted by Qiuen Yu in partial fulfillment of the requirements for the degree of Master of Arts.

Supervisor

Date: January 4, 1989

ABSTRACT

Both appealing to Gödel incompleteness, Lucas' anti-mechanist argument [26-30] and Priest's anti-consistency argument [37-39] set mechanicalness and consistency in apparent tension. With the two arguments expressly reconstructed, it is shown that they are logically equivalent in stating that consistency and mechanicalness are incompatible due to Gödel incompleteness. By way of refuting both arguments, this thesis is designed to defend the compatibility of consistency, mechanicalness and incompleteness without appeal to the absolute truth of either mechanism or the thesis of consistency. For this purpose the thesis pursues a representational approach, which has its origin in works by, among others, Benacerraf [6], Myhill [31], Webb [44-45] and Arbib [2]. The point of the approach is that the logic of the mind can overcome its own Gödel incompleteness not by itself being inconsistent or beyond the mechanistic but by its being capable of representing stronger systems in itself; and so can a proper machine. In light of the approach, it is shown that the two arguments share the same pitfall of misidentifying the logic of the mind with the system or systems in which the Gödel sentence for the logic of the mind is provable. It turns out that the logic of the mind can be both mechanical and consistent because it is both literally

Gödel incomplete and representationally Gödel complete. Concepts such as Gödel representability, quasi Gödel representability, representational provability, the restricted universal Turing machine, and so on, are introduced along the way. About one third of the volume is given to a review, which helps show, among others, why the two arguments have not ceased to be challenging in spite of there having been many criticisms of them.

Acknowledgement

I would like to take this opportunity to thank my dear Professors from whom I received great assistance and encouragement for my thesis work. I am most grateful to Professor B. Linsky, my supervisor, who gave me invaluable guidance and help, not only for my thesis work, but also for all my graduate study at the University of Alberta. I am also deeply indebted to Professors M. Matthen and P. Schouls, who instructed and helped me in many ways, and especially, to Professor J. Pelletier, whose generous comments on my thesis draft were invaluable.

Q. Yu

Table of Contents

0. **List of Terms and Symbols** vii

1. **Introduction** 1

2. **A Review of Some Criticisms of the L- and the P-argument** 17
   - 2.1. Soundness, Formalizability, and Chihara's Considerations 18
   - 2.2. Axiomatizability, and Napoli's Observation ..27
   - 2.3. Absolute Undecidability, Butrick, and Lacey and Joseph's Viewpoints 30
   - 2.4. Consistency, and Putnam's Criticism 34
   - 2.5. Turing machine Specification, and Dennett's Reasoning 37
   - 2.6. Finite Mechanism, and Kirk's Approach 41
   - 2.7. The Sense of Provability, Benacerraf's Idea and His Argument 45
   - 2.8. Effectiveness, and Myhill's Theorem..........52

3. **An Examination of the P-argument** 57
   - 3.1. A Preliminary Examination and the P-identification 58
   - 3.2. Recursiveness in General and the Special System Identification 61
   - 3.3. The P-inference 65
   - 3.4. O-provability and The Containment Principle of Interpretation 67
   - 3.5. R-provability and The Representational Principle of Interpretation 72
   - 3.6. The QG-inference and Its Justification 76
   - 3.7. The Conclusion 88
   - 3.8. Appendix I (The Assumption of Non-classical Logic) 90

4. **An Examination of the L-argument** 94
   - 4.1. A Preliminary Examination 94
   - 4.2. Gödel Incompleteness and Turing Machines 96
   - 4.3. Appendix II (Four Technical Proofs) 110

Bibliography 113

List of Terms and Symbols

Assumption of classical logic. 90
Acceptability. 73, 85
B-argument. 48
B-idea. 47
Conditional anti-mechanism. 10
C-principle. 69
Finite mechanism. 41
Formalizability assumption. 19
G1'. 60
Gödelian system. 6
Gödel's first incompleteness theorem (G1). 5
G-representability. 74
Gödel sentence for LM (G_{LM}). 6
Gödel sentence for system S (G_S). 5
Gödel representational completeness. 110
Infinite RUT-machine. 103
JC. 14
JI. 13
$MI1_C$. 69
$MI1_R$. 77
$MI1_R$". 80
$MI1_R$'. 80
O. 82
O-provability. 68
P-argument. 6
P-identification. 61
P-inference. 66
Point of effectiveness. 53
$Provability_1$. 61
$Provability_2$. 61
QG-representability. 74
QG-representationally maximal. 77
QGR-provability. 76
R. 82
Representability. 74
Restricted universal Turing machine (RUT-machine). 100-101
R-inference. 81
R-principle. 72
R-provability. 72
S*. 48
Soundness principle. 18
Strong finite mechanism. 43
Theoremhood representative set (TR-set). 99
Thesis of anti-mechanism (TA). 3, 12
Thesis of consistency (TC). 2, 12
Thesis of inconsistency (TI). 3, 12

Thesis of mechanism (mechanism, TM). 2, 12
Thesis of recursive capacity (TRC). 6, 12
Thesis of the recursiveness of proof. 62
Thesis of true inconsistency. 8
T-consistency. 106
T-equivalence. 99
Uniformity assumption. 83
Z. 9

1. Introduction

The general topic of this thesis is the logical nature of the mind. In approaching the logical nature of the mind we can, as we will in this thesis, address the mind as a logic, a system of logical means including possibly non-logical axioms or their equivalents. Let us call it the logic of the mind, LM.[1] With reference to the logic of the mind there are two principal issues, the issue of mechanicalness and the issue of consistency. The former issue is significant to the way the logic of the mind operates; whereas the latter is pertinent to the outcome of the operation of the logic of the mind. Correspondingly, there are two principal questions concerning the logic of the mind: First, can the workings of the logic of the

[1] Here, the logic of the mind is supposed to be a static logical system. This supposition does not contradict the fact that a real mind may change itself by receiving and processing information from outside. For so long as we are supposed to be able to imagine the whole truth of arithmetic, we are entitled to address the logic of the mind as complete as possible with respect to arithmetic. We admit that the arguments based on this supposition may not be directly applicable to many aspects of real minds. However, it is the ideal mind that is understood as the primary concern of this thesis. In this regard, our term 'the logic of the mind' means the same as does J.R. Lucas' term 'a complete or adequate model of the mind',([26], Lucas, 1961, p. 44) G. Priest's terms 'the naive canons of proof', and 'the proof procedures of proof',([41], Priest, 1987, p. 51, p. 55) or R.J. Nelson's term 'the logic of mind'([37], Nelson, 1982, p. 2) in their contexts respectively, with the provision that it is an open question as to what logical rules and axioms (logical or non-logical) are in the logic of the mind.

mind be explained completely in terms of recursion? Second, may the outcome of the logic of the mind be consistent? Positive answers to these two questions are the following two theses:

> **The thesis of mechanism** (or mechanism). The logic of the mind is essentially mechanistic in the sense that it can be explained exclusively in terms of recursion.[2]

> **The thesis of consistency**. The logic of the mind is essentially consistent in the sense that it does not necessarily contain any contradiction.[3]

---

[2] There are various formulations, versions and supposed connotations of mechanism. We think that the above formulation of mechanism is in essence the same as is presented by J. Webb ([48], 1983, p. 310). A brief review of some of the versions of mechanism can be found in ibid, pp.310-312; and [47], Webb, 1980, p. 235. An analysis of a couple of supposed connotations of mechanism is given by Nelson ([37], 1982, pp. 2-6). Nelson himself subscribes a version of finite mechanism to the effect that the logic of the mind can be characterized as a Nondeterministic finite automaton (NFA).(ibid, p. 2.) A very interesting historical investigation of mechanism and its associates from a contemporary perspective is contained in Webb's [47], Chapter 1. There is meaning in looking for better and better formulations of mechanism. But we think the above formulation is adequate enough for us to examine Lucas' argument of antimechanism as will be given shortly.

[3] The thesis of consistency should be distinguished from the law of contradiction, or rather of non-contradiction. Strictly speaking, the law of contradiction, being a normal rule of logic, only decrees that any contradiction, once uncovered, ought to be consciously eliminated from our rational universe. The law of contradiction per se does not entail that it is always possible to eliminate all uncovered contradictions from our rational universe.

Recently, two arguments have been put forward respectively against mechanism and the thesis of consistency. The first is initiated by J.R. Lucas, to the effect that Gödel's first incompleteness theorem (hereafter, G1) implies antimechanism --- the negation of mechanism. ([28], Lucas, 1961)[4] The second is presented by G. Priest, to the effect that G1 implies the thesis of inconsistency --- the negation of the thesis of consistency. ([39-41], Priest, 1979, 1984 and 1987) As for the implications of the two arguments, we simply quote from Lucas and Priest

---

It is the tenet of the thesis of consistency to state that it is always possible to do so regardless of whether one really intends to do so. Of course, these two principles are closely related. The law of contradiction would lose almost all its meaning if the thesis of consistency did not hold, That is why Professor G. Priest suggests, after being convinced that the thesis of consistency does not hold (though he has not used the term 'the thesis of consistency`), that the law of contradiction should be rejected.([39], Priest, 1979, p. 220 and [41], Priest, 1987, p. 258-259) Moreover, the thesis of consistency does not carry the connotation that the logic of the mind must be consistent. For, that the logic of the mind is not necessarily inconsistent does not imply that it is necessarily consistent. Rather, it says just that the logic of the mind is possibly consistent. In section 4.2, we will touch this point again in light of our explication of the correlation between formal systems and Turing machines.

[4] According to Webb, E. Post was the first, in 1924, to prove undecidability and to conceive of its consequences taken to be against mechanism. Webb quotes Post, 'We see that a machine would never give a complete logic; for once the machine is made we could prove a theorem it does not prove.' But upon consideration Post says, to the contrary, 'The conclusion that man is not a machine is invalid.' (See [38], Davis (ed), 1967, p. 417 and p. 423; and [47], Webb, 1980, p. 229.)

the following two passages:

> 'If the proof of the falsity of mechanism is valid, it is of the greatest consequence for the whole of philosophy.'([28], Lucas, 1961, p. 59)

> '[t]hey (Priest's points made in his book of 1987 against the thesis of consistency, of which the above mentioned is a substantial one) are a sustained attack on the dominant logical theory of our times, the logic of Frege, Russell and their successors, or, as it has come to be known, classical logic.'([41], Priest, 1987, p. 257)

We need not say more than the above in this regard. In opposition to both Lucas and Priest's arguments, this thesis is devoted to defending mechanism and the thesis of consistency together by way of refuting them both in one line.

Obviously, if either Lucas or Priest is correct, then, other conditions granted, either mechanism or the thesis of consistency is incompatible with G1. Since the validity of G1 is beyond question. So, if either Lucas or Priest is correct, then either mechanism or the thesis of consistency must be rejected. At first sight, there seems no significant connection between Lucas and Priest's arguments. However, it can be observed that the two arguments are so closely related that they both appeal to G1 and the thesis of recursive capacity (see below) and both purport to show that mechanism and the thesis of consistency are incompatible. To see this, we explicate Lucas and Priest's arguments through the following two reconstructions of them:

A. The L-argument (a reconstruction of Lucas' anti-mechanist argument)

A1 (G1). For any system[5] S which is a). recursively enumerable (i.e.) and b). capable of representing all means of recursion there is a Gödel sentence G_S for S which says that G_S is unprovable; further, it would indeed be unprovable in S if S is consistent.

A2. G_S, because of its meaning and its unprovability in S provided that S is consistent, can be seen to be true and therefore in this sense be provable <u>in the logic of the mind</u>.

A3. The system S addressed in A1 is arbitrary so lang as it satisfies the two clauses a) and b)

---

[5] All theoretical systems addressed in this thesis are first order ones unless otherwise indicated. This technical restriction will not in essence affect the generality of our argument. For any higher order system in a hither order language can in a way be represented equivalently by a first order system in a first order language. A discussion about this point is given in ([4], Bell and Machover, 1977, pp. 14-15). As for the logic of the mind, presumably, there is no prior reason for us to suppose that it has to be of first order. However, it is still meaningful to investigate whether or not the logic of the mind as restricted to the first order dimension (supposing that it has other dimensions which are irreducible to the first order) has to be beyond the mechanistic or has to be inconsistent, in view of Gödel incompleteness. To our understanding, Lucas, Priest and almost all their critics admit that their arguments can be, as they are really presented, developed under the first-order restriction.

in A1. Let any such system be called a **gödelian system** and let LM refer to the logic of the mind. If LM is itself a gödelian system, there would be a Gödel sentence for LM, say G_{LM}. By A2 G_{LM} would be provable in LM if LM is consistent. By A1 however, the provability-in-LM-of G_{M} means that LM is inconsistent.

A4. The thesis of consistency holds. So, LM is essentially consistent.

A5. From A3 and A4 LM cannot be a gödelian system.

A6 (The thesis of recursive capacity). LM is capable of representing all means of recursion.

A7. From A5, A6 and the definition of gödelianness, LM cannot be r.e.[6] So, mechanism is false.

B. The P-argument (a reconstruction of Priest's anti-consistency argument)

B1. The same as A1 (G1).

B2. The same as A6 (the thesis of recursive capacity)

B3. Mechanism holds. So LM can be explained and realized exclusively in terms of recursion.

B4. From B2 and B3 LM is a gödelian system, or

---

[6] Note here, a system which is capable of representing all means of recursion may not be itself recursive.

can be completely formalized as a gödelian system.

B5. From B1 and B4 LM has its own Gödel sentence, say G_{LM}, which says that G_{LM} is unprovable in LM and it would indeed be unprovable in LM if LM were consistent.

B6. G_{LM}, because of its meaning and its unprovability in LM (provided that LM is consistent), can be seen to be true and therefore in this sense is provable in LM.

B7. From B5 and B6 LM is inconsistent. Moreover, suppose LM is consistent, then G_{LM} is both provable and unprovable in LM. So the thesis of consistency is false.

The above two reconstructed arguments can be regarded as faithful simulations of Lucas' and Priest's original arguments to the effect that they are virtually refutable if and only if so are their original counterparts.[7] The case for Priest's is obvious. Priest's original argument is explicitly equivalent in essence to that formulated above.(Priest, [41], 1987, p. 56; [40], 1984, p. 165; and [39], 1979, p. 221) Here, one subtle as well as important

7 There are many succinct reconstructions of Lucas' original argument. See for example, [8], Boyer, 1983, p. 148; and [25], Kirk, 1986, pp. 437-438. A brief reconstruction of Priest's original argument is given in [34], Napoli, 1985, p. 405.

point about the meaning of the thesis of inconsistency needs explaining. That is, the thesis of inconsistency does not connote that contradictions which are provable are true in any sense. The thesis which connotes just this may be called the **thesis of true inconsistency**, in contrast to the thesis of inconsistency. Priest does attempt to distinguish them and to argue for them both, though he neither calls them by the same names as we do now nor highlights the distinction between them.([40], Priest, 1984, p. 171) However, in this paper only the thesis of inconsistency is our central concern. We are not going to address the thesis of true inconsistency directly, though we do tend to think that this thesis is either untenable or pointless. In view of the above point, the P-argument, i.e., Priest's argument of inconsistency as reconstructed above, is only a part of his whole argument of true inconsistency. One thing is clear, merely maintaining the thesis of inconsistency is not sufficient for those who want to defend the thesis of true inconsistency, whereas a rejection of the thesis of inconsistency is enough for those who want to dismiss the thesis of true inconsistency.[8]

[8] If the thesis of inconsistency is rejectable, there will be no contradiction which is necessarily provable, let alone to be necessarily true. In section 2.1 we will touch this point again.

The case for Lucas needs more explanation. First, Lucas has developed many arguments for antimechanism through a long process of his rejoinders with his opponents. Not all of his arguments are even partially reflected in the L-argument as formulated above.([28-32], Lucas, 1961, 1968, 1970, and 1984) However, we believe that the core of his arguments is well represented in the L-argument. Second, exactly speaking, Lucas does not seem to have the thesis of consistency invoked explicitly in his original argument in the way in which it is above, though he does invoke the thesis of consistency in his own way. Taken succinctly, his reasoning consists of the following two points: First, in order to show that, in view of G1, the logic of the mind is indeed essentially different from any gödelian system, it is necessary to show that the elementary Peano arithmetic (hereafter, Z) is consistent. For otherwise Z and hence any extension of Z would be inconsistent, making everything including G_Z (the Gödel sentence for Z) provable in them and therefore rendering G1 irrelevant to the concern of mechanicalness. In this case, there is no way for LM to be distinct from Z in virtue of its having G_S for any gödelian system S including Z provable in LM. Second, in order to have a sound proof of the consistency of Z, it is necessary to show, or at least to assume that LM is consistent.

Otherwise Z may still be inconsistent even though its consistency is 'provable' in LM. Thus Lucas invokes the thesis of consistency in order to show the consistency of Z from (the demonstration of) the consistency of LM. To be sure, Lucas is completely correct in doing so, regardless of whether his whole argument is correct or not. We did not let this piece of Lucas' reasoning appear in the L-argument merely for the sake of simplicity with a view to intensifying the main contrast between Priest and Lucas. On the other hand, few people would really worry about the consistency of Z. At least, without appeal to the consistency of Z we could establish a conditional version of anti-mechanism, say **conditional anti-mechanism**, to the effect that if Z is consistent, then LM is essentially nonmechanistic. Nonetheless conditional anti-mechanism would be enough for many anti-mechanists, Lucas included. For conditional antimechanism to hold, it is not necessary to invoke the thesis of consistency to prove the consistency of Z. In this connection the L-argument shows that the thesis of consistency must be invoked in a way even for conditional antimechanism to hold, without regard to the consistency of Z. In fact, there is presumably a possibility that Z is consistent whereas LM is inconsistent. If that is the case, even though we can prove a priori that Z is consistent and in consequence that G_Z is indeed unprovable in Z, we may still not be able to show that LM is essentially distinguished from being gödelian in virtue of the fact that the Gödel sentence G_S for any gödelian system S is provable in LM. For presumably, there are two possible cases in which LM may be able to prove the Gödel sentence G_S for each gödelian system S. One case is that LM is inconsistent. In this case everything is, according to our common understanding[9], provable in LM. The other case is that LM is capable of something beyond the capacity of any gödelian system. In this case, LM may be capable of proving something of which no gödelian system is capable. Since what Lucas intends to show is certainly the second case, i.e. that LM can be used to prove G_S for every gödelian system S just because LM is capable of something beyond the mechanistic, he has to invoke the thesis of consistency to exclude the first case. The fact that Lucas fails to invoke the thesis of consistency in this way explicitly does not mean that the L-argument as constructed above must fail to be a faithful simulation of Lucas' original argument. Rather, it may well be viewed as a defect of Lucas' original argument. Besides, there are places at which Lucas invokes the

---

[9] We confine ourselves within the discipline of classical logic. Later in section 3.8 we will show that the assumed distinction between the classical and the non-classical is not significant for our concerns from a representational point of view.

thesis of consistency in other ways. A prominent one is seen in his reply to a critical point made by Hilary Putnam.([42], 1960, p. 77) We will review it in some detail in section 2.4. What these show is that the role of the thesis of consistency in Lucas' various arguments for anti-mechanism is crucial in many ways.[10]

Now, in order to highlight some interesting aspects of the two reconstructed arguments, we use some informal symbolization, namely:

TM --- The thesis of mechanism,

TC --- The thesis of consistency,

TA --- The thesis of antimechanism,

TI --- The thesis of inconsistency,

TRC --- The thesis of recursive capacity.[11]

Using the above symbols, we can draw from the L- and the P-arguments two informal formulas:

L1. [G1 & TC & TRC] => TA (-TM)

and

---

[10] Any way, at the very least, the value of the L-argument and hence its criticisms are independent of who for what first formulates it as such. Moreover, one who is convinced of conditional antimechanism may not be convinced of antimechanism; whereas if we can succeed in refuting conditional antimechanism, we will be in a better position to refute antimechanism.

[11] Note, as we indicated before, anti-mechanism is the negation of mechanism and the thesis of inconsistency is the negation of the thesis of consistency. So, TA≡-TM and TI≡-TC.

P1. [G1 & TM & TRC] => TI (-TC).

Both L1 and P1 are equivalent to a third one:

JI. -G1 v -TRC v -TM v -TC.

JI means exactly the joint incompatibility of TM, TC, G1 and TRC, or say the joint incompatibility of TM and TC given G1 and TRC. Now, since the validity and the truth of G1 and TRC are much less controversial than that of TM and TC, the apparent tension or incompatibility as shown by JI is supposed to be primarily between TM and TC. This incompatibility has been recognized by Priest. He says, in a brief comment on Lucas' antimechanism:

> 'Both the anti-mechanist and I agree that the recursiveness and the consistency of proof are not compatible.'([40], Priest, 1984, p. 168)[12]

For both Lucas and Priest, this incompatibility is necessary regardless of which one of the two theses TC and TM is true.

L1 and P1 each captures the logical relations amongst TM, TC, G1 and TRC as demonstrated in the L- and the P-

---

[12] In this connection Priest adds also that 'to infer the nonrecursiveness of proof, therefore, invokes consistency and hence begs the question.'([38], Priest, 1984, p. 168) This added remark is somewhat misleading. For the joint incompatibility of the recursiveness and the consistency of proof is not given a priori. It is only to be proved or disproved. As the following discussion shows, the problematic point here is not of begging the question, but of taking what is uninferable for inferable. If Priest were right in this regard, then for the same reason, to infer inconsistency from mechanism begs the question, too, though in a converse way. But that is just what Priest has consciously done.

arguments. Nonetheless they each also miss some contents which make the two arguments themselves in tension. To have these contents included, we use the following informal formulas:

L2. G1 & TC & TRC & [[G1 & TC & TRC] => TA (-TM)]

and

P2. G1 & TM & TRC & [[G1 & TM & TRC] => TI (-TC)].

Since L2 and P2 are equivalent respectively to

L2'. G1 & TC & TRC & TA (--TM)

and

P2'. G1 & TM & TRC & TI (--TC),

the tension in question is obvious. This tension is a predicament for both mechanism and the thesis of consistency.

In our view, neither L2 nor P2 nor JI may be true; what may be true is the negation of JI, i.e.

JC. G1 & TM & TC & TRC.

In this thesis we will not argue for mechanism or the thesis of consistency directly, instead, towards establishing JC as our conceived final goal we will try to demonstrate that TM and TC are compatible with each other, and with G1 and TRC as well. In other words, we intend to show that there is no logical incompatibility between mechanism and the thesis of consistency under G1 and TRC even though it is impossible for us to establish the truth of either of mechanism and the thesis of consistency absolutely.

In the literature, both the L- and the P-argument have been criticized in many respects, though very few if any of these critiques are designed explicitly with the view of defending both mechanism and the thesis of consistency. In order to have an appropriate outlook of the present status of the issues raised by Lucas and Priest, we will make a brief review of some criticisms of them in chapter 2. The main body of this thesis will be formed in chapters 3 and 4, in which we present our own examinations of the L- and the P-argument, with the larger part being against the P-argument in chapter 3.

Our general viewpoint is that since the provability concept is in general system-relevant, the provability concept according to which G1 holds and the one according to which G_{LM} is viewed as provable in LM may well be different; and because of this both the L- and the P-arguments prove to be invalid. To vindicate this viewpoint, we take what may be termed a representational approach, which owes its origin to works by, among others, Benacerraf ([6], 1967), Myhill ([31], 1964), Webb ([47-48], 1980, 1983), and Arbib ([2], 1987). The central idea is that the provability in LM of G_{LM} as is usually recognized means and only means that G_{LM} is provable in a system which is representable in LM. Accordingly, the logic of the mind can overcome its own Gödel incompleteness not by itself being inconsistent or beyond the mechanistic, but by its being capable of representing stronger systems in itself. For the sake of convenience, we develop this central idea mostly by way of examining the P-argument in chapter 3. In chapter 4 we will show that the idea developed in our examination of the P-argument is applicable to that of the L-argument. If we are right in this regard, it then means that, among other things, the two arguments share the same pitfall: misidentification of the two provabilities.

2. A Review of Some Criticisms of the L- and the P-argument

In this chapter, we review briefly some critical remarks relevant to the L- and the P-argument respectively, with a view to getting a proper background to develop our own points and to examine them by contrast. Originally, some points reviewed in this chapter might not be made with reference particularly to the L-and the P-arguments, for these two are but our present reconstructions of Lucas and Priest's original arguments. Nonetheless these points will be considered primarily in connection with the L- and the P-argument. We do so because we think that our reconstructions do reflect the essences of their original counterparts and our concern here is just confined to these reconstructions. We have no intention to cover all the issues raised by Lucas' and Priest's original arguments. Hopefully, this review will help show that although both of the arguments have been seriously challenged from various perspectives, neither of them has ceased to be itself challenging.[13] Some

---

13 To many, what is challenging is not the question whether these two arguments are really wrong, but rather the question where and why they go wrong. For example, Kirk remarks, with reference to the attempt to use Gödel incompleteness against mechanism, 'there is not much agreement on just what is wrong with it.'([25], Kirk, 1986, p. 437) We write this thesis just to explore the second question.

explicatory points such as those about the logical connections between mechanism and consistency will be especially attended to in this chapter.

We choose first to review some criticisms of Priest's argument.

2.1. Soundness, Formalizability, and Chihara's Considerations

Charles S Chihara was among the very first to challenge the P-argument.([11], Chihara, 1984) The main critical points he makes in his pioneer examination of the P-argument can be summarized as two considerations as shown below.

The first consideration is of soundness. As Chihara understands, in order to prove that the Gödel sentence G_{LM} for LM is true Priest must first, not only assume, but also <u>prove</u>

> **The soundness principle.** Whatever is provable in LM is true.

In Chihara's mind, the ultimate point Priest intends to prove is the truth of G_{LM}, but without the soundness principle the truth of G_{LM} may not be granted even though G_{LM} is provable in LM. However, even though this soundness principle is indeed true, it does not automatically follow that we could prove or would realize it to be so.([11], Chihara, 1984, pp. 120-121) Besides, if the soundness principle really holds, LM may be weaker than is needed to prove the truth of G_{LM}.(ibid, p.120 and p. 123)

The second consideration is of formalizability. Chihara observes that the P-argument presupposes

> **The formalizability assumption** The logic of the mind can be completely formalized in a system that is open to G1.

In Chihara's view, this assumption is untenable.(ibid, pp. 121-122)

In our view, Chihara's considerations, though instructive or correct in some respects, are in general not relevant to or less than fatal to the P-argument. Moreover, in some respects they are even questionable. Our reasons for this general view are given in the following.

We first consider the relevance of the soundness principle. The soundness principle is necessary only for the proof that G_{LM} for LM is true, but not for the proof that G_{LM} is provable in LM. Now, in order for the thesis of inconsistency to hold, it is needed only to demonstrate, without necessary reference to the thesis of true inconsistency, the provability in LM of G_{LM}. As was indicated in chapter 1, the thesis of inconsistency should be distinguished from that of true inconsistency. Certainly, the first thesis does not imply the second, nor does the refutation of the second imply the refutation of the first. Since our primary interest lies only in a refutation of the first thesis, Chihara's consideration, for all its strength in questioning the thesis of true inconsistency, is not very relevant to our present concern. In fact, checking the two arguments that Priest gives for the truth and the provability in LM of G_{LM} in two places, we find that all the appeal to the soundness principle can be omitted if the purpose of the proofs is limited to establishing the thesis of inconsistency. At the first place, Priest tries to spell out the intuition that g 'expresses its own unprovability and hence is true' by giving the following argument:

(1) $\exists x Prov(xg) \Rightarrow$ ''$\exists x Prov(xg)$' is true
(2) $\Rightarrow$ g is provable
(3) $\Rightarrow$ g is true
(4) $\Rightarrow \neg \exists x Prov(xg)$.

Then he concludes:

Hence $\neg \exists x Prov(xg)$.

(where $\neg \exists x Prov(xg)$, i.e. g is the Gödel sentence in question) He remarks:

> 'Step (2) depends upon the fact that 'Prov' really does represent the proof relation. Step (3) depends on the fact that whatever is provable in P is true. Steps (1) and (4) follow from the Tarski bi-conditional T('p')<-->p (where 'p' is the code of p).([39], Priest, 1979, pp. 222-223)

But, clearly, if this proof for the truth of g is correct, then so should the following proof, which is merely for the provability of g be, i.e.

(1) $\exists x\, Prov(xg) \Rightarrow g$ is provable

(2) $\Rightarrow \neg \exists x Prov(xg)$.

Hence $\neg \exists x Prov(xg)$.

In this proof, step (1) requires only that 'Prov' really represents the proof relation; step (2) follows merely from the definition of g. So no appeal to the soundness principle is necessary.

At the second place, Priest says:

> 'In outline the proof is this. let the sentence be A. Either A is provable in T, or it is not. If it is provable then, assuming soundness, it is true. Suppose it is not provable, Then for every integer n, $\neg P(n,g)$ is provable. Assuming that the system of proof is sound, it follows that every integer satisfies $\neg P(x,g)$, and hence that $(x)\neg P(x,g)$ is true.'([40], Priest, 1984, p. 166)

Supposing the correctness of the above proof, we can make the following one merely for the provability of A:

> Either A is provable in T, or it is not. If it is provable then, we need nothing more. Suppose it is not provable, then for every integer n, $\neg P(n,g)$ is provable. It follows that $(x)\neg P(x,g)$ is provable in LM. (Notice that LM here is supposed to be Ω-complete.)

We get the same result as the above.

Regarding the provability of the soundness principle, Priest argues that it is easy and even trivial to prove it.[14] Here we need not care much whether Priest is all right in this regard or not. If it were just the unprovability of the soundness principle that baffled the P-argument, one would be able to claim that the thesis of inconsistency could hold so long as so could the soundness principle. This means, for one thing, that if LM is sound, then there are, in LM, contradictions which are true. We guess Priest would like, but Chihara would not like, to accept this consequence.[15]

As for the relation between the soundness principle and the thesis of true inconsistency, the unprovability of the soundness principle is not enough for us to reject the thesis of true inconsistency. For what we need for a rejection of the thesis of true inconsistency is not that there is no possible formulation of LM whose soundness can be proved, but that for any sound formulation of LM no contradiction can be proved to be necessarily in it and true.

Concerning the capacity of LM in case it is sound, Chihara distinguishes N -- the system of the real naive

---

[14] See [40], Priest, 1984, p. 166, and [41], Priest, 1987, pp. 62-63, where he presents his proofs of this principle.

[15] In fact, since LM is the ideal model of the human mind, if there is no reason for us to assume soundness principle for LM, there will be no reason for us to assume the soundness of anyone's mind. Thus, the soundness principle would become meaningless.

proof procedures -- from A -- the system of the apparent proof procedures. According to him, apparent principles such as Tarskian truth formula (roughly, 'p' is true if and only if p.) should not belong to N. He thinks that 'a more fundamental objection' to Priest's argument is that if LM=N, G_{LM} may no longer be provable in LM.([11], Chihara, 1984, p. 123) Although Chihara is not sufficiently specific as to why he think so, we agree with Chihara to the extent that there should be some N such that in case LM=N no contradiction is provable in LM. (For this we need only to restrict a candidate N' as much as needed.) But we also admit, probably in disagreement with Chihara, that in a sense there is certainly a proof of G_{LM} in LM. What is needed is then to analyze the sense and show to the effect that the sense in which G_{LM} is provable is different from the one in which it is not, thus resolving the apparent contradiction. Chihara's idea is that LM must become weaker if LM is altered so as to be sound from being unsound. This idea may be true if 'weaker' is taken in the sense of containment, but may not be true if it is taken in the sense of representation.[16] Even in the sense of containment Chihara's idea may not be tenable unless he provides it with other support. For Priest just insists

---

[16] For detail please see chapter 3, especially, sections 3.4 and 3.5.

that principles such as the Tarskian truth formula be kept in LM. In view of this, the dispute between Priest and Chihara would be transformed into whether and in what sense principles such as the Tarskian truth formula should be kept in LM. Here, we agree with Chihara that the Tarskian truth formula should at least be somewhat qualified before being included in a correct formulation of LM. However, even though the defender of the P-argument admitted this, he could still question whether the proof of the provability and the truth of G_{LM} must employ the Tarskian formula any more than many other proofs do. So, if Chihara wants to falsify Priest's argument by applying the above idea, he has to show concretely, among other things, why any proof of G_{LM} cannot be carried out in a sound embodiment of LM. Since Chihara has not done so, his point in this regard is at most inconclusive.[17]

We now turn to Chihara's consideration of formalizability. Here also we beg to differ. It is true that the logic of the mind must be formalizable in order to be subject to G1. We can also be certain that **practically**, the logic of a real mind cannot be completely formalized.

---

[17] Besides, it can be pointed out that if all that matters hinges on the soundness of LM, it can then be claimed that if LM is sound or can have a sound embodiment, then Priest's argument holds. We do not know whether Chihara or Priest would like to accept this case or not. However, in this thesis we just want to show that Priest's argument does not hold even though LM is sound.

(Maybe the uncertainty principle of physics plays a role in this regard.) However, here we are concerned not with **the practical formalizability** of logics of real minds, but with **the formalizability-in-principle** of the logic of the ideal mind. What matters to our concern is whether there is a possible mind the logic of which satisfies the formalizability assumption, and its Gödel sentence is provable in a sense which ensures a contradiction. To study this question, we may first presuppose that the logic of a possible mind satisfies the first condition, i.e. the formalizability assumption, and then see whether this logic can at the same time satisfy the second condition, i.e. having its own Gödel sentence provable in a required sense. Of course, we may also first presuppose for a possible mind that it satisfies the second condition and then see whether it also satisfies the first one. Now, since Priest intends to prove the thesis of inconsistency by appeal to mechanism, if Chihara wants to falsify the P-argument, he should follow him in presupposing mechanism and then show that the thesis of inconsistency does not follow. But, if mechanism holds, or is presupposed, there should be no reason why the logic of the mind cannot be in principle formalizable and therefore be subject to G1. Chihara expresses his doubt that the axioms of LM may not be recursively enumerable.(ibid, p. 122)[18] But, again, if mechanism holds, the logic of the mind should be recursively enumerable by definition.

Chihara also questions the formalizability assumption by citing Robinson's infinitary arithmetic. According to Chihara, if the formalizability assumption held, then Robinson's infinitary arithmetic would be proved to be true from self-evident truths. But since he is one of those who believe that they need only to follow David Hilbert and Hartry Field in this regard, this is for him impossible. (ibid, p. 122)[19] Here we are not sure what is the exact sense or senses in which Hilbert and Field do reject any infinitary arithmetic as Chihara means. Perhaps Chihara means that any ideal mathematics is only instrumental (here following Hilbert) and any mathematics is

---

[18] In his text, Chihara uses the term 'effectively decidable' for being effectively, or recursively enumerable, or axiomatizable. According to standard terminology, this is a misexpression. Being effectively decidable is stronger than being axiomatizable, which and only which is necessary of a theory to be open to G1, other conditions being given. (See [4], Bell and Machover, 1977, p. 332, Theorem 7.5.4, 7.5.2; p. 356, Theorem 7.11.8.) Maybe Chihara thinks of the original version of G1 as was given by Gödel([17], Gödel, 1931) when he says that 'to be open to Gödel's theorem the axioms of P must be effectively decidable.' But, since that original version has already been strengthened to being applicable to any recursively enumerable theory which can represent all recursive functions, it would be better to invoke the strongest version of G1 in this regard.

[19] The references Chihara makes of Hilbert and Field are [22], 1925 and [15], 1980.

only instrumental (here following Field). But these kinds of instrumental viewpoints themselves are not beyond question. For example, Gödel has a tendency toward Platonism, which is at odds with instrumentalism.([18], Gödel, 1964) Moreover, even if Hilbert or Field is right in this regard, it may not legitimate to apply instrumentalism directly against the thesis of inconsistency because this thesis does not involve the matter of truth directly. Merely from the thesis of inconsistency it is not deducible that there are some truths in arithmetic, inconsistent or not.

From the above we see that Chihara's points, partly because they are mainly directed to the thesis of true inconsistency, have little force against the P-argument.

2.2. Axiomatizability, and Napoli's Observation

Succeeding Chihara and recognizing well the tension between Lucas and Priest, Ernesto Napoli tries to argue that no presupposition of consistency is required to show the invalidity of the P-argument, which is, according to him, 'nothing less than a sheer and circular presumption of inconsistency'.([34], Napoli, 1985, p. 403) In Napoli's view, there are two cases in which G_{LM} may be proved in LM. Case 1: LM is axiomatizable. In case 1, the provability of G_{LM} in LM means that LM is inconsistent. So in case 1 the provability of G_{LM} in LM is but the supposition of what is to be proved. Case 2: LM is not axiomatizable. In case 2, since the provability relation of LM is not recursive, even if G_{LM} is provable in LM, there is no contradiction because what is established (by G1 and the provability of G_{LM} in LM) is only that

(a). G_{LM} is not effectively provable in LM; and

(b). G_{LM} is non-effectively provable in LM.

Napoli's implicit view is that any proof conducted in a nonaxiomatic, i.e. non-r.e. theory is non-recursive; so that if G_{LM} is only non-effectively provable in LM, then there may be no contradiction which is necessarily involved, for clauses (a) and (b) do not constitute a direct contradiction. Here, Napolil sees that the pitfall of the P-argument may lie in the P-identification.[20] His observation contains some truth. But this truth is not enough to invalidate the P-argument. First, as Napoli notices, Priest himself claims mechanism. Since mechanism means that any proof of LM is recursive, it follows that, unless special explanation or qualification is given, case 2 should be excluded from consideration. In fact, if G_{LM} could be non-recursively provable in LM, there would be no Gödel sentence for LM. For in that case, the provability predicate for LM would be non-recursive, and hence G1

---

[20] Please recall the last paragraph of chapter 1; also see the beginning paragraph of chapter 3, and especially section 3.1 for detail.

would not be applicable to LM and the Gödel sentence for LM could not even be constructed. Priest provides several powerful arguments to the effect that any proof conducted in LM must be recursive and hence in particular so must the proof of G_{LM} be. Agreeing with Priest in this regard, we take the view that any proof is recursive so long as it is a proof, whether or not it is conducted in a non-r.e. theory.[21] If Napoli thinks that the P-argument in this regard is incorrect, he should, as he did not, let his reader know why. Second, pointing out the equivalence between the claim of the recursive provability in LM of G_{LM} and the claim of the inconsistency of LM in case 1 is not enough to show that G_{LM} cannot be recursively provable in LM. For certainly, circularity is not invalidity. Moreover, just because of this circularity (or better, equivalence), if Napoli wants to renounce the inconsistency of LM without begging the question, he has to give proper reason to show why G_{LM} cannot be recursively provable in LM without appeal to the consistency of LM. Otherwise Priest would also have good reason to accuse Napoli of begging the question in return. Since Napoli has not done so, his argument cannot be viewed as a satisfactory refutation of the P-argument,

---

[21] See [40], Priest, 1984, 167-170. Our detailed discussion about this will be given in section 3.2.

though his observation is appreciable.

2.3. Absolute Undecidability, Butrick's, and Lacey and Joseph's Viewpoints

We now consider the problem of undecidability of the Gödel sentence directly. Concerning this Richard Butrick holds a special viewpoint, which, though not raised for the purpose of attacking Priest's argument, is nonetheless quite relevant to it. This viewpoint is

> **The absolute undecidability viewpoint.** '[t]he Gödel formula, (i.e. Gödel sentence) is not decidable in any system, and in that sense it is a pseudo sentence'.([9], Butrick, 1965, p. 144)

Butrick's argument for his viewpoint is the following:

> The Gödel formula contains a designatory function, and the claim that the number named by that function does not stand in the relation "Dem" to any other Gödel number. To determine whether what it says is true, one must determine whether the formula associated with that number is demonstrable (either an axiom or obtained from the axioms by the rules of substitution and transformation). That formula, it turns out, has the Gödel number of the Gödel formula and hence contains a designatory function and the claim that the Gödel number named by that function does not stand in the relation "Dem" to any other Gödel number. To determine whether this latter claim is true one must proceed again in the same manner, and so on endlessly, obviating the notion that any decidable claim is made at all.'[22]

Obviously, if Butrick's argument were correct, the P-

---

[22] See ibid above, p. 414, wherein 'Dem' is the expression which Gödel used as the provability relation in his original article in which G1 was first proved. See [17], Gödel, 1931.

argument would have stumbled at step B6, at which it is stated that G_{LM} can be proved in LM.

The validity of Butrick's argument bears very much on what the Gödel sentence designates by its designatory function and what it claims about its designation. Hugh Lacey and Geoffrey Joseph challenge Butrick in just this regard.([27], Lacey and Joseph, 1968) According to Lacey and Joseph, in examining the decidability of the Gödel sentence we need to consider three systems: the object theory, say P, i.e. our Z; the intuitive, unformalized arithmetic, say A; and the metatheory of P, say M. Correspondingly, the Gödel sentence G (or any other sentence of the formal object language of P) can be read in three ways: as G, which is significant to P; as G_A, which is significant to A, and as G_M, which is significant to M.(ibid, p. 77-79)[23] In Lacey and Joseph's view, the problem of the undecidability of the Gödel sentence should be settled with respect to the three systems separately and independently. Thus viewed, G is indeed undecidable in P according to them, but G_A is provisionally decidable in A (in the sense of depending upon how 'strict a mathemati-

---

[23] The distinction between G and G_M is also discussed by, among others, Webb without mentioning G_A. ([48], Webb, 1983, pp. 328-329.) This distinction is enough for our concerns. In any case, we disagree with Lacey and Joseph in some aspects on the G-G_A-G_M distinction, though we will not go into detail about it in this thesis because it is not relevant to our present concern.

cal finitism one advocates'), and G_M must be decidable in M.(ibid, p. 82) When Butrick argues for his viewpoint, he obviously invokes G_M, for only G_M in M can be taken to be saying something other than what the mere technical and syntactic form of G means. But, according to Lacey and Joseph's analysis, G_M claims something only about G, not about G_M itself. Thus there is no self reference involved, and hence Butrick's argument is misleading.(ibid, p. 80) Now, we do not believe that Lacey and Joseph's argument is sufficiently tight. For example, there could still be doubt why G_M can only say something about G, but not about G_M itself, though to our understanding, there might be no decisive means for us to dissolve this doubt. At least, Lacey and Joseph do not say anything specific on this point. Nonetheless, we do think that their argument shows that Butrick's viewpoint of absolute undecidability is hardly plausible, for after all there are definitely senses in which the Gödel sentence is decidable, and moreover, true.[24]

From the above, we can say that Butrick's viewpoint is untenable. However, even a conclusive refutation of the

---

24 In fact, probably because they fail to attend to Butrick's general viewpoint adequately, Lacey and Joseph miss a simple point against Butrick. That is, the Gödel sentence read as G is simply decidable in many object formal systems (in contrast to meta ones) such as any maximal extension of P, or trivially, in any object formal system containing G.

absolute undecidability viewpoint does not imply that Priest' argument must hold. For though the above reasoning is applicable to the Gödel sentence for any gödelian system, it does not say anything specifically about the sense or senses in which the Gödel sentence for a system may be decidable, whereas what the P-argument presupposes is not simply that the Gödel sentence is provable in <u>some</u> sense, but that it is provable in the exact sense which guarantees that LM is inconsistent.

In summary, Chihara's first consideration is primarily about the truth of G_{LM}; his second consideration purports to show that LM may not qualify as a system which is subject to G1; Napoli's observation yields an analysis of the meaning of the provability in LM of G_{LM}, but only in regards to recursiveness and non-recursiveness; and finally, Butrick's viewpoint concerns the meaning of G_{LM} directly. As the above discussion shows, none of these criticisms may be justified as a satisfactory refutation of the P-argument. It is worth noting that though these critical points are diverse in themselves, they all leave untouched the question of system identification: In which system is G_{LM} provable in an ordinary sense? We will answer this question in section 3.2.

Now, we turn to some of criticisms of the L-argument.

2.4. Consistency, and Putnam's Criticism

Putnam's criticism of antimechanism concerning consistency and incompleteness has been widely appreciated.[25] According to this criticism, we can not simply say, merely by appeal to Gödel incompleteness, that there is a Gödel sentence G_T for a Turing machine T which is unprovable in T but is provable in LM. The reason is that

> 'Given an arbitrary machine T, all I (LM) can do is find a proposition U such that I can prove:
>
> (3) If T is consistent, U is true,
>
> where U is undecidable by T if T is in fact consistent. However, T can perfectly well prove (3) too! And the statement U, which T cannot prove (assuming consistency), I cannot prove either (unless I can prove that T is consistent, which is unlikely if T is very complicated)!'([42], Putnam, 1960, p. 77)

The upshot of this reason is that there is no difference between LM and a proper Turing machine T at least as regard Gödel incompleteness. This is why Putnam's criticism is also considered having force against the L-argument, though the former appeared prior to the latter.

However, since the L-argument takes the thesis of consistency as one of its antecedents, we can moderate it by adding to it some simple supplementary argument so that

---

[25] For the criticism, see [42], Putnam, 1960, p. 77. This criticism is made in a comment on Nagel and Newman. For its citations and comments, see [10], Chihara, 1972, p. 507-508, [8], Boyer, 1983, p. 148); and [48], Webb, p. 329). Even earlier, Good expressed largely the same point of criticism.([20], Good, 1969, p. 357).

it becomes immune from Putnam's criticism. The required supplementary argument, which is nicely presented by Lucas himself, goes like this: T is either consistent or inconsistent. If T is inconsistent, T cannot be equivalent to LM because LM is consistent according to the thesis of consistency. If T is consistent, then U is unprovable in T according to G1, but it is provable in LM. Again T cannot be equivalent to LM.([29], Lucas, 1968, p. 153-154) This supplementary argument shows that since the thesis of consistency is presupposed, for the validity of the moderated L-argument it does not really matter whether the consistency of T in question is eventually provable or not. In other words, the moderated L-argument could still pass intact even though the consistency of T for a class of Ts would never be provable. Or at least it would pass unless and until further proper criticisms of it were made.

To make the above point clearer, we can think of it in the contrapositive. If the thesis of consistency is not presupposed, then T and LM might be equivalent to each other precisely because they were both inconsistent and therefore U would be provable in both of them. So, the thesis of consistency really matters in this regard.[26]

---

[26] The thesis of consistency (for LM) and the actual provability of the consistency of T are different things. Lucas' moderated argument is that since the thesis of

Besides, needless to say, few mechanists except those like Priest would like to defend themselves at the sacrifice of giving up consistency.

The above discussion shows that in order to find the real pitfall of the L-argument, we have to fully appreciate the logical implications of the thesis of consistency as an antecedent of the L-argument.

Webb observes that Lucas' concept of consistency is not a simple and standard one. It is characterized as 'halting, tentative and self correcting'. Webb therefore doubts whether or not it is formalizable and provable.- ([48], Webb, 1983, p. 329) Webb's observation here is correct, but it does not serve as a case against the L-argument. For Lucas is an anti-mechanist, he need not care whether the logic of the mind or any of its aspects is formalizable or not. He might well intend the concept of consistency invoked in the L-argument not to be formalizable. As for the contents of consistency, since Lucas gives diverse arguments for the thesis of consistency, he makes the defended consistency open to certain qualifications which are not part of consistency as usually understood. However, in our view, what Lucas

---

consistency is presupposed, T must be different from LM, whether or not T is consistent. Conversely, if the thesis of consistency is not presupposed, T may be equivalent to LM.

really intends to show by arguing for the consistency of real minds is beyond the real. That is, theoretically, the existence of real minds, which are consistent only in some self correcting sense, implies that there is an ideal mind which is consistent in the usual sense, with all the capacities that the real ones may possess, save, of course, those merely attributable to inconsistency. To our knowledge, there is so far neither empirical evidence nor theoretical argument nor any combination thereof that has succeeded in disproving the existence of such an ideal mind. Besides, with regard to the L-argument, it should be noticed that the thesis of consistency is presupposed in the L-argument, and is not to be proved by the L-argument. So, even if the logic of the mind be consistent only in a 'self correcting' sense, it would be already enough for the logic of the mind to be distinct from all inconsistent formal theories. This means, for one thing, that the validity of the L-argument might not be seriously affected by what Webb notices concerning consistency in this context.

2.5. Turing Machine Specification, and Dennett's Reasoning

In Daniel C. Dennett's view, the multiplicity of Turing machine specifications of a physical object shows where and why the L-argument is wrong. He says:

> 'The fundamental error behind attempts to apply Gödel's Theorems to philosophy of mind is

> supposing that objective and exclusive determinations of the activities and capacities of concrete objects are possible which would determine uniquely which Turing machine specification (if any) is <u>the</u> specification for the object.'([14], Dennett, 1978, p. 264)

His reasoning can be summarized as the following: A Turing machine is in itself only a symbolic abstraction. Any concrete physical object is at the same time capable of receiving different Turing machine specifications, i.e., realizing different Turing machines, corresponding to different events which we wish to interpret as input/output symbol tokens rather than 'noises', and different physical states which we wish to interpret as logical machine states. Especially,

> 'it ought to be possible to interpret any man as any Turing machine -- indeed as all Turing machines at the same time.'(ibid, p. 262)

At the same time, there are no absolute or objective criteria by which to determine for an arbitrary object which Turing machine specification of it is the unique one. Even though an object is designed with the intention to realize a certain Turing machine, that intention does not provide legitimate grounds for us to say that that intendedly designed Turing machine is the unique correct Turing machine specification for that object. Because of this a person can in principle prove the Gödel sentence for any Turing machine (which is equivalent to a proper formal system) that he realizes under one interpretation in virtue of his realizing another Turing machine under another interpretation. Thus viewed, then, what can be done by a person can be so done by him merely in virtue of his capacity of Turing machine realization. That is, a person need not be beyond the mechanistic to be a person.

The above is Dennett's reasoning. This reasoning may well be correct in its own right, with some provisions, but it can hardly be right for the purpose of refuting the L-argument. There are three reasons for this. First, the L-argument is, strictly speaking, directed only to the relation between the logic of the mind and any _one_ abstract Turing machine, not to the relation between a concrete person and a concrete object, both as being simultaneous realizations of the many, and in fact infinitely many, Turing machines as Dennett conceives. Presumably, the equivalence between a person and a non-living object with respect to the capacity of Turing machine realization does not imply by itself the e-quivalence between the logic of the mind which is realizable by a person and any _one_ Turing machine. It is the latter equivalence that is primarily under the attack of the anti-mechanist. What Dennett does is to defend the latter equivalence by arguing for the first equivalence. That strategy does not work well. Perhaps Dennett thinks that that strategy must work. But at least this is just what he has to prove before he can justify his reasoning. If Dennett wants to do so, he should give further arguments, but he has not done so yet. Second, when Lucas claims that a person is able to prove the Gödel sentence for every gödelian system, he obviously means that a person is able to do it consciously. If Dennett's reasoning is correct and a person is able to prove the Gödel sentence for every gödelian system just because he can realize all the required Turing machines physically , it would require that a man is able to switch among infinitely many different Turing machine schemata consciously. However, empirically, there has been so far no hard evidence as to whether or not a person can do this. To the contrary, it seems natural to assume that because of certain physiological constraints, any concrete person cannot do this. Besides, by definition, the function of any Turing machine can be performed by a universal Turing machine. In this sense, a universal Turing machine can do any job which is recursive. So a person need only realize <u>one</u> universal Turing machine to do whatever he is supposed to be able to do, provided that mechanism holds. For, by definition, the function of any Turing machine can be performed by a universal Turing machine. Third, even it were possible for a person to realize infinitely many Turing machines consciously, it would still be very hard for the advocate of Dennett's reasoning to explain why a person can get to know that the Gödel sentences for all universal Turing machines are true, given the fact that no single Turing machine as he understands can prove the Gödel sentences for all universal Turing machines simultaneously.[27]

2.6. Finite Mechanism, and Kirk's Approach

Recently, Robert Kirk takes a somewhat peculiar approach in attacking the L-argument and its like. Kirk demonstrates his approach by giving two sub-arguments. The first sub-argument, which may be viewed as a complex of considerations, leads to a special version of mechanism:

> **Finite mechanism.** The logic of the mind can be adequately modelled by a finite automaton.([25], Kirk, 1986, p. 445)[28]

Here, a finite automaton is understood as a device having 'finitely many possible (internal) states, finitely many possible inputs and finitely many possible outputs; and for any given state and any input its output and next

---

27 For this point, see Turing, 1964, in [1], Anderson (ed), 1964, p. 16) and [48], Webb, 1983, p. 339.

28 Nelson subscribes to another version of finite mechanism in his book The Logic of Mind where he argues that human beings are nondeterministic finite automata. ([37], Nelson, 1982, p. 4 and p. 35)

following state will be determined.'[29] In contrast to Kirk's finite mechanism, our version of mechanism as we have so far formulated and talked about can be reformulated as

> **Standard mechanism (or Turing mechanism)** The logic of the mind can be adequately modelled by an infinite automaton.

Here, an infinite automaton, e.g., a Turing machine, is equipped with finitely many possible internal states but it may be equipped with unlimitedly or potentially infinitely many inputs and outputs.[30] Clearly, the difference between finite mechanism and standard mechanism lies in that what is capable of being modelled by an infinite automaton may not be so by a finite one.

The second sub-argument is a direct refutation of the L-argument given the truth of finite mechanism. It can be compressed as the following: No finite automaton is capable of all elementary arithmetic, or in our terminology, the thesis of recursive capacity does not hold for any finite automaton. In order to apply G1 to LM in the antimechanistic way, LM has to satisfy two conditions: first, being formalizable, i.e. being recursive; and

---

[29] The definition of a finite automaton can be found in [26], Kobrinskii, 1965, p. 3, or [2], Arbib, 1987, p. 57.

[30] For an exact definition, see [35], Nelson, 1968, p. 62, and p. 99), and [16], George, 1973, pp. 6-7, 69-70.

second, being capable of all elementary arithmetic. Under Kirk's approach, LM satisfies the first condition, but since finite mechanism holds, LM does not satisfy the second condition in the manner in which it satisfies the first one. So the conclusion: G1 cannot be applied to LM in the anti-mechanist way as it is intended to be by Lucas.(ibid, pp. 444-445)

We can see that either Kirk's second argument is invalid, or it presupposes a stronger version of finite mechanism, i.e.:

> **Strong finite mechanism.** LM can be adequately modelled by a finite automaton, but it cannot be adequately modelled by any automaton which is not equivalent to a finite one.

Clearly, if LM can also be adequately modelled by an infinite automaton, Kirk's second argument will lose all its force due to this point.[31]

Before checking the truth of strong finite mechanism,

---

[31] We say so because there is so far no proof that what is capable of being adequately modelled by a finite automaton cannot be so by an infinite one. Such a proof will depend partly upon what 'adequately modelling' means. Anyway, the main point of Kirk's argument will not be changed if there is such a proof. Strong finite mechanism appears to be a silly position. But, as reasoned in the text, Kirk's approach cannot hold without it. So we guess that the concept of adequately modelling in Kirk's mind is very special. It may have the connotation of modelling as much as possible by means as little as possible.

we first examine what Kirk's second argument would be about if strong finite mechanism held. To be sure, Kirk's second argument as a conditional does hold regardless of whether strong finite mechanism does as well. Indeed, if strong finite mechanism held, then so would the conclusion of the argument. However, strong finite mechanism does not hold. For when we say that LM can be completely or adequately modelled by such and such means, we straightforwardly mean that LM _with its full capacity_ can be so modelled by such means. Since we know that any finite automaton does not have the full capacity of representing all means of recursion as LM has, any finite automaton cannot model LM with its full capacity adequately. If Kirk thought that to model LM adequately did not mean to model LM with its full capacity, he could establish strong finite mechanism. But this kind of mechanism would not vitiate Lucas' anti-mechanist position any more. For what Lucas insists is just that LM with its full capacity cannot be modelled by any machine. Thus viewed, Kirk's approach is hardly acceptable to the standard mechanist.

As we suggested above, Kirk in fact gives no decent argument for finite mechanism. In Kirk's view, grantedly, the mechanist 'will probably concede' that there are only finite number of possible 'discriminably different' inputs, outputs and internal states for the human mind.(ibid, p. 441 and p. 444) But this view needs much clarifying. For this we consider the case of input only.[32] There are two senses in which it is meaningful to talk about the number of the inputs of a machine or a mind: the actual and the potential. Two points are clear in this regard. First, the number of the inputs of a mind is actually finite in the sense that the number of the discernable inputs that the mind actually receives in any limited time is finite. Second, the number of the inputs of a mind is potentially infinite (or unlimited) in the sense that for any ordinary mind there is no previously given finite set of inputs which can cover all the inputs that the mind may receive in its undefinitely long potential life. Kirk accepts the second point as well as the first one. For he admits that the human mind is capable of all elementary arithmetic 'by interacting with the outside world'.(ibid, pp. 442-443) Now the question is, to let an automaton model the human brain as adequately as possible, should it be a) both actually and potentially finite, or rather, b) only actually finite but

---

[32] The case of output is parallel to the case of input; and the case of internal states can be omitted for there is no difference between a Turing machine and a finite machine with respect to internal state per se. Like a finite automaton, a Turing machine has only finite internal states; whereas unlike any finite automaton, a Turing machine may have infinite inputs and outputs.

at the same time potentially infinite? Certainly, the answer must be b). The next step is obvious: it is only a proper Turing machine, not a finite automaton, that is capable of b); and hence a finite automaton cannot best model the human mind. From this we conclude that the best explanation of Kirk's approach is, though Kirk could have not been very clear about this as he should be, that he considered the human mind only to be something which is isolated from, or unaided by, the outside world. Assuming that the human mind is isolated from the outside world, then of course Kirk approach is right. But since the human mind addressed in the L-argument is legitimately supposed to be interacting with the outside world, Kirk's approach is just beside the point with respect to the L-argument.

2.7. Representability, Benacerraf's Idea and His Argument

In an attempt to clarify the provability concept invoked in the L-argument, Paul Benacerraf construes it so that

> 'to prove a formula is to derive it as a formal theorem of a consistent system which includes the postulates of arithmetic.'([6], Benacerraf, 1967, p. 20)

Based on this construal, and noticing that every formal system equivalent to a Turing machine can be enumerated by a universal Turing machine, say Maud, under some stipulation about vocabulary and Gödel numbering, he reasons as follows:

> 'Identifying the i-th formal system in the enumeration as W_i, it follows from Gödel's work that the relation '$F_1, \ldots, F_n$ is a proof of F_n in W_i' has its counterpart in Maud: i.e. whenever a statement of that form is true, its translation into Maud is provable by Maud. Calling '$Maud_1$' the result of adding H (the Gödel sentence for Maud) as an axiom to Maud, then 'H is a theorem of $Maud_1$' is provable by Maud. So, thus interpreted, what Lucas can 'prove' does not differentiate him from Maud.(ibid, pp. 20-21)

Here, Benacerraf expresses an important idea:

> **The B-idea.** Some proof in LM of G_{LM} can be interpreted as a proof in a consistent formal theory which is not necessarily the same as LM but which can in a sense be represented in LM.[33]

We believe that the B-idea, if fully developed and thoroughly applied, can be used to refute both the L- and the P-argument in a way which is most coherent with, and beneficial to, the background of contemporary logic and philosophy of mind.

Nonetheless, there are two shortcomings with Benacerraf's presentation of the B-idea. First, he is not specific as to how to determine the Gödel sentence for Maud if Maud is a universal Turing machine. But this must be made specific to meet Benacerraf's requirements. We will expand this point in chapter 4 but omit it here.

---

[33] According to J.J.C. Smart, W.V. Quine had presented to him roughly the same as the B-idea before 1961.(See [43], Smart, 1961, pp. 109-110)

Second, and more importantly, he does not seem to appreciate his own B-idea sufficiently. The emphasis of the paper in which he presents this idea is given to his preferred reconstruction of Lucas' original argument; let us call it the B-argument. The B-argument presupposes quite the opposite of the B-idea, namely:

> **The containment presupposition** Any sentence which is provable in LM is contained in the logical closure of LM.(ibid, pp. 23-30)[34]

Obviously, since according to the B-idea, a Turing machine can prove certain sentences by virtue of representing some other Turing machines, and hence without necessarily containing those sentences (i.e. proving them directly), the B-idea is incompatible with the containment presupposition. Benacerraf himself does not recognize such an implicit presupposition in his argument. But the evidence for this is very clear, once noticed. For example, throughout his twenty step argument, he defines S^*, the closure of all what 'I' (roughly, our LM) can prove as the set which contains all provable sentences; whereas if he took the B-idea seriously, he would not define S* that way.

The conflict between the B-idea and the B-argument

---

[34] This presupposition is similar to the C-principle to be formulated in section 3.4.

shows at least that Benacerraf does not take the B-idea very seriously. This is not surprising. His B-idea has been almost totally overlooked in the literature. There are in the literature comments and criticisms, from different perspectives, of Benacerraf's paper and the B-argument, some of them being instructive or correct in themselves. But none of them, it seems to us, have been made with regard to the B-idea. The authors of these comments and criticisms, Lucas included, simply do not pay attention to the B-idea, apparently thinking that whether or not Benacerraf is correct with his B-argument has nothing to do with the B-idea.[35] Our view and appreciation are rather different. Because of this, we feel a need to review the B-argument in a way that highlights the B-idea directly.

In Benacerraf's own view, the only assumption of the B-argument is, at step 9,

> **The tripartite assumption.** '[t]here is a recursively enumerable set W_j such that
> a) '$Q \subseteq W_j$'ϵS^*
> b) '$W_j \subseteq S^*$'ϵS^*
> c) $S^* \subseteq W_j$

---

[35] See [29], Lucas, 1968; and [11], Chihara, 1972. Further, there is no allusion to the B-idea in Boyer's review of some problems with Lucas which have attracted attention; ([8], Boyer, 1983, p. 148) neither is there specific mention of Benacerraf and his B-idea in Kirk's recent list of some of the most popular suggestions for mechanism in view of Gödel incompleteness.([25], Kirk, 1986, p. 451)

where Q is Robinson's arithmetic, and S^* is the same as introduced above (i.e.,the logical closure of the theorems that 'I' can prove).(ibid, p. 25) The B-argument ends with a contradiction at step 20. So, the conclusion of the B-argument, step 21 of the B-argument, is, instead of the negation of (9c) as Lucas suggests, the negation of the tripartite assumption. That is,

> 'At best Gödel's theorems imply the negation of the conjunction of 9a, 9b, and 9c.'(ibid, p. 29)

To Benacerraf, the B-argument

> 'fairly represents what underlies the vague ones that Lucas presents...';(ibid, p. 23)

and it is

> 'the best case that one can make for a view such as Lucas's and extract from that what would seem to be the import of the Gödel theorems for the philosophical thesis of Mechanism.'(ibid, p. 17)

We will make four points concerning the B-argument, but will not go through all the details of the argument. First, although perhaps the tripartite assumption is the only <u>explicit assumption</u> of the B-argument, there are implicit ones, among which the containment presupposition is prominent. Because of this, if the containment presupposition is unacceptable, the B-argument may not be valid. Second, if the containment presupposition were acceptable, the L-argument would be able to survive in spite of its criticisms reviewed in this thesis.[36] Were this the case, the L-argument instead of the B-argument would probably still be the best case for Lucas. Third, as we will show in chapter 4, the containment presupposition also underlies the L-argument. In spite of this similarity, the B-argument and the L-argument are qualitatively different, for they lead to different conclusions. Benacerraf gives no specific reason to show why and in which sense the B-argument 'fairly represents what underlies the vague ones that Lucas presents'. Especially, we do not see that there is a proper sense in which the B-argument more fairly represent Lucas' original argument than the L-argument does. So we think that the L-argument will still need to be refuted whether or not the B-argument is valid. Fourth, Benacerraf takes the negation of the tripartite assumption to mean further that

> 'If I am a Turing machine not only can I not ascertain which one, but neither can I ascertain of any instantiation of the machine that I happen to be that it is an instantiation of that machine.' (ibid, p. 29)

Here we do not argue whether or in which sense the B-argument really implies the above meaning. Instead we want to point out that even if the B-argument were correct and really implied the above meaning, it would say nothing as

---

[36] For this point, please see our discussion of the common pitfall of the L- and the P-arguments in chapters 3 and 4 for detail.

to whether or not another human being other than '*I*' could ascertain which Turing machine is equivalent to '*I*'. Understandably, if the B-argument entails nothing significant to the latter question, its significance will be much less than otherwise expected. Besides, when '*I*' is said to be able or unable to ascertain himself as a specific Turing machine, it seems to be presupposed that '*I*' already possesses an adequate concept of his self identity. It seems no easy job to deal with problems related to this presupposition.

2.8. Effectiveness, and Myhill's Theorem

For the mechanist, B-idea needs to be justified in a way in which effectiveness is essentially involved. For, to refute the L-argument, the mechanist needs to show not only that what can be done in a way by a brain can be done in a way by a machine, but also that the two ways in which the job is done are both effective. Now, to refute the L-argument, we need not show directly that the way in which any job is done by a brain has to be effective; we need only show that the job of overcoming Gödel incompleteness can be done in a effective way by a machine. For if we can show this, then it is at least possible that the way in which LM is able to overcome Gödel incompleteness is effective. As Webb observes

> 'the real source of Lucas' feeling of superiority here is the very effectiveness with which

> Gödel's diagonal argument will enable him always to find an achilles' heel in each machine... if Lucas can effectively stump every machine, then by (CT) (i.e. Church's thesis), there should be a machine that does this too!'([48], Webb, 1983, p. 339)

In this regard, a relevant and significant result is

> **Myhill's theorem** There is a total recursive function g such that for the Turing machine $Z_{h(\Sigma)}$ that prints out theorems of the adequate consistent arithmetic Σ, the Turing machine $Z_{g(h(\Sigma))}$ is $Z_{h(\Sigma')}$ for an adequate consistent arithmetical logic Σ' with more theorems than Σ.([33], Myhill, p. 115; also see [2], Arbib, 1987, pp. 167-168)

As Arbib comments,

> 'while Gödel's incompleteness theorem points to an inevitable limitation of any axiomatization of arithmetic, Myhill's theorem points out the much less well known fact that this limitation can be effectively overcome.'(ibid, Arbib, p. 168)

Thus understood, what Myhill's theorem shows is precisely

> **The Point of Effectiveness.** The procedure of finding out an appropriate machine which is able to overcome the Gödel incompleteness of another arbitrarily chosen machine is effective.

The point of effectiveness implies that there is some sense in which it can be claimed not only that for each machine M_x there is another M_y which can overcome the

Gödel incompleteness of M_x, but also that there is one machine M which can, of course, effectively, overcome the Gödel incompleteness of every appropriate machine. So in this sense what can be done by the human mind with respect to Gödel incompleteness can be done by a Turing machine.[37]

The point of effectiveness shown by Myhill's theorem, and Webb' and Arbib's comments is very important. It suggests that the limitation of any effective means as is shown by Gödel incompleteness can be itself overcome in a effective way. However, it still leaves something to be desired for a full refutation of the L-argument with the thesis of consistency retained. There are two points in need of further explanation. First, Myhill's theorem shows only that the procedure of extending consistent arithmetic systems is effective. Suppose a proper Turing machine, say M, can perform this procedure and represent all produced systems, then, concerning Gödel incompleteness, we need to show what is the technical sense in which the Gödel sentence for M can be determined and proved in M itself, before we can say that M can overcome its own Gödel limitation in the manner that Arbib conceives. Second, G1

---

[37] Good also conceives this point, though in a vague sense.([19], Good, 1967, p. 145) It is worth noting that the point of effectiveness addresses the Turing machine and the human mind directly, but not physical devices capable of realizing Turing machines. Because of this, it is in sharp contrast to Dennett's reasoning (see section 2.5).

shows that the Gödel sentence for any gödelian system is unprovable in that system unless that system is inconsistent. Granting that a universal Turing machine is equivalent to a gödelian system, how can we show that a universal Turing machine can prove its own Gödel sentence without being inconsistent, i.e. without being equivalent to an inconsistent system? Certain explanations for these two points may be implicit in Myhill, Webb and Arbib's mind. But they have not expressed them explicitly yet. In chapter 4, we will give our own explanations of these two points, in light of the B-idea and the point of effectiveness.

In summary: Putnam's criticism is intended to show that a machine may not be different from the mind with respect to Gödel incompleteness because we may not be able to prove the Gödel sentence for that machine through the (only possible) way of proving its consistency. Dennett's reasoning appeals, instead to a universal Turing machine, to an object's 'universal' capacity of realizing arbitrary machines. Kirk's approach equates LM to a finite automaton, and thereby omits LM's unlimited potential. Benacerraf's B-idea has the provability in S of G_S interpreted as the provability in one system which is representable in S. Myhill's theorem helps show that the B-idea accords well with mechanism. In our view, the B-idea implies a representational approach toward refuting the L-argument; whereas Webb and Arbib's viewpoint, the point of effectiveness supported by Myhill's theorem, represents mechanism in its proper sense; moreover, if we want to have both the point of effectiveness and the thesis of consistency, the only way for us is to go along with a proper representational approach.

3. An Examination of the P-Argument

The critical examination of the P-argument presented in this chapter constitutes a self-contained refutation of the P-argument. To locate the pitfall for the P-argument, section 3.1 starts with a preliminary examination of the P-argument, which leads to the discovery that Priest tacitly identifies two provability concepts: one according to which G1 holds, and the other according to which G_1 is viewed in the usual sense as provable in LM. Call this identification the P-identification. Based on a discussion of the recursiveness of proof, section 3.2 shows that, the crux of the examination of the P-identification is not recursiveness in general, but the special identification of two systems of which one is LM, and the other is one of those in which G_{LM} is provable. Section 3.3 formulates Priest's underlying inference needed for the P-identification. Call it the P-inference. It is shown that the truth of the P-inference hinges primarily upon how we interpret its premise that any sentence which is provable in a sound system is provable in LM. Refer to the premise of the P-inference as I1. Regarding the meaning of I1, section 3.4 first formulates and then dismisses the containment principle of interpretation according to which

I1 is interpreted as stating that LM contains as its subset the axioms of S for any sound system S. Section 3.5 presents the representational principle of interpretation, according to which I1 is interpreted as holding that LM can represent the axioms of S for any system S. Correspondingly, concepts such as representational provability and its technical counterpart QGR-provability are introduced. The task of section 3.6 is to show how the entire P-inference is interpreted according to the representational principle and why this interpretation is justifiable. Section 3.7 contains the conclusion that since the representational provability concept is plausible and distinct from the ordinary provability concept, the P-identification is not the only option in question, and hence the whole P-argument is invalid.

3.1. A Preliminary examination and the P-identification

Upon a preliminary examination of the P-argument, we can be sure of the following points about it. First, B1 is just an informal version of G1; hence there should be no problem with B1. Second, B2, i.e., the thesis of recursive capacity, is highly plausible, or in Priest's words, 'relatively unproblematic.'[38] Although B2 may not be true

---

[38] See ([40], Priest, 1984, p. 167), where he explains, 'For our naive canons of proof (i.e. our LM) contain those of ordinary arithmetic, in which all recursive functions are specifiable in the usual way.'

for some real minds, it is true for certain ideal minds. In fact, our interest and also Priest's lie just in whether the thesis of inconsistency is true for these minds. So we can pass B2 without making a difference to our and Priest's interest. Third, B3, i.e., mechanism, is certainly open to examination. However, in the P-argument, B3 serves only as an assumption. In examining the P-argument we are interested only in whether this argument holds <u>given</u> B3, not in whether B3 itself holds or not. Since our intention is to defend the joint compatibility of mechanism, the thesis of consistency, G1, and the thesis of recursive capacity, even though the arguments available for mechanism are not sufficient for us to confirm it, it is still useful to show that mechanism is jointly compatible with the thesis of consistency and G1. Fourth, B4 follows from B2 and B3 regardless of whether B2 and B3 are true. So it is valid as a conditional. Fifth, B5 holds as well for it is just an application of G1 in case the gödelian system is LM. Sixth, B7 as a sub-argument is also valid, for what B7 says is only that the conclusion of B7 follows from that of B5 and B6 regardless of whether one or both of the two themselves hold or not. Thus, finally, seventh, the above points suggest that if the P-argument is indeed invalid, the pitfall is most probably in B6.

We now try to locate the pitfall of the P-argument by examining B6. To begin with, we express our agreement with Priest that there is a definite sense in which B6 holds, i.e. G_{LM} is provable in LM. To help make it clearer in which sense B6 holds and what are the bearings of this sense on his whole argument we reformulate G1 by the following conditional:

G1'. For any system S, if

(a). S is r.e.,

(b). S satisfies the recursive capacity assumption, and

(c). S is consistent,

then

(d). G_S is not provable in S.

Put in terms of G1', the P-argument goes like this: In the case where S = LM, we have

(e). G_{LM} is provable in LM.

So the negation of (d) holds. Since (b) holds as well, a choice has to be made between (a) and (c). Since (a) is indispensible, (c) has to be rejected. The crucial underlying question is: in which sense can the provability in LM of G_{LM} be demonstrated? To be precise, we let (d) and (e) above be re-expressed as

(d'). G_{LM} is not provable in LM according to some provability concept.

and

(e'). G_{LM} is provable in LM according to some provability concept.

Let us call the provability concept imposed upon (d') **provability$_1$** and that imposed upon (e') **provability$_2$**. At least nominally, the two provabilities can be regarded as being different. The question is, are they indeed essentially different? To Priest, they **must be** essentially the same. For he thinks that the provability of G_{LM} implies the inconsistency of LM according to G1, and that the provability of G_{LM} and that of the unprovability of G_{LM} (provided that LM is consistent) constitute a contradiction.([40], Priest, 1984, p. 165) Looking at the P-argument we see that only by assuming the identity of the two provabilities can he reach his desired conclusion. Let us call Priest's assumed identification of the two provabilities **the P-identification**. Obviously, if we can show that the P-identification is not true, then we can choose to keep both (a) and (c) as mentioned above without necessarily causing inconsistency. We will show in the following that the P-identification is indeed a misidentification.

3.2. Recursiveness in General and the Special System Identification

We need a long argument to show why the P-identification is misleading. At the first sight, the crux of the matter of P-identification appears to be recursiveness. In fact, without specifying everything that provability$_1$ connotes, we can safely say that it carries recursiveness as its primary connotation. Definitely, when we say that G_{LM} is not provable in LM, we mean that G_{LM} is not recursively provable in LM. Thus, the truth of the P-identification seems to be essentially contingent upon whether provability$_2$ connotes recursiveness as well. Priest entertains such a view. To him, the P-identification is justifiable simply because G_{LM} is recursively provable.[39] Prompted by this view, Priest gives a series of reasons for a more general thesis:

> **The thesis of the recursiveness of proof.** Any proof is recursive.

Priest's reasons for this thesis are elegant and forceful.[40] We agree with Priest completely on this thesis

---

[39] In fact, many philosophers (including Napoli and Lucas) think that this is the most significant question regarding provability$_2$. Napoli suggests that Priest is wrong just because, among other things, G_M is not recursively provable. (See section 2.2) Lucas in fact agrees with Napoli on this point, and it forms his pivotal point against mechanism.

[40] See [40], Priest, 1984, pp. 167-170; and [41], Priest, 1987, pp. 51-55. The term 'the thesis of the recursiveness of proof' appears in ibid, p. 54. The origin of this thesis can be traced back to at least to Hilbert's viewpoint that proofs are finite objects. See [23], Hilbert, 1927; also [47], Webb, 1980, p. 122, p. 152, p. 175, p. 186).

and generally on his reasons for this thesis. For the sake of this paper then, we omit mentioning the details of his reasons except for quoting two telling passages, from A. Church and Priest respectively:

> '... consider the situation which arises if the notion of proof is non-effective. There is then no certain means by which, when a sequence of formulas has been put forward as a proof, the auditor may determine whether it is in fact a proof. Therefore he may fairly demand a proof, in any given case, that the sequence of formulas put forward is a proof; and until this supplementary proof is provided, he may refuse to be convinced that the alleged theorem is proved. This supplementary proof ought to be regarded, it seems, as part of the whole proof of the theorem...' ([12], Church, 1956, p. 53.)

> 'If the proof relation is effectively recognizable, then by Church's thesis (which seems entirely reasonable in the context), it is recursive.'([40], Priest, 1984, p. 167)

From a theoretical point of view, of course, there are non- recursively enumerable systems, (non-r.e. systems), as well as r.e. systems. For a r.e. system, plainly, any concrete proof in it is recursive. For a non-r.e. system, the set of its axioms are not r.e., so the entire proof relation for it cannot be r.e. However, this does not mean that none of its axioms is discernable or identifiable. For example, second order arithmetic is non-r.e., but we certainly can identify some of its axioms. Precisely because of this, we can still say that any concrete proof, even though it is conducted in a non-r.e. system, is recursive in spite of the fact that the entire provability relation for the non-r.e. system is not r.e.. Besides, naturally, any proof which is conductible in a non-r.e. system should be so in an appropriate r.e. system.

From the above we can see that if the crux of the examination of the P-identification were merely recursiveness, the P-identification would indeed be justifiable by the thesis of the recursiveness of proof. However, the story is not so simple. Recursiveness is at most one perspective from which the P-identification should be examined; there is another perspective: system identity. The P-identification holds if and only if LM is identical to one of the systems in which G_{LM} is provable.[41] As Benacerraf notices, whatever sentence is provable is provable necessarily in a system.(See the beginning of section 2.7 Although the thesis of the recursiveness of proof guarantees that G_{LM} is recursively provable if it is indeed provable, it does not and can not decree or determine the system or systems in which it is provable. For this reason we would rather say that the real crux of the P-identification is not recursiveness in general, but the special identification of two systems of which one is

---

[41] We need not assume that the system in which G_{LM} is provable is unique. If G_{LM} could be provable in a system which LM properly includes, then G_{LM} could be so in LM as well.

LM, and the other is one of those in which G_{LM} is provable.

3.3. The P-inference

We now need to return to G1. From G1 and the 'gödelian suite' of its related results[42], the Gödel sentence for any consistent gödelian system can be proved in another consistent gödelian system, but in no consistent gödelian system can all the Gödel sentences for all the gödelian systems be provable, and especially, in no consistent gödelian system can the Gödel sentence for that system itself be provable.

Because of the above, there are two alternatives: either

(1) The P-identification is true and in this case LM is inconsistent;

or

(2) The P-identification is false and in this case there is a system S such that G_{LM} is provable in S but S is not included in LM and hence may be consistent.

Presumably, with the concerns shared by Priest and us, but without any prejudice against either alternative, we can only claim, without specifying what system S is:

---

42 I use the phrase 'the gödelian suite' after Webb. See [48], Webb, 1983, p. 309, 'Gödel's incompleteness theorems and their suite'.

C1. G_{LM} is provable in a system S.

C1 is merely compatible both with (1) and with (2). But Priest must go further for the sake of his whole argument, and say that S in C1 could be LM , for he must claim

C2. G_{LM} is provable in LM according to the same provability according to which G1 holds for LM.

Presumably, C1 is a necessary but not sufficient condition for C2. But Priest thinks that it is legitimate to infer C2 from C1. Let us call the inference from C1 to C2 the P-inference. How could Priest make the P-inference? Priest has said nothing explicitly regarding this except suggesting the vague and ambiguous point that whatever is provable in a sound system is provable in LM.([40], Priest, 1984, p. 165) He seems to think that he need not to say anything specific about the P-inference. However, it may well be illuminating to explicate the P-inference. To us, the most probable explication of the P-inference is like this:

The P-inference

I1. Any sentence which is provable in a sound system is provable in LM.[43]

---

[43] Drawn from Priest's context, a system is sound if there is an intended model in which the appropriate axioms are true.([38], Priest, 1984, p. 166, where he talks about the soundness of Peano Arithmetic) Priest is not specific, it seems to us, as to whether there are sound systems which are, like Euclidean geometry and non-Euclidean geometries, mutually inconsistent. It would be

I2. There is a sound system in which G_{LM} is provable.

I3. From the above G_{LM} is provable in LM.

We now examine the P-inference as explicated above. There should be no problem with I3 as a conditional. Whether I2(i.e., C1) is true is still subject to examination from a skeptical point of view. But we do not want to be such skeptics as to be bothered with I2. So in our view, the truth of the P-inference hinges on I1. However, it does seem natural to view I1 as a meta-axiom about our LM, and we do think, or at least prefer to stipulate, maybe with some provisions, that I1 is true. Thus what may be left open and nonetheless hinges on the truth and the meaning of the P-inference is the interpretation of I1.

3.4. O-provability and The Containment Principle of Interpretation

To interpret I1 correctly, we need a correct principle. There are at least two rival principles for our present concerns. We consider the first one in this section. To begin with, we make some elementary concepts

---

unfair to Priest to think that he would deny the existence of the intended models for non-Euclidean geometries. But if Priest did admit the existence of such models, and hence to admit the corresponding systems as being sound, we seem to have one more point against his view. (See later the discussion about the second reason against $MI1_C$.)

or terms sufficiently explicit. First, by definition, a sentence p is provable in a system S if and only if there is a finite series of sentences $p_1, \ldots, p_k$ such that p is p_k, and for any $i \leq k$ p_i is either an axiom of S or deducible from the preceding ps according to the rules of S. Second, ordinarily, when a sentence p is said to be an axiom of a system S, p is meant primarily to be _included_ or _contained_ in S. Finally, we introduce the concept

> **Ordinary provability (or O-provability).** A sentence is O-provable in S if and only if there is one proof of p such that all the axioms required for that proof are contained in S.

The above definition of O-provability seems superfluous. For ordinarily, the provability one addresses is just O-provability. However, since there may be meaningful non-ordinary kinds of provabilities, it is helpful to have O-provability defined explicitly. Later, another kind of provability will indeed be introduced. At any rate, a little bit of superfluity does not create much trouble. In the following, O-provability and provability will be meant to be the same, but in case we feel a need to emphasize the distinction of the ordinary provability concept from the representational provability concept (see next section), we will use the term 'O-provability' or its cognates instead of the term 'provability' or its cognates.

We are now ready to formulate the first principle.

The containment principle of interpretation (or the C-principle). The general provability concept for LM should be understood such that a sentence is provable in LM if and only if it is O-provable in LM.

According to this principle, the meaning of I1 is

$MI1_C$. Any sentence which is O-provable in a sound system is O-provable in LM.[44]

In other words, the logical closure of LM contains as a subset of the closure the set of the axioms of S for every sound system S.[45]

---

[44] We use 'sound' here in $MI1_C$ (and also in $MI1_R$ and so on, see later) because it appears in I1, the first premise of the P-inference. We would rather like to use 'acceptable' instead of 'sound', though. The meaning of 'being acceptable' is now vague and flexible, and it may just mean being sound or something else. However, our point is that we need not specify the meaning of 'being acceptable' in order to carry our further discussion. Please see sections 3.5 and 3.6 for relevant points.

[45] The logical closure of a system is the set of all the consequences of the system. There seems no hard evidence as to whether Priest accepts $MI1_C$. However, since only according to $MI1_C$ can the provability in LM of G_M and the unprovability in LM of G_{LM} constitute a contradiction, and Priest, holding his P-identification, does think this contradiction has been derived wherefrom, he virtually accepts $MI1_C$. A most likely possibility is that Priest is inadvertent about $MI1_C$ in contrast to other possible ones. However, since when one says a sentence is provable in a system, he ordinarily means that that sentence is O-provable in that system, Priest should be ordinarily understood as meaning $MI1_C$ when he holds I1 without

We have two reasons against $MI1_C$ and hence against the C-principle. The first reason comes in part from the existence, or rather, the recognition of the existence, of non-r.e. systems. Concerning $MI1_C$ we have the following conditional:

CD. If

(a) $MI1_C$ is true,

(b) the theory of recursion is true, and

(c) there exists at least one non-r.e. system,

then

(d) mechanism is false.

Of this conditional CD, (a) is in question; (b) is certainly viewed as holding; and (c) is also commonly recognized as true, and probably held by Priest. Since mechanism is claimed to be true by both Priest and us, $MI1_C$ should be rejected according to both Priest and us. We are satisfied with this contraposition of the conditional CD, not going to go into any detail about the truth of (b) or that of (c).

Our second reason is found in the existence, or at least the recognition of the existence, of systems which are inconsistent with one another while each of which can,

---

indicating which special interpretation should be given to I1. Besides, when he talks about the meaning of proof, the arithmetic capacity of LM, he seems to presuppose $MI1_C$.(See [39], Priest, 1979, pp. 220-221; and [40], Priest, 1984, p. 165, and especially, p. 167) in virtue of the fact that they can be given intended models, be viewed as sound. For example, the continuum hypothesis, CH for short, has been proved to be independent of ZFC. This means, for one thing, that both T_a = ZFC U {CH} and T_b = ZFC U {-CH} are consistent if so is ZFC. Since there is no reason for us to claim that one is sound while the other is unsound, we should accept them both as sound. For another example, as we mentioned before, both Euclidean and non-Euclidean geometries have intended models, and hence are consistent by themselves, but they are mutually inconsistent. If $MI1_C$ were true, then LM would be a priori inconsistent in virtue of the fact that LM contained all the axioms of these mutually inconsistent systems. One might think that this consequence would not baffle Priest, for inconsistency is just normal to him. However, since every sentence except the negation of a tautology is self consistent, if $MI1_C$ is correct, then LM would contain all sentences except the negations of the tautologies, making every contradiction virtually provable in LM. This consequence is certainly unacceptable even by Priest, for he holds nonetheless the viewpoint that contradictions should be kept to a minimum.([41], Priest, 1987, p. 144-145)

Given the two above reasons, it should be obvious that $MI1_C$ is implausible.[46]

3.5. R-provability and The Representational Principle of Interpretation

In contrast to O-provability, we now define

Representational provability (or R-provability). A sentence p is R-provable in a system S if and only if p is O-provable in a system which is representable in S.

The representability concept in the above definition is itself unspecified. We will define it later. Now, using the concept of R-provability, we formulate

The representational principle of interpretation (or the R-principle). The general provability concept for LM should be understood such that a

---

[46] We anticipate that there are some who are still not convinced of the implausibility of $MI1_C$ even after being given the above reasons. However, for our main purpose at the very least, we need not to exclude $MI1_C$ as a possible interpretation of I1 in order to justify the compatibility of mechanism and the thesis of consistency. For this we need only to demonstrate that there is a plausible interpretation of I1 which leads to the justification of this compatibility. For if this is the case, we can show at least that inconsistency is not the only choice which we have to make under the apparent pressure of Gödel incompleteness. Besides, even whether $MI1_C$ works for Priest's purpose is still a question. For, if LM contains axioms which make G_{LM} provable in LM, how can it be possible that G_{LM} is not provable in LM at the same time? What G1 states, with reference to LM, is that if LM is gödelian and consistent, then G_{LM} is not provable in LM. Surely, we should have, if G_{LM} is provable in LM and LM is gödelian, then LM is inconsistent, then G_{LM} cannot be unprovable in LM.

sentence p is provable in LM if and only if p is O-provable in a system which is representable in, and acceptable for, LM.

There are two key concepts in the above formulation: (system) representability and (system) acceptability. We will concentrate on representability. But, before doing this, we try to explain the acceptability concept briefly first. For acceptability, we stipulate here roughly that

a system S_1 is **acceptable** for S_2 if S_1 has not been (corresponding to a weak sense), or cannot be (corresponding to a strong sense), found inconsistent, unsound or the like by virtue of S_2.

It is easy to see that there are other possible weaker or stronger or intermediate senses in which the acceptability concept can be defined along this general line. However, the essence of all these can be viewed as the same for our present concerns. We will discuss the acceptability concept in more detail in section 3.6.

Now we consider representability. Since there are many concepts of representability, there may be many senses in which the R-principle can be meaningful. In this thesis we just choose one. To show what our choice is we introduce two concepts: gödelian representability (or G-representability) and quasi-gödelian representability (or

QG-representability). For any two systems S_1 and S_2:

S_1 is said to be **G-representable** in S_2 if the syntax predicates of the language of S_1 and the provability predicate for S_1 are <u>representable</u> in S_2; and

S_1 is said to be **QG-representable** in S_2 if **for every r.e. aspect A** of the syntax or of the provability predicate of S_1, A is G-representable in S_2.

Here, a predicate P is said to be representable in a system S if and only if there is a formula f in the language of S such that

$$P(x) \equiv S \vdash f(x).$$ [47]

To be sure, the above two concepts are meaningful. For G-representability, from Gödel's results, we know that, first, the syntax of any r.e. language is G-representable in any system which is capable of

---

[47] This concept of the representability of a predicate is understood according to the definition given in Peter Aczel (1977), in [2], Barwise (ed), p. 754. This definition is equivalent to that of 'weak representability (of a predicate)' given in [3], Bell and Machover, pp. 324-325. In the definition of the G-representability, the condition of the representability of the syntax predicates of S_1 can be strengthened to that of strong representability of those predicates. Here, a predicate P is said to be strongly representable in a system S if and only if there is a formula f in the language of S such that

if P(x), then S ⊢ f(x); if -P(x), then S ⊢ -f(x).

(See ibid) Since we think that the technical difference between these two represenabilities is insignificant for our present concerns, we just pass it by.

representing all means of recursion; and second, the provability predicate for any r.e. system S is G-representable in any system satisfying the recursive capacity assumption.

For QG-representability, it is important to recognize that though we can talk about non-r.e. systems, and seem to be able to address these systems in the same way we do r.e. systems, all we can express are, supposing mechanism[48], but the r.e. aspects of such non-r.e. systems. To be convinced of this, let us imagine how a non-r.e. system can be described or defined. Let S be an arbitrary but intended non-r.e. system. By contemporary standards, the best and standard way to define S is to define or construct it in precise axiomatic-set-theoretic terms. In fact, the language of S, the axioms of S, the rules of S and so on all can and can only be adequately defined in these terms which are provided and conditioned by a proper axiomatic set theory, ZFC, for example. This means that all the necessary steps for an adequate definition of S can be taken in ZFC. In other words, all the dimensions of S can be defined in ZFC so long as they can be precisely described or defined, period. Nonetheless ZFC is itself a r.e. system. This very fact implies that

---

[48] Mechanism here serves to exclude the possibility that the logic of the mind is capable of non-recursive means.

although we can prove within ZFC that S is not a r.e. system, we cannot say anything beyond the r.e. aspects of S. This is so not because ZFC is short of something, but because ZFC is capable of providing the definition of any r.e. or non-r.e. apparatus whereas any definition given in ZFC has to be itself recursively given. In this sense we say generally that if for any r.e. aspect x of S x can be G-representable in a system, then S can be viewed as best representable in that system, nothing better could be achieved in any formal system or in LM, provided that mechanism holds.

As for the relation between G- and QG-representability, the former is extensionally included in the latter. Besides, in case the system in question (to be represented) is r.e., the QG-representability of the system is automatically reduced to G-representability of the system. Because of this, in what follows we will be generally concerned with QG-representability.

In terms of QG-representability we can refine R-provability into

QGR-provability. A sentence p is QGR-provable in a system S if and only if p is O-provable in a system S' which is QG-representable in S.

Correspondingly, we can get a special version of the R-principle in terms of QG-representability. We omit its full expression.

Back to the problem of interpreting I1, we now can, according to the R-principle and in terms of QGR-provability, interpret I1 as

$MI1_R$. Any sentence which is O-provable in a sound system is QGR-provable in LM.

To see the difference between MI1R and $MI1_C$, we contrast the QG-representational relationship and the containment relationship. There are many differences between them. For example, there are two systems S and S' such that S is QG-representable in S' but not containable in S'. For another example, an inconsistent system cannot be contained in the logical closure of any consistent system, but it is QG-representable in a proper consistent system. Moreover, a most striking difference between the two relationships is that there is no system which is containmentally maximal if language expansion is always allowed.[49] In contrast however, there are systems which are **QG-representationally maximal** in the sense that any formal system is in principle QG-representable in them.[50] For

---

[49] Here, a system S is containmentally stronger than a system S' if S' is the consequence of S but not vice versa. For any theory S' it is always possible by expanding the language in question to find S such that S' is a proper subset of S and S is not the consequence of S'.

[50] We define that system S as QG-representationally stronger than system S' if and only if S' is QG-representable in S but S is not QG-representable in S'.

example, Z, because it is capable of representing all means of recursion, is such a QG-representationally maximal system. Because of these differences, $MI1_R$ is definitely different from $MI1_C$.

With regards to non-r.e. systems, $MI1_R$ seems to be contrived as an interpretation of I1. But technical details aside,[51] this is to the general effect the only possible interpretation of I1 provided that mechanism holds, and a fortiori, Church's thesis holds. Psychologically analyzed, $MI1_R$ seems to be contrived as an interpretation of I1 only because the existence of non-r.e. systems seems to be a priori and have no bearing on the capacity of the human mind. This is however not the case. Here, we need not enter the ontological controversies over the existential status of non-r.e. systems. It suffices for us to consider how we can come into any knowledge or recognition of these non-r.e. systems. If mechanism holds, then clearly, we can recursively only recognize the r.e. aspects of non-r.e. apparatuses even if these apparatuses really exist in some proper sense. Here, mechanism seems to lead to a paradox of non-r.e.-ness. On the one hand, non-r.e.-ness is beyond r.e.-ness; on the other hand, any non-r.e. system can only

[51] We view as technical details such differences as result from the different versions of representational interpretation, as discussed above.

be described by some r.e. systems. But, this seeming paradox can be resolved by the observation that the existence of non-r.e. systems are proved recursively by and only by being represented in some r.e. systems. To emphasize, in and only in the sense that they can be represented in some r.e. systems that their existence can be proved. It is a most natural implication of the representability in gödelian systems of non-r.e. systems that the distinction between the r.e. and the non-r.e. systems should and can be representable in gödelian systems. That is why though any non-r.e. system is only representable-in-a-way in a gödelian system, the non-r.e.-ness of this system is nonetheless representable in the same way in the same gödelian system as well.[52]

To have a still more adequate conception of $MI1_R$, we should acknowledge fully that $MI1_R$ is, as indicated above, not unique as an interpretation of I1 according to the R-principle. There are others. For example, we can have

$MI1_R$'. Any sound system is **finitely G-**

---

[52] One may object to this point by citing the observation that if the representational interpretation were reasonable, then there would be no sentence which is unprovable in a gödelian system but provable in a nonrecursive system. Here, the observation cited is correct, but it does not serve as an objection. For what it shows is only that contrary to the antimechanist' imagination, any seemingly nonmechanistic job can be done in a mechanistic gödelian system. We will return to this point in chapter 4.

representable in LM.

Here, a system S_1 is said to be finitely G-representable in another system S_2 if any **finite** aspect of the syntax or of the provability predicate of S_1 is G-representable in S_2. We can see that $MI1_R'$ might be weaker in a sense than $MI1_R$ in that S_1 might be finitely G-representable but not be QG-representable in S_2.[53] For another example, recognizing that virtually all structures including the intended model for Z and so on can be conveniently defined in ZFC, we may as well be led to interpret I1 as

$MI1_R$". Any sentence which is O-provable in a sound system is **generally QGR-provable** in LM.

Here, a sentence p is said generally QGR-provable in a system S if and only p is O-provable in a system S' which is QGR- provable either in S directly or in an intermediate system S" which is again QG-representable in S. Needless to say, much explicatory or auxiliary work has to be done before we can have an adequate conception of $MI1_R$". However, we believe this work can be done, though we will not do it in this paper. What is important is that since any gödelian system is itself QG-representable in any other gödelian system, it seems unlikely that there

---

[53] Here we notice that an exact definition of finite G-representation requires an exact definition of being a finite aspect. There may be controversies over this point. But we can pass this point without weakening the illuminating power of our example.

should be any essential difference between $MI1_R$ and $MI1_R$"
with respect to our present concern.

3.6. The R-inference and Its Justification

We now consider the whole P-inference in light of the R-principle and its associates. Accordance to $MI1_R$, the P-inference should be read as

The R-inference

I'1 ($MI1_R$). Any sentence which is O-provable in a sound system is QGR-provable in LM.

I'2. There is a formal sound system in which L_M is O-provable.

I'3. From the above G_{LM} is QGR-provable in LM.

The R-inference as a conditional is valid, for I'3 follows logically from I'1 and I'2. But we still need to justify I'3. Since I'3 follows from I'2 and I'1 and the truth of I'2 is generally granted, in order to justify I'3, we need only to justify I'1. According to the known results of gödelian numbering, in order to justify I'1 we require only that LM is capable of all means of recursion. But this is just what B2 (of the P-argument) claims. We discussed at the beginning of section 3.1 already that B2 should be viewed as acceptable. At least, since B2 is well accepted by both the mechanist and the anti-mechanist, we can simply assume it to be true if we have not been convinced of its truth, without stretching the issue in question. Thus, I'1 and hence I'3 can be viewed as justifiable.

The above shows that I'3 is justifiable as itself, whereas we need to show further that it is so as a proper interpretation of I3. In this regard, since QGR-provability is a typical notion of provability corresponding to the R-principle, we can simply take QGR-provability to be in essence the exclusive alternative to O-provability. Thus we have either

(R) (i.e. the conclusion of the R-inference).

G_{LM} is QGR-provable in LM,

or

(O) (i.e. the conclusion of the P-inference).

G_{LM} is O-provable in LM.

To our knowledge, there has been no special reasons presented for (O) and against (R) in this context except for begging the question to appeal to the thesis of inconsistency. To the contrary, since the truth of (O) is based on the C-principle but the C-principle has been proved implausible, we seem to have one reason for (R) and against (O). This reason becomes more convincing when the following assumption is made:

The uniformity assumption. G_{LM} is provable in LM in the same way in which the Gödel sentence for each gödelian system is provable in LM.

The uniformity assumption is highly plausible. Now, if G_{LM} is O-provable in LM, then, according to the uniformity assumption the Gödel sentence for every gödelian system is O-provable in LM. But we know clearly that no consistent gödelian system could be such that the Gödel sentence for every gödelian system is O-provable in it. The above consideration highlights, and in a sense repeats, that in accordance with G1, G_{LM} cannot be O-provable in LM, not only because any system in which G_{LM} is O-provable has to be non-r.e., but also because no consistent r.e. system can be such that the Gödel sentence for every gödelian system is O-provable in it. This systematic constraint just is demonstrated by G1.

Of course, we do not think the above reason is conclusive. So, with the purpose of providing explanation we would like to consider further some imaginable points which would be taken as against the R-principle and hence against (R).

One point is of the observation that since any inconsistent system is QG-representable in LM and everything is provable in an inconsistent system, any sentence would be QG-representable in LM. This observation is correct. But it does not jeopardize the plausibility of the R-principle. The reason for this is that the pass-or-fail test for the reasonableness of the R-principle is not whether all inconsistent systems are QG-representable in LM, but (a). whether any really sound or consistent theory is QG-representable in LM, and more crucially, (b). whether LM is capable of ascertaining them as acceptable (being consistent, sound, or whatever) to a proper extent. First, for part (a) of the test, it is easy to see that the R-principle cannot fail it.[54] Second, for part (b) of the test, it is true that in many cases LM is incapable of ascertaining sound or consistent systems as such absolutely. But this fact does not mean that LM must fail part (b) of the test in the legiti-mately required sense. For what is and should be required of LM in this regard is that LM be capable of ascertain-ing sound, consistent, or any required systems as such **to and only to the extent to which any possible system can be.** Obviously, if mechanism and Church's thesis hold, then no doubt LM will pass part (b) of the test. For in this case, any two gödelian systems will have the same capacity in this regard, and the capacity of any non-r.e. system in this regard will have to be conditioned upon that of the gödelian systems, in which the non-r.e. system may be representable and hence may gain its existence and any capacity therefrom.

---

[54] As we showed before, any system is QG-representable in any gödelian system. Moreover, there is no essential difference between consistent and inconsistent systems with respect to being represented in a gödelian system.

Since in our justification of the R-principle, mechanism and Church's thesis are presupposed, we can continue to elaborate our theory in spite of the above point of observation.

Here is also a good place for us to consider the second key concept in the R-principle, namely, acceptability. In section 3.5 it was stipulated that system S_2 accepts S_1 if S_1 has not been or cannot be found inconsistent or unsound or the like by virtue of S_2. We admit that this stipulation or explanation is not thorough and clear enough. What is however clear enough in this regard is that we need not necessarily have a thorough definition of this concept before we can meaningfully use it. We need only be sure that for any satisfactory acceptability concept, if S_1 is acceptable for S_2 where S_2 is a consistent gödelian system, then S_1 must be acceptable for any consistent system. For any defect of S_1 concerning its acceptability that can be by any system S or LM can be found by S_2. Again from the above discussion we know that this is guaranteed, granting mechanism and Church's thesis.

A further point which would cause (R) and the R-principle trouble is drawn from the question of whether for an arbitrary sentence p it is ascertainable when p is O-provable and when p is merely QGR-provable. It is natural to ask this question. Since the meaningfulness of QGR-provability does not exclude but rather presupposes that of O-provability, it seems natural that if a sentence p may either be O-provable in LM or be merely QGR-provable in LM, it should be ascertainable in a way which one the case is; but, supposedly against the idea of QGR-provability, it seems that it is in general not ascertainable. We appreciate this point; but at the same time we think that this point, like the first one, only helps spell out implications of (R) and the R-principle; it creates however no real pressure against our viewpoint. The reason is that although it is in general not ascertainable for an arbitrary sentence whether or not it is only QGR-provable in LM, this is so only because we do not know exactly which gödelian system LM happens to be. To be sure, special identification is rather different from general classification. The fact that LM can be reasonably classified as a gödelian system does not entail that it can be identified with all its logical contents. Conversely, the possibility that LM may not be identifiable with all its logical contents does not imply that it cannot be gödelian. As we discussed before on the thesis of the recursiveness of proof, for any sentence p, if p is provable, either O-provable in LM or merely QGR-provable in LM, a proof of p must be effectively specifiable with all the required axioms. It is trivially true that p is O-provable in LM if and only if the axioms required for at least one proof of p are all contained in LM. Thus viewed, then, our inability of ascertaining between O-proofs and QGR-proofs conducted in LM is merely due to our inability of ascertaining all the axioms of LM. Since the latter inability will always be with us even though only O-proofs in LM are justifiable, the former inability should not be a blame of trying to accommodate QGR-proofs as well as O-proofs in LM. We believe that the above discussion has already shown that the reason for accommodating only O-provability is no more convincing than the reason for accommodating both O-provability and QGR-provability. If we have to accept these two provabilities and their possible implications, we should not be bothered much by the ascertainability problem for our present concerns.

Related to the above discussion, it is instructive to consider a question which concerns the logical contents of LM directly. Suppose that LM can be viewed as a gödelian system, how can we specify which gödelian system it is? For example, Z is primarily an artificial formal system. Should LM be Z exactly? If LM is not equivalent to Z (in the sense of containment), is it equivalent to some other gödelian system which is specifiable? It seems that there is no reason why LM has to be some specified gödelian system. Granted that LM should be an ideal model of all the real human brains, LM could be any consistent gödelian system without gaining or losing any rational capacity which some variant of LM would have. Further, generalizing the spirit of the R-principle, we need not even presuppose that LM contain any particular arithmetic truth at all. It suffices for LM to be equivalent to the pure first order predicate calculus, PC. For Z itself can be viewed as the set of all the tautological conditionals in Z's language in each of which the antecedent is a conjunction of some axioms of Z. In this sense Z can be conditionally represented in PC. In fact, it is an open question in which sense LM contains the axioms of Z, supposing that LM contains them in a sense. At least there is so far no reason for excluding the case that LM contains Z only in some conditional sense.

We stop here discussing the reasonableness of (R) and the R-principle directly, not because we believe that enough has been given to establish their truth, but because we think that enough cannot be given for their truth until enough of their applications has been made. We acknowledge that the truth of our whole argument hinges essentially upon their truth.

3.7. The Conclusion

From the above, then instead of claiming the O-provability in LM of G_{LM}, we claim the QGR-provability in LM of G_{LM}. Since provability$_1$ and provability$_2$ introduced near the end of section 3.1 can be interpreted respectively as O-provability and QGR-provability, and it is fairly probable that without begging the question to invoke the thesis of inconsistency G_{LM} should not be taken to be O-provable in LM, we can reach

> **The conclusion of this thesis:** The P-identification is not the only possibility in question, and hence the P-argument is not valid. Put positively, the joint compatibility of mechanism, the thesis of consistency, the thesis of recursive capacity and G1 may well hold in spite of all consequences of G1.

If the above conclusion is correct, then it is easy to explain why it is at least possible both that LM is consistent and that G_{LM} can be viewed as provable in LM. Suppose LM is gödelian and consistent, then according to G1, G_{LM} is not O-provable in LM. However, since G_{LM} is O-provable in another system which is representable in a sense (maybe just QG-representable) and acceptable for LM, G_{LM} may well be R-provable (maybe just QGR-provable) in LM.

To be sure, so far we have not claimed that provability$_2$ can only be interpreted as QGR-provability. To do so we first need to present an independent proof of the thesis of consistency, but we have not done so yet.[55] On the other hand, to show that that could not be the case, one would have to first give an independent proof of the thesis of inconsistency, but no one has done so yet either! In view of this, we just express our present viewpoint, to be further confirmed, that if mechanism holds as Priest thinks, then the only reasonable alternative in this regard is (R), namely, G_{LM} is QGR-provable, instead of O-provable, in LM. In other words, should Priest be correct in claiming the provability in LM of G_{LM}, he would have to mean it in the sense of the QGR-provability in LM of G_{LM}.

3.8. Appendix I (The assumption of non-classical logic)

In advancing the P-argument Priest has in his mind

> **The assumption of non-classical logic:** the logic of the mind is non-classical.

So, to understand the above conclusion better, we should take into consideration the distinction between classical logic and non-classical logic. In this section, we try to make some brief points about this distinction.

One presupposition of our above discussion is that

---

[55] If LM were inconsistent, then provability$_2$ certainly could be interpreted as O-provability, for then every sentence would be provable in LM.

the logic of the mind is of the classical as is typically embodied by the standard first order logic plus certain possible non-logical axioms. Since there is essential difference between classical logic and non-classical logic,[56] we should reformulate our conclusion reached above as to the effect that

> if the logic of the mind is classical, then the joint compatibility of mechanism, the thesis of consistency, the thesis of recursive capacity, and G1 holds.

Now the question is, what would turn up if the logic of the mind is non-classical? Here we anticipate that the defender of the P-argument would make the rejoinder that since the logic of the mind may be non-classical, the above conclusion may lose its meaning. In reply to this rejoinder, we present the following points:

Point 1. There has been so far no hard evidence, and will probably be none as we believe, as to whether the logic of the mind is classical or non-classical. Because of this, it is meaningful to explore the logical nature of the mind under the presupposition that the logic of the

---

56 For our present concerns, the major difference between the classical logic and the non-classical logic lies in that the disjunctive syllogism

$$(p\&(\neg pVq) \vdash p$$

holds for the former but not for the latter whereas inconsistencies are allowed in the latter but not in the former.

mind is classical.

Point 2. If the logic of the mind is non-classical, or especially, just paraconsistent, then the thesis of consistency should of course be rejected or at any rate be revised. However, as our above discussion shows, Gödel incompleteness has no force to entail the thesis of inconsistency. In order to establish this thesis, one cannot appeal to Gödel incompleteness. But this is just what we intend to show.

Point 3. In Priest's view, his dialetheic logic, a kind of paraconsistent logic, is preferable to classical logic as the embodiment of the logic of the mind. His reason is,

> '[a] theory has greater power, or a wider range of applications, or, in general, subsumes a rival is a well known methodological test for that theory to be preferable to its rival. In many ways, dialetheic logic relates to classical logic as does Special Relativity to Newtonian Dynamics.' ([41], Priest, 1987, p. 148)

In our view, although the preference criterion Priest cites is acceptable, his reason is itself problematic. Priest insists on mechanism. But, if his dialetheic logic is mechanical, it should, as we demonstrated above, be representable in any consistent gödelian system. This means, for one thing, that any power that Priest's dialetheic logic possesses will be possessed in a representational sense by the logic of the mind qua a consistent gödelian system, even though the logic of the mind is not itself dialetheic. We have shown that any consistent gödelian system is QG-representationally maximal.(See section 3.5.) This very truth will baffle any attempt of showing that there is a non-classical system which is preferable to the classical as an embodiment of the logic of the mind according to the preference standard cited in Priest's reason. It may be the case that Priest's dialetheic logic is also representationally maximal in a special sense. If this is the case, then we say Priest's dialetheic logic is representationally equivalent to any consistent gödelian system. In this sense, the mere possibility that LM is non-classical is insignificant for our present concerns. For if Priest wants to use Gödel incompleteness to exclude the possibility that LM is classical, he begs the question; whereas if it is possible that LM is classical, then there is possibility that LM is consistent, mechanical, Gödel incomplete, and capable of representing all means of reursion, all at once. But this is just what we try to show by way of refuting the P-argument, without any prejudice asainst Priest's dialetheic logic or any kind of paraconsistent logic.

4. An Examination of the L-argument

The critical examination of the L-argument in this chapter, in close connection with that in last chapter, leads to a refutation of the L-argument. In the P-argument, mechanism is presupposed whereas the thesis of consistency is to be rejected. In contrast, in the L-argument, the thesis of consistency is presupposed, and mechanism is to be rejected. Thus, as the informal formula JI in chapter 1 shows, the P- and the L-argument abstracted as two conditionals are equivalent to each other. In this sense, a refutation of the P-argument is in essence a refutation of the L-argument as well. Because of this, our examination of the L-argument is restricted only to showing that our basic points made on the P-argument are in essence applicable to the L-argument as well. For this purpose we first make an preliminary examination of the L-argument in section 4.1 to locate its pitfall. Then we will concentrate, in section 4.2, on how the representational approach can be carried out in the parlance of Turing machines. Besides, in section 3 we will give the technical proofs of the theorems stated in section 3.

4.1. An Preliminary Examination of the L-argument

Based on considerations similar to those we have for our examination of the P-argument, we can make the following points for preliminary examination of the L-argument. First, there should be no problem with steps A1 and A6. They are the assumptions of the P-argument as well. Second, steps A3, A5 and A7 as conditionals are all valid regardless of whether their antecedents hold or not. Third, step A4, i.e., the thesis of consistency, should be viewed as holding in our examination of the L-argument with a view to defending its compatibility with mechanism under G1 and the thesis of recursive capacity. To emphasize, this compatibility is meaningful and logically examinable even though neither the thesis of consistency nor mechanism could be absolutely confirmed.

From the above points we naturally think that if the L-argument is indeed invalid, its pitfall may probably be found in step 2. The question for the L-argument is the same as for the P-argument: Assuming LM has its own Gödel sentence, in which sense can it be provable in LM? Now, in terms of O-provability, the L-argument means largely the following reasoning: if LM had its own Gödel sentence, say G_{LM}, G_{LM} would be provable in LM; if G_{LM} were provable in LM, it must be O-provable in LM; if G_{LM} were O-provable in LM, LM must be inconsistent; since LM is not inconsistent, LM must have no Gödel sentence of its own, and therefore not be a gödelian system; finally, since LM is capable of all means of recursion, LM must not be mechanical. On examination we can see clearly that G_{LM} is O-provable in LM if and only if the P-identification holds. So, in a sense we can say that the L-argument is valid if and only if the P-identification holds, other conditions given.[57] However, as shown in chapter 3, since the R-principle is acceptable, the P-identification does not hold, for whereas $provability_1$ is certainly O-provability, $provability_2$, the one according to which the Gödel sentence for each gödelian system is provable in LM, may well be QGR-provability. Thus, the case may well be that or only that G_{LM} is O-provable in a system which is QG-representable in LM but not identical to any subsystem of LM. Obviously, in this case, there will be no necessary tension between mechanism and the thesis of consistency unless some argument other than the P-argument could prove valid to the opposite effect.

4.2. Gödel Incompleteness and Turing Machines

In section 2.8, we indicated that in order to refute the L-argument in a satisfactory way, we should show

---

57 Lucas once said that the provability concept invoked in A2 is 'deliberately uncouth' in order to ([Lucas, 1984, p.) But, however uncouth it is, it must be O-provability, for otherwise his argument cannot pass through. It is true that if mechanism were really rejectable, LM could not have its own Gödel sentence, and hence the problem of the P-identification would not arise. However, since invoking Gödel incompleteness is an essential step in the L-argument, the P-identification has to be presupposed for the validity of the L-argument.

explicitly how the Gödel sentence for each machine as a gödelian system in a sense can be determinable and provable in LM without invoking or creating inconsistency. Granting Church's thesis, our discussion of the R-principle and its associates in last chapter serves this purpose in principle, providing a full basis for us to achieve this goal in the representational approach. However, that discussion is not conducted directly in the parlance of Turing machines. Since from the mechanistic point of view the mechanism of the human brain is treated primarily as that of a proper Turing machine and then, only consequently, as that of a formal system due to an equivalence relation between formal systems and Turing machines, an explicit demonstration of the representational approach in the parlance of Turing machines is still needed. We feel this need to be very pressing, especially in view of the fact that in the literature almost all discussions, mechanist or anti-mechanist, about the implications of G1 in the parlance of Turing machines are just too vague or even ambiguous to fit the issue in question. For example, there has been much talk of the Gödel sentence for a Turing machine, but it is just unclear as to how to determine it and its technical implications in correspondence with their counterparts in terms of formal systems. In the following, we try to remedy this situation by explicating what is the case for Gödel incompleteness in the parlance of Turing machines. Our explication consists of three points.

Point 1: the correlation between Turing machines and formal systems.[58] A Turing machine can be viewed as a natural number generator which generates or 'prints' a natural number as its output if it really stops at some stage after being given a natural number as its input. The range of a Turing machine is the set of all the numbers that it can eventually generate as its outputs, and its domain is the set of all the numbers at each of which as its inputs it generates a number as its output. Assuming Church's thesis, then, a Turing machine is equivalent to, and hence can be characterized as, a r.e. function. In this spirit, if a Turing machine M generates b as its output at a as its input, we write M(a)=b. Hereafter we will restrict ourselves to one-place Turing machines and correspondingly one-place universal Turing machines. There is, however, no loss of generality for us to submit to this restriction.[59] Now, in order to develop the same ideas for Turing machines as we do for formal systems, we first need a proper correlation or correspondence relation

---

[58] This point has been well established. We specify it here for the sake of integration.

[59] A discussion about this point is found in [45], Wang, 1963, pp. 152-153.

between formal systems and Turing machines. This correlation can be established through a Gödel numbering. Suppose we are given a proper (first order) language, in which every formal system we address in the following is expressed, and an arbitrary but given Gödel numbering g for that language. We can see that for each formal system S there is a set W_S such that s is a theorem of S if and only if g('s') (i.e. the Gödel number of 's' according to g) belongs to W_S. Let us call W_S **the theoremhood representative set**, or **the TR-set** for short, of S according to g. Now we can establish the correlation between formal systems and Turing machines regarded as theorem proving devices. The key idea is that we can view the correlation in such a way that

> M is **Turing-equivalent** (or **T-equivalent) to S according to g** if and only if the range of M is the TR-set of S according to g.

Up to equivalence in a sense, this is the only proper way to conceive the correlation between Turing machines (or r.e. functions) and formal systems.[60]

Point 2: determination of the Gödel sentence for a Turing machine. According to G1 and the above discussion, for any gödelian system S and any Gödel numbering g, if S is consistent and G_S is the Gödel sentence of S, then

---

60 For a discussion of this point, see ibid, p. 155.

$g('G_S')$ does not belong to W_S where $g('G_S')$ is the Gödel number of G_S according to g and W_S is the TR-set of S. This implies that if a Turing machine M is T-equivalent to S, then the Gödel sentence for M according to g is G_S and its Gödel number $g('G_S')$ is not in the range of M. Such an understanding of Gödel sentence determination seems adequate enough, but it creates a difficulty for us to determine the Gödel sentence for a universal Turing machine. For, naturally, we tend to think that a gödelian system should be equivalent to a universal Turing machine. But, on the one hand, the range of any universal Turing machine is N (the set of all the natural numbers); on the other hand, if a universal Turing machine is T-equivalent to a consistent formal system, the Gödel number of its Gödel sentence has to be outside of N. This observation seems to suggest that either any universal Turing machine can only be T-equivalent to an inconsistent system; or it does not have its own Gödel sentence at all; or the correlation relation as introduced above is improper. However, neither of the three is the case. For although we cannot talk about Gödel sentences for universal Turing machines directly under a usual stipulation, we can do so indirectly under an unusual stipulation. To show this we introduce the concept of

Restricted universal Turing machine. A (two-place) Turing machine M with range R_M is a (one-place) restricted universal Turing machine (a RUT-machine) if and only if for any (one-place) Turing machine M' with range $R_{M'}$, if $R_{M'} \subseteq R_M$, then there is k such that $M(k,y) \equiv M_k(y) \equiv M'(y)$.

Corresponding to the above definition, an ordinary universal Turing machine can now be defined as a RUT-machine with range N. The following points about RUT-machines are significant to our concerns. First, by definition a RUT-machine with range V is capable of representing any Turing machine with range V' such that $V' \subseteq V$. In our view, this fact shows that there is a natural and proper sense in which RUT-machines are regarded as genuine universal Turing machines. Second, the concept of RUT-machine is, due to its allowance of restrictedness, quite general. Any ordinary universal Turing machine is a RUT-machine. Third, the existence of non-trivial RUT-machines other than the ordinary universal Turing machines is also certain. In fact, we have

Theorem A For any r.e. subset V of N, there is a RUT-machine with range V. (A proof is given in section 4.3 (Appendix))

Theorem A guarantees that for any gödelian system S and any Gödel numbering g there is a RUT-machine M such that M is T-equivalent to S according to g. Let us call a RUT-machine with an infinite range an **infinite RUT-machine**. We now can meaningfully talk about Gödel sentences for infinite RUT-machines directly.[61] Fourth, any RUT-machine can be effectively converted into an ordinary universal Turing machine. That is, we have

> **Theorem B** For any infinite RUT-machine M there is a partial 1-1 Turing machine Me such that Me*M, the concatenation of M with Me, is an ordinary universal Turing machine. (A proof is given in Appendix.)

Theorem B implies that there is a good sense in which an infinite RUT-machine is equivalent to an ordinary universal Turing machine.

Point 3: representational provability and consistency. From the above, we know that the Gödel number of

---

[61] For finite RUT-machines we still cannot talk their Gödel sentences directly unless we make some further provisions. For, a closed formal system, being an infinite set of sentences, cannot be T-equivalent to any finite set according to any Gödel numbering. One might think that the concept of RUT-machine is too artificial to be taken seriously. In our view, however, it is a useful concept. True, it is somewhat artificial, but the choice of N as the field (domain plus range) of Gödel numberings and Turing machines is itself artificial, too. Any recursive infinite subset of N can serve as the field of Gödel numberings and Turing machines. The status of the RUT-machines with the same range with respect to N, the standard field of the standard Gödel numberings is isomorphic to that of the ordinary, universal Turing machines with respect to a certain non-standard field of a class of non-standard Gödel numberings. In view of this, one may become more tolerant toward the concept of RUT-machine.

the Gödel sentence for a RUT-machine may not belong to its range. Since an infinite RUT-machine is in essence of the same recursive capacity as a direct universal Turing machine, there is possibility that an infinite RUT-machine is T-equivalent to a consistent system whereas at the same time is able to prove its own Gödel sentence in a representational way which does not ensure inconsistency. This possibility is real. Let us go through the following:

Given

S -- a consistent gödelian system with W_S as its TR-set,

g -- a Gödel numbering for the language of S,

M -- a RUT-machine which is T-equivalent to S according to g, and

G_S -- the Gödel sentence of S according to g.

Since S is consistent, g('G_S') does not belong to W_S. Since M is T-equivalent to S, there is no x and y such that M(x,y)=g('G_S'). This means that M cannot generate g('G_S'). Since S is consistent, there is a r.e. or finite consistent extension S' of S such that G_S is provable in S'. Since S' is r.e. its provability predicate is representable in S. (As we emphasized before, any r.e. set or predicate is representable in any gödelian system.) Now, suppose Pr is the predicate representing the provability relation in S of S'. Then, there is a number i such that i= the Gödel number of 'Pr(j)' according to g where j is the Gödel number of 'G_S' according to g. Let

$Tr_{S'}$={b/ S⊢ Pr(a) for some constant a and b is the Gödel number of 'Pr(a)' according to g}.

Then, $Tr_{S'}$ is the representative Tr-set (in S) of S'. Clearly, $Tr_{S'}$ is r.e.; $Tr_{S'}$ is properly included in the TR-set of S (i.e. W_S); and $i \in Tr_{S'}$. Because of this, there are two numbers k and l such that a) $M(k,y) \equiv M_k(y)$ is a Turing machine; b) the range of M_k is $Tr_{S'}$; and c) $M_k(l)=i$. This means that M is capable of generating all the members of $Tr_{S'}$ including the one which means that G_S is provable in S'. In other words, this is tantamount to saying that all the theorems of S' including G_S are provable by M in a representational sense. To repeat the same thing in still another way, although M does not generate j (i.e. the Gödel number of G_S), M does generates i which means the provability in S' of G_S and S' is representable in M (in the sense of T-equivalence). So, G_S can be viewed as being QGR-provable in M in the same sense in which G_S is QGR-provable in S. Here, needless to say, since S and hence S' is consistent, S' should be acceptable for S, or T-equivalently, M_k is acceptable for M.[62]

---

62 It is interesting to notice that though W_S is properly included in $W_{S'}$, the representative Tr-set in S of S', i.e., $Tr_{S'}$ is nonetheless properly included in W_S.

Our explication of the representational approach in the parlance of Turing machines ends here.

One might wonder whether a RUT-machine would be T-equivalent to a consistent system as well as to an inconsistent system according to different Gödel numberings. The answer to this question is negative. In concrete, for any infinite RUT-machine M, if there is a Gödel numbering g according to which M is T-equivalent to a consistent formal system, then there is no Gödel numbering g' according to which M could be T-equivalent to an inconsistent formal system; and vice versa. This has to be the case because we have

> **Theorem C** For any gödelian system S and any Gödel numbering g, the TR-set of S according to g is a) recursive if S is inconsistent; and b) merely r.e. if S is consistent. (A proof is given in Appendix.)[63]

According to Theorem C and the T-equivalence correlation, all the infinite RUT-machines can be classified into two classes α and β: class α contains those whose ranges are

---

There is no inconsistency or magic here. All this can be well explained in view of ordinary relations between infinite sets, and especially, between r.e. infinite sets.

63 Please notice that Theorem C addresses only gödelian systems. For non-gödelian systems it simply does not hold. For example, there are non-gödelian while decidable and consistent systems whose TR-sets are, according to any Gödel numbering, recursive.

all recursive sets; class β contains those whose ranges are all merely r.e. sets.[64] Theorem C provides us with a criterion by which to bisect infinite RUT-machines with respect to consistency. Specifically, there is reason for us to accept

The Definition of Turing-consistency (for RUT-machines). A RUT-machine is **Turing-consistent** or **T-consistent** if and only if it is T-equivalent to one consistent system according to one Gödel numbering if and only if its range is merely r.e..[65]

The above definition is meaningful. From Theorem B we know that there are T-consistent RUT-machines as well as T-inconsistent RUT-machines. For the former, any RUT-machine with a merely r.e. range is such a one. For the latter, any RUT-machine with a recursive infinite range is such a one.

For a further interesting observation, we have

Theorem D Any RUT-machine in class α can be

---

64 To be sure, a gödelian system cannot be at the same time both consistent and inconsistent. This necessity is reflected in the fact that if the Tr-set of a gödelian system S according to one Gödel numbering g is recursive (or merely r.e.), then there is no other Gödel numbering g' such that the Tr-set of S according to g' is merely r.e. (or recursive).

65 Please note, the above definition is applicable only to RUT-machines.

> converted into one in class β, and vice versa, by being concatenated with a simple Turing machine. (A proof is given in Appendix.)

This fact seems to suggest that any RUT-machine and hence any infinite Turing machine can always be regarded in a way as being inconsistent as well as being consistent. In our view, however, what this fact means is not that the above explanation is improper, but that

> the logic of the mind is neither necessarily inconsistent nor necessarily consistent.

This is an extremely important point by itself. A consequence from this point is, informally, that no Turing machine is crazy or inconsistent in nature. However, since this point is already well beyond the concern of this thesis, we would like to expand it only in another paper.

The above explication is, it seems to us, enough to show that the basic points we made about the P-argument are completely applicable to the L-argument. The only difference is of emphasis. For the P-argument, we emphasize that LM <u>can be consistent</u> under the mechanistic constraints. For the L-argument, we stress that LM <u>can be mechanistic</u> under the constraints of consistency.

We now try to expand the philosophical significance of the above technical points by remarking on some typical viewpoints.

Nelson once expressed the viewpoint that

> 'The proper retort (to Lucas) (I [Nelson] think) is that there is no reason why computers cannot also in principle "produce" such sentences (Gödel sentences) by heuristic means'([36], Nelson, 1980, p. 449)

However, the above explication shows, it seems to me, that not only by heuristic means, but also by effective means of proof only, though in a representative sense, can computers produce their own Gödel sentences.

Hofstadter once observed that Lucas had for his L-argument the idea that

> 'we are always outside the system, and from out there we can always perform the "Gödelizing" operation, which yields something which the program, from within, can't see is true.'([24], Hofstadter, p. 472)

In contrast, the idea drawn from the above explication as against the L-argument is that neither the human mind nor the machine can really perform anything outside their own systems; nonetheless, both can represent their own systems as proper parts of themselves and hence create certain logical environments in which they are really capable of performing things outside their own systems.

Defending mechanism, Arbib once said that

> 'Gödel theorem limits a human as much as a machine'.([2], Arbib, 1987, p. 183)

Now, based on the above explication, we can also say that the representational means as is revealed by, among others, Gödel numberings, empowers a proper machine as much as a human.

According to Wang, Gödel views as a rigorously proved result the following viewpoint:

> 'Either the human mind surpasses all machines (to be more precise: it can decide more number theoretical questions than any machine) or else there exist number theoretical questions undecidable for the human mind.'([46], Wang, 1974, p. 324)

Gödel also agrees, still according to Wang, with Hilbert on rejecting the second disjunct of the above disjunction.(ibid) However, based on our examination of the L-argument, we would rather hold a different viewpoint. That is, any one RUT-machine (the human brain included) can surpass another or itself simply by representing it as a proper part of its own whole program, but none (the human brain being of no exception) can do this in a way in which others cannot do it.

With the purpose to show that

> 'Gödel's work was perhaps the best thing that ever happened to both mechanism and formalism,([44], Webb, 1980, p. vii)

Webb once pointed out that

> 'The existence of a universal machine is essential to Turing's thesis (for mechanicalness of the rationality) and the undecidability of its halting problem is essential to its universality.'(ibid, p. 232)

In Webb's line and for our special concerns, we can say that the existence of a RUT-machine is essential to the compatibility of consistency and mechanicalness; whereas the undecidability of its halting problem is essential to its restricted universality.

Finally, expressing a widely accepted viewpoint, J. Van Heijenoort once pointed out,

> 'Mathematics cannot be completely and consistently formalized in one system.'([2], Heijenoort, 1967, p. 356)

This viewpoint is correct in a sense. Nonetheless we would also like to advocate another viewpoint, namely, that mathematics can be represented in one formalized system to the same extent of completeness and consistency to which it can be represented or entertained in the logic of the mind. The former viewpoint is implied by Gödel incompleteness; whereas the latter viewpoint is implied by QG-representational maximalness or what we would like to call Gödel representational completeness.[66]

4.3. The Appendix

A Proof of Theorem A

Let

M be any ordinary universal Turing machine,

V be any r.e subset of N,

M_V be a total Turing machine which enumerates V,

and

M' be such that $M'(x,y)=M(x,y)$ if $M(x,y)=M_V(z)$

---

[66] We stipulate that a system S is Gödel representationally complete if and only if the Gödel sentence for any gödelian system can be R-provable in S.

for some z, and otherwise undefined.

Then M' is a Turing machine with range V and satisfies the RUT-machine condition.

A Proof of Theorem B

Let

M be any RUT-machine with infinite range V,

M_V be a total 1-1 order-preserving Turing machine which enumerates V,

M_V^- be a Turing machine such that $M_V^-(x)=y$ if $M_V(y)=x$ for some y, and otherwise undefined, and

M' be such that $M' \equiv M_V^- * M$.

Then for any Turing machine M_a, $M_b \equiv M_V * M_a$ is a Tmachine and its range is a subset of V. We have

1) $M_a(y) \equiv (M_V^- * M_V) * M_a(y)$

2) $\equiv M_V^- * (M_V * M_a(y))$

3) $\equiv M_V^- * M(j,y)$ for some j

4) $\equiv M'(j,y)$ for some j

For the above deduction, 1) holds because $M_V^- * M_V$ is a total identity Turing machine; 2) according to the association law; 3) because the range of $M_V * M_a(y)$ is included in V and M is a RUT-machine with range V; and 4) according to the definition of M'. Since M_a is arbitrary, so, M' satisfies the ordinary-universal-Turing-machine condition that for any Turing machine M there is number k such that $M'(k,y) \equiv M_k(y) \equiv M(y)$; and M_V^- is the required Me for M.

A Proof of Theorem C

First, the TR-set of any inconsistent system is, according to any Gödel numbering, the set of the numbers for all the sentences of the language in question. Such a set has to be recursive if so is the whole language in question. To be sure we can simply stipulate that our language in question is recursive. Second, if the Tr-set of a consistent gödelian system were, according to any Gödel numbering, recursive, the system itself would be decidable. However, According to G1, it is impossible.

A Proof of Theorem D

Immediate by application of Theorems A and B.

Bibliography

[1] Anderson, Alan Ross (ed), Minds and machines -- Contemporary perspectives in Philosophy Series, Prentice-Hall, Inc (1964), Englewood Cliffs, N.J..

[2] Arbib, Michael A., Brains, Machines, and Mathematics, Springer-Verlag (1987), New York.

[3] Barwise, J. (ed), Handbook of Logic, North-Holland (1978), Amsterdam.

[4] Bell, J.L. and Machover, M., A Course in Mathematical Logic, North Holland Publishing Company (1977), Amsterdam.

[5] Benacerraf, Paul and Putnam, Hilary (eds), Philosophy of Mathematics: Selected Readings, Cambridge University Press (1983), Cambridge.

[6] Benacerraf, Paul, 'God, The Devil, and Gödel', in The Monist V. 51 (1967), pp. 9-32.

[7] Boolos, George and Jeffrey, Richard C., Computability and Logic, Cambridge Univ. Press (1980), London.

[8] Boyer, David L., 'J.R. Lucas, Kurt Gödel, and Fred Astaire', Philosophical Quarterly Vol. 33 No. 131 (1983), pp. 147-159.

[9] Butrick, Richard, 'The Gödel Formula: some Reservations', Mind, Lxxiv (1965), pp. 411-414.

[10] Chihara, Charles S, 'On Alleged Refutations of Mechanism Using Gödel's Incompleteness Results', Journal of Philosophy V. LXIX (1972), pp. 507-526.

[11] Chihara, Charles S, 'Priest, the Liar, and Gödel', in Journal of Philosophical Logic 13 (1984) pp. 117-124.

[12] Church, A., Introduction to Mathematical Logic, Princeton University Press (1951).

[13] Dennett, Daniel D., 'J.R. Lucas: The Freedom of the Will, in J. of Philosophy Lxix, No. 17 (1972), pp. 527-530.

[14] Dennett, Daniel D., Brainstorms, Bradford Books, Publishers (1978).

[15] Field, Hartry H., Science Without Numbers: A Defense of Nominalism, Basil Blackwell Publisher (1980), Oxford.

[16] George, F.H., The Brain as a Computer, Pergamon Press (1973), New York 10523.

[17] Gödel, Kurt, 'On formally Undecidable Propositions of Principia Mathematica and Related Systems I' (1931), in Van Heijenoort, Jean (ed) [41], pp. 596-616.

[18] Gödel, Kurt, 'What is Kantor's Continuum Problem', in Philosophy of Mathematics: Selected Readings, P. Benacerraf and H. Putnam (eds), Englewood Cliffs (1964), pp. 258-273, New Jersey.

[19] Good, I.J., 'Human and Machine Logic', in British Journal for Philosophy of Science V. 18 (1967), pp. 144-147.

[20] Good, I.J., 'Gödel's Theorem is a Red Herring', in British Journal for Philosophy of Science V. 19 (1969), pp. 357-358.

[21] Heijenoort, J. Van. 'Gödel's Theorem', in Encyclopedia of Philosophy V. 3, Paul Edwards (ed), Macmillan Publishing co., Inc. & the Free Press (1967), New York.

[22] Hilbert, David, 'On the Infinite' (192[illegible]), in Philosophy of Mathematics: Selected Readings P. Benacerraf and H. Putnam (eds), Englewood Cliffs (1964), pp. 134-151.

[23] Hilbert, David, 'The Foundations of mathematics' (1927), in Van Heijenoort, Jean (ed) [41], pp. 446-463.

[24] Hofstadter, Douglas R., Gödel, Escher, Bach: An Eternal Golden Braid, Basic Books, Inc, Publishers (1979), New York.

[25] Kirk, Robert, 'Mental Machinery and Gödel', Synthese 66 (1986) pp.437-452.

[26] Kobrinskii, N.E., and Trakhtenbrot, B.A., Introduc-

tion to the Theory of Finite Automata (translated from Russian), North Holland Publishing Company (1965), Amsterdam.

[27] Lacey, Hugh and Joseph, Geoffery, 'What the Gödel Formula Says', in Mind V. 77 (1968), pp. 77-83.

[28] Lucas, J.R., 'Minds, Machines and Gödel', Philosophy, Vol. XXXVI (1961), reprinted in [1], Alan Ross Anderson (ed) pp. 43-59.

[29] Lucas, J.R., 'Satan Stultified: A Rejoinder to Paul Benacerraf', Monist Lii (1968), pp. 145-158.

[30] Lucas, J.R. The Freedom of the Will, Clarendon Press (1970), Oxford.

[31] Lucas, J.R., 'Metamathematics and the Philosophy of Mind: a Rejoinder', in Philosophy of Science, 38, 1971.

[32] Lucas, R.J., 'Lucas Against Mechanism II: A Rejoinder', Canadian Journal of Philosophy Xiv, No. 2, June (1984), pp. 189-191.

[33] Myhill, John, 'The abstract Theory of Self Reproduction', in Views on General Systems Theory, M.D. Mesarovic (ed), John Wiley & Sons (1964), pp. 106-118.

[34] Napoli, Ernesto, 'priest's Paradox', in Logique et Analyse Annee, 28 (1985), pp. 403-407.

[35] Nelson, R.J., Introduction to Automata, John Wiley & Sons, Inc. (1968), New York.

[36] Nelson, R.J., 'Margraret Boden, Artificial Intelligence and Natural Man, Harvest Press, Hassocks, Gr. Britian (1977)', in Synthese 43 (1980), pp. 433-451.

[37] Nelson, R.J., The Logic of Mind, D. Reidel (1982), Dordrecht, Holland.

[38] Post, Emil. L., 'Absolutely Unsolvable Problems and Relatively Undecidable Propositions: Account of an Anticipation, in M. Davis (ed), The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions, Raven Press (1967), pp. 338-433, Hewlett, New York.

[39] Priest, Graham, 'The Logic of Paradox', Journal of Philosophical Logic Vol. 8 (1979), pp. 219-241.

[40] Priest, Graham, 'Logic of Paradox Revisited', in Journal of Philosophical Logic 13 (1984), pp. 153-179.

[41] Priest, Graham, 'In Contradiction --- a Study of the Transconsistency, Martinus Nijhoff Publishers (1987), Dordrecht.

[42] Putnam, Hilary, 'Minds and Machines: A Symposium', Sidney Hook ed, New York University Press (1960), in Alan Ross Anderson (ed) [1] pp. 72-97.

[43] Smart, J.J.C., Gödel's Theorem, Church's Theorem and Mechanism, in Synthese V. 13 (1961), pp. 105-110.

[44] Van Heijenoort, Jean (ed), From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931, Harvard University Press (1967), Cambridge, Massachusetts.

[45] Wang, Hao, A Survey of Mathematical Logic, Science Press (1963), Peking.

[46] Wang, Hao, From Mathematics to Philosophy, New York Humanities Press (1974), Routledge & Kegan Paul, London.

[47] Webb, Judson Chambers, Mechanism, Mentalism, and Metamathematics --- An Essay on Finitism, D. Reidel (1980), Dordrecht, Holland.

[48] Webb, Judson Chambers, 'Gödel's Theorem and Church's Thesis --- A Prologue to Mechanism', in Language, Logic, and Method, R.S. Cohen and M.W. Wartofsky (eds), D. Reidel (1983), Dordrecht, Holland.