Some Comments on Slezak's Analysis of Descartes, Gödel and Lucas.

Matthew C. Clarke, 1988

1. Introduction

In this essay, I have broken two rules. Firstly, it is an attempt to refute two articles written by the person marking the essay. Secondly, it attempts to present some original material. I hope that my knowledge built up over the last seven years regarding Descartes, computability and Gödel is sufficient to support my audacity.[1]

The essay is in two (largely disconnected) parts. Section 2 describes some difficulties with the article "Descartes's Diagonal Deduction" (referred to throughout this essay as "DDD"). Section 3 gives Lucas some support against the challenges of "Gödel's Theorem and the Mind" (referred to as "GTM").

The overall thesis is that Slezak's arguments, while fascinating, are not rigorous. The faults in Slezak's two articles lead one to judge that his mechanistic conclusion is unjustified.

2. "Descartes's Diagonal Deduction"

In this article, Slezak presents a novel approach to the Cogito which attempts to mesh with "some recent independent discussions of the mind" [DDD p14]. These recent discussions include an apparent cast of thousands : Gunderson, Craik and Minsky, Nagel, Herzberger and others.

One of the central lines of reasoning is to represent all possible human beliefs as an enumerated list of propositions. We need to analyse more closely than Slezak has, whether it is either valid or useful to apply a diagonal argument to such a list.

Following that are some comments on the contribution of Hintikka to the Cogito debate.

2.1 You cannot enumerated a complete set of beliefs

Diagonal arguments are only applicable to infinite sets but they do not apply to all infinite sets. For instance, a diagonal proof can be used to show that the real numbers are not enumerable, but cannot be used to show that the integers are not enumerable.

Hence, it is important that the enumerated list of beliefs proposed by Slezak is infinitely long. However, just because we have an endless list of beliefs does not mean that a diagonal argument can be validly applied.

But let us suppose that arguing diagonally is valid. Would the conclusion be useful ? The outcome of a diagonal proof is that there exists an object which should be on the list, yet is not. Hence the enumeration is incomplete. So if a diagonal proof is appropriate to the

list of beliefs, then the conclusion would be that beliefs cannot be completely enumerated.

What we should conclude from all of this is that enumeration is not a rich enough representation of beliefs. But if that is the case, then the use to which Slezak wants to make of this enumerated list is flawed from the start.

2.2 (x*) would not be on the enumerated list

Once Slezak has started generating beliefs, he inserts onto the list the set -

(x) I doubt (n)

and then the sentence -

(x*) I doubt (x*) [DDD p24]

But (x*) is precisely one of the propositions which cannot be on the list. Once the generation of set "(x) I doubt (n)" is commenced, you are continually generating more propositions to be doubted. Hence we must have the following situation (though the numbering is arbitrary) -

(100) I doubt (1)
(101) I doubt (2)
.
.
(199) I doubt (100)
(200) I doubt (101)
.
.
(299) I doubt (200)
.
etc

The set (x) is infinite and there is no way that the generator will ever produce "(x*) I doubt (x*)".

2.3 (x*) is groundless and does not represent the Cogito

In a footnote, Slezak remarks about the construction of (x*) that he is "not so much defending the argument here as suggesting that it reflects Descartes reasoning" [DDD p24]. But is this the case?

Using the enumeration procedure suggested, if we wish to represent Descartes's words "I doubt just because I cannot doubt that I do so" the best we can do would be -

(80) I doubt

(179) I doubt (80)

(179) can be re-written as "I doubt that I doubt"; and Descartes wants to say that (179) is impossible to be believed. Our intuition agrees with Descartes: it is undoubtable that we doubt. But is this provable?

Suppose we provide an explanation of (80) which fits with our standard usage of English -

"I doubt" is defined to be true if ever "I doubt p" is true for some thought p.

Now if I believe (179) to be true, then it is true that I doubt something (namely (80)), and by the above definition, "I doubt" must be true. Conversely, if I believe (179) to be false, then it must be that I do not doubt (80). Either way it turns out that I must believe (80).

This demonstration that "I doubt" is indubitable is simpler than Slezak's demonstration using (x*), primarily because both (80) and (179) are perfectly well grounded sentences whereas (x*) is groundless[2].

In paraphrasing "I doubt just because I cannot doubt that I do so" as "what (x*) says is true just because I cannot doubt what it says" [DDD p25], Slezak has substituted the grounded phrase "I doubt" with the groundless (x*). This change in groundedness is not surprising since the two propositions "I doubt" and (x*) are not synonymous. (x*) does not represent "I doubt", but the more specific "I doubt something " where the something being doubted isn't even the claim that "I doubt", but rather "I doubt that I doubt that I doubt that ...".

Slezak claims that the enumeration approach climaxing in (x*) "is hardly more than a restatement of Descartes's own reasoning in a more perspicuous form" [DDD p24]. In contrast, the conclusion of this section is that if we are forced to use the enumeration of beliefs, (x*) is not the best representation of the essence of the Cogito.

2.4. (x*) can neither be true nor false

Even if the process of enumeration held sufficient representational power and even if (x*) was equivalent to Descartes's justification of the Cogito, the claim that (x*) is indubitable does not follow.

Slezak argues -

> (x*) involves entertaining the possibility that (x*) is false. But if (x*) is false, this means that it is not the case that I doubt (x*). In other words, the attempt to doubt (x*) which involves entertaining the possibility of its falsity, leads directly to the conclusion that I do not doubt it. In this way (x*) seems to be a proposition which is immune from doubt. [DDD p24]

But one could equally well argue -

> To deny (x*) involves entertaining the possibility that (x*) is true. But if (x*) is true, this means that it is the case that I doubt (x*). In other words, the attempt to accept (x*) which involves entertaining the possibility of its truth, leads directly to the conclusion that I must doubt it. In this way (x*) seems to be a proposition which is infected with doubt.

It is fruitless to argue in this way about groundless sentences, because their very nature refuses to allow the assignment of any truth value.

2.5. The Cogito must be first person and introspective

Slezak is well justified in emphasising that the Cogito is essentially first person and introspective. Descartes does not attempt to prove to his readers that he (Descartes) exists. Rather, he indicates how he convinces himself that he exists and thereby reveals a method by which they may convince themselves that they exist.

Even if the logic of the Cogito is accepted by person P, s/he need not be convinced about its relevance to person Q. Even if Q recites the Cogito, P would have no reason to believe in Q's existence until Q could prove that Q was thinking.

The two go hand in hand: if we are ever convinced that an entity thinks that we are immediately convinced that they also exist. Introspectively we can satisfy ourselves that we think (indeed we find it impossible to deny it) and this leads us to be sure of our existence. However, we cannot apply this to other minds: those other minds must find some alternate way to convince us that they think.

But Hintikka sees this the opposite way -

> If one of these days I should read in the morning paper, "There is no more De Gaulle any more," I could understand what is being said. But no one who knows Charles de Gaulle could help being puzzled by these words if they were uttered by De Gaulle himself; the only way of making sense of them would be to give them a nonliteral meaning. ...
>
> [N]obody can make his hearer believe that he doesn't exist by telling him so; such an attempt is likely to have the opposite result. [Hintikka p12]

There is a strong sense in which this is true in the way we choose to deal with everyday life. But this assurance of the existence of other minds is more a useful social assumption than a logical necessity. When compared to the justification of scepticism in Meditations, we see that Hintikka's reasoning fails to accommodate either Descartes's claim that -

> ...sleep and waking can never be distinguished [Anscombe and Geach p62]

nor his recognition that it is possible…

> ...that there is an evil spirit, who is supremely powerful and intelligent, and does his utmost to deceive me. [Anscombe and Geach p65]

Hintikka throws together the two issues of a person's own existence and the existence of other minds. eg -

> The reason why Descartes could not doubt his own existence is in principle exactly the same as the reason why he could not hope to mislead anybody by saying "I don't exist". [Hintikka p18]

But Descartes explicitly treats these separately. Descartes believes in his own existence because of the Cogito (in the Second Meditation); his belief in the existence of other minds is more closely related to his discussion in the Sixth Meditation (based on his assurance that God does not deceive). Introspection is crucial to the Cogito: it has no power in trying to convince others of one's existence.

In the light of this, Hintikka's third person examples (for instance regarding the existence of Homer and Hamlet) have no relevance in a discussion about the source of the Cogito's power.

2.6 Hintikka covers his tracks

While agreeing that -

> By assimilating the logical force of the Cogito argument to the force it has for a hearer, Hintikka weakens it radically [DDD p28]

there are at least two cases which reveal that Hintikka is not so far off the straight and narrow as Slezak suggests.

Firstly, he upholds the "singularity" of the Cogito and accepts that "Descartes's insight is not generalizable" [Hintikka p20]. The sense in which he means this is that while everyone may construct a sentence proving their own existence (to both themselves and to any audience), we cannot construct a general sentence proving the existence of all other minds. The proof depends on the performance.

This is at least partially true, though it would be better to say that the proof depends on a self-performance.

Second, on the question of whether the logic of the Cogito is enthymatic. Though Hintikka's reconstructs the missing premise as -

B(a) … $x (x = a) [Hintikka p6] (ie if I think, then I exist)

he does not propose that this is adequate (which Slezak seems to imply [DDD p27]). In fact he recognises that Descartes denies that the Cogito is an enthymeme [Hintikka p20] and gives several reasons why the above reconstruction does not lead to any significant conclusions.

2.7 Whatever happened to ergo sum ?

The model based approach of Craik-Minsky and the enumeration approach both lead to conclusions about doubting. At best, Slezak deduces from these two approaches that "I cannot doubt that I doubt", and hence that "I think" is necessarily true ("for what is doubting but thinking in a certain way?" [DDD p25]).

But this conclusion is not worth achieving, on two counts -

First, it must be noted that both the Craik-Minsky and the enumeration approach implicitly assume that the mind being represented has thoughts. For what is a model of the world (W*) apart from a way of representing beliefs (which are thoughts) about the world? And how can a list of thoughts be enumerated unless there are thoughts to enumerate?

So then the deduction "I think" (even with my formulation using (80) and (179)) is a *petito principii*.

Second, to concentrate on "I think" is to overlook half the Cogito. If it is true (as Slezak claims) that Hintikka over-emphasises *sum* , then it is equally true that Slezak over-emphasises *cogito* . It is not the deduction "I think" (nor "I doubt"), but the step between thinking and existing which is the illusive point of the Cogito. But this step is left undiscussed.

3. "Gödel's Theorem and the Mind"

This article (which pre-dates DDD) is a direct attack on J.R.Lucas' paper "Mind's, Machines and Gödel". There are certainly holes in Lucas' argument which Slezak quite rightly picks up. However, I do not believe that the central proposal of Lucas is disproved by GTM.

Gödel's theorem can be extended to strengthen Lucas' case in a way which makes it immune to Slezak's refutation.

3.1 The extended Gödel theorem

The standard rendition of Gödel's theorem is that "in any consistent system which is strong enough to produce simple arithmetic there are formulae which cannot be proved-in-the-system, but which can be seen to be true [from outside the system]". [Lucas p43]

Such a formula can be formed by representing the claim "no demonstration of this sentence is possible" (or more understandably "this sentence is unprovable") within the system. This may be stated symbolically as -

(G) ~(Ex) Dem(x, G)

However, I believe that a corollary can be deduced from (G) which further supports the separation of minds from machines. Let us substitute the formula for G in the place of the symbol G (with a suitable new quantifier) to produce the "extended Gödel theorem"[3] -

(GG) ~(Ex) Dem(x, ~(Ez) Dem(z, G))

GG is more than a piece of meaningless horsing around, for it can be read as -

It cannot be demonstrated that there is no demonstration of G.

The implication of this is that not only are there formulae which cannot be proved-in-the-system, but it is impossible to demonstrate within the system that these formulae cannot be proved-in-the-system.[4] Translating this into the machine metaphor we see that any particular machine, when confronted with its Gödel sentence -

1. Can neither prove nor disprove the sentence (this is the standard result of Gödel's theorem).

but also -

2. Cannot make a response to the effect that "I can neither prove nor disprove that sentence".

The significance of this result for the comparison of minds to machines will become apparent after I have clarified Lucas' idea of out-Gödeling.

3.2 The power of out-Gödeling[5]

Lucas challenges the mechanist with -

> Gödel's theorem must apply to cybernetical machines, because it is of the essence of being a machine, that it should be a concrete instantiation of a formal system. It follows that given any machine which is consistent and capable of doing simple arithmetic, there is a formula which it is incapable of producing as being true - ie the formula is unprovable-in-the-system - but which we can see to be true. [Lucas p44]

This challenge in no way relies on the confusion between the dual meanings of the phrase "producing as true", for in this case Lucas clearly means that although the formula is unprovable (within the system), it is none the less true (from the outside perspective).

Lucas is only restating Gödel's result, but not without intent. For he then explains the practical significance of his restatement -

(1) Every machine can be "out-Gödelled" by a human mind[6].

But if the mechanist claim (that minds are a particular type of machine) is true, then by substitution the following two sentences must be true -

(2) Human minds can be out-Gödelled by other human minds.

(3) Human minds can be out-Gödelled by at least some machines.

I don't think (1) is under dispute, but I will show that (2) and (3) are false and hence that minds cannot be a type of machine.

Slezak claims that a machine with a Gödelizing operator could out-Gödel other machines and hence (1) is not significant [GTM p48]. But it is only insignificant if one first assumes an identity of mind and machine. For consider the two possible cases -

1. The mind is the same type as a machine (only a different token).

   We know that certain machines can out-Gödel certain other machines, so it is insignificant to point out that a mind (which is after all only a machine) can out-Gödel machines.

2. The mind is of different type that a machine.

   Then it is significant to point out that one of the differences between mind and machines is that minds are better Gödelers. In this case, even positing a machine which can out-Gödel every other machine makes no difference to Lucas since even this super-machine could be out-Gödeled by the mind.

The really significant by-product of this is that a necessary condition for mechanism to be true is that the mind is out-Gödelable. It is important then to examine whether the mind is likely to fall prey to Gödel's theorem.

If it were possible to produce a Gödel sentence for a mind, what would it look like? Following the procedure laid out by Gödel we would need to find some way of representing "this sentence cannot be proved" within whatever formal system the mind instantiates.

This may not be as difficult as it sounds: since the mind is capable of natural language manipulation, it may be adequate to present the Gödel sentence in English. The sentence most likely to be Lucas' Gödel sentence seems to be[7] -

(L) Lucas cannot consistently assert (L)

It is clear that Lucas can neither prove nor disprove (L), and precisely because of that we can see that (L) is true. Either a person or a machine can confront Lucas with (L) and successfully out-Gödel him. It is also true that Lucas can out-Gödel his opponent, but this only goes to show that they are different "tokens", not different "types".

Though this is a strong argument, it fails when the "extended Gödel theorem" (GG) is taken into account.

How will Lucas respond when faced with (L)? After thinking for a while, he will discover his dilemma and declare "I can neither prove nor disprove (L)". But according to GG, machines can never recognise that their Gödel sentences are undecideable. Lucas' realisation that (L) is undecideable indicates that he has not been truly out-Gödeled by the proposer of (L).

While it is possible that there are other sentences which could truly out-Gödel the mind, it is far from proven that they must exist. But one must be proven to exist before Gödel's theorem can be used to support mechanism. My observation about people is that they can recognise the paradoxical nature of certain sentences: it is this ability which renders them immune to the extended Gödel theorem.

If it is true that humans can recognise any potential Gödel sentences, it follows that they cannot be an instantiation of any formal system.

4. Slezak's "intriguing speculations" are unsupported

Slezak speculates about the possibility of using diagonalisation and Gödel's theorem to support mechanism. He very cleverly points out some fascinating relationships between Descartes's Cogito, Gunderson's investigational asymmetry, self-reference, Gödel's theorem, Craik and Minsky's model, diagonalisation, and Herzberger's groundedness. This vast network of interconnectedness may lead to some important conclusions, but the strength of the connections will first have to be established more rigorously.

I have shown that at least some of the connections are very brittle: applying a diagonal argument to an enumerated list of beliefs does not lead to Gödel's theorem as Slezak suggests. I have also shown that a corollary of Gödel's theorem strengthens Lucas' criticism of mechanism. Consequently it is extremely unlikely that -

Far from refuting mechanism, Gödel's theorem may even provide the most persuasive support for it. [GTM 51]

Bibliography

Anscombe, E., Geach, P.T.; "Descartes : Philosophical Writings"; pub. Nelson's University Paperbacks, 1970.

Herzberger, H.G.; "Paradoxes of Grounding in Semantics" in The Journal of Philosophy, 1970 vol 67/6.

Hintikka, J.; "Cogito, ergo sum : Inference or Performance ?" in The Philosophical Review, 1962 vol 71.

Hofstadter, D.R.; "Gödel, Escher, Bach : an Eternal Golden Braid"; pub. Penguin 1979.

Lucas, J.R.; "Minds, Machines and Gödel" in Philosophy, 1961 vol 36. Reprinted in "Minds and Machines", ed. A.R.Anderson, pub. Prentice-Hall, 1964.

Slezak, P.; "Descartes's Diagonal Deduction" in British Journal for the Philosophy of Science, 1983 vol 34.

Slezak, P.; "Gödel's Theorem and the Mind" in British Journal for the Philosophy of Science, 1982 vol 33.

End Notes

[1] I promise that my second essay (when I get to it) will be much more orthodox.
[2] Along with Slezak, I draw the concept of "groundedness" from Hintikka.
[3] I can see no reason why this substitution should not be valid, but I believe the result could be arrived at by other means as well. The result seems intuitively safe: surely the only way to produce proofs about proofs is in a meta-system, not within the system itself.
[4] If expressed in terms of machines, I would imagine GG could be proved in a fashion similar to the Halting Problem. The situation where a machine is asked whether a certain sentence is provable or not parallels the situation where a machine is asked whether another machine will halt or not. The latter is demonstrably not do-able (particularly in certain self-referential cases) and I expect the same could be concluded for the former.
Lucas hints at this, but I don't think he appreciates the implications - "The machine cannot answer the question whether it can prove - or cannot prove - the Gödelian formula in the system" (quoted in GTM p47).
[5] Definition : "A can out-Gödel B" abbreviates "System A can concoct a sentence which is a Gödel sentence for system B". (As the linguists say "you can verb anything".)
[6] It doesn't need to be claimed that every human mind is capable of out-Gödelling machines (I know many people who wouldn't have a hope of understanding Gödellisation). It is enough to say that at least some minds have the capability of out-Gödelling any machine.
[7] Originally proposed (I believe) by C.H.Whitely, quoted in Hofstadter p477