On the Question of Whether the Mind can be Mechanized

Peter Koellner
Harvard University

Wuhan, April 21, 2021

Introduction

In this talk I would like to address the question of whether the incompleteness theorems imply that "the mind cannot be mechanized."

This is a large question with a long history.

It is also a question on which it has been difficult to make genuine progress. For without a precise analysis of the concept of "mechanism," it would be hard to even get started. And, even if one did have such an analysis, it would be hard to see how one could give a definitive argument for or against the claim that the mind can be mechanized.

The sitation changed somewhat in the 1930s, through two major developments in mathematical logic.

(1) Turing's analysis of computability.
   - Provided a sharpening of the vague notion of "an idealized finite machine" in terms of the precise mathematical notion of *a Turing machine*.

(2) Gödel's discovery of the incompleteness phenomenon.
   - Demonstrated that for any consistent formal system ($\sim$ Turing machine), there are mathematical truths that cannot be captured by that formal system ($\sim$ computed by that Turing machine).
   - While at the same time made it look to readers of those theorems as though *we* (and hence "the idealized human mind") could capture the mathematical truths that escaped the reach of the formal systems ($\sim$ Turing machines).

So we have here an interesting take on the question of whether

"the mind can be mechanized".

For if we focus on the more precise and specific question of whether

"the mathematical outputs of the idealized human mind can coincide with the mathematical outputs of an idealized finite machine".

we have a hope of making progress.

In this talk I would like to focus on this specific version of the question. I would like to stress:

- I will not be concerned with the performance of actual human minds; I will only be concerned with what the *idealized* human mind can do *in principle*.
- I will not be concerned with a broad array of outputs; I will only be concerned with *mathematical* outputs.
- I will not be concerned with the creative process by which an agent discovers mathematical facts; I will only be concerned with the *outputs*.
- I will not be concerned with other capacities (such as the capacity to form normative judgements, to fall in love, and so on); I will only be considering the ability to produce mathematical outputs.

In short, I will only be addressing the following question:

Question

Do the incompleteness theorems imply that "the mathematical outputs of the idealized human mind do not coincide with the mathematical outputs of any idealized finite machine (Turing machine)."

Gödel actually refrained from answering "Yes" to our question—he did *not* argue that the incompleteness theorems implied that "the mind cannot be mechanized." Instead he argued that his theorems implied a weaker, disjunctive conclusion, namely:

Gödel's Disjunction

Either "the mind cannot be mechanized" or "mathematical truth outstrips human reason."

He thought that each disjunct had important philosophical consequences, each quite different, but both "very decidedly opposed to materialist philosophy."

Namely, if the first alternative holds, this seems to imply that the working of the human mind cannot be reduced to the working of the brain, which to all appearances is a finite machine with a finite number of parts, namely, the neurons and their connections. So apparently one is driven to take some vitalistic viewpoint . . .

... On the other hand, the second alternative, where there exist absolutely undecidable mathematical propositions, seems to disprove the view that mathematics is only our creation; for the creator necessarily knows all the properties of his creatures, because they can't have any others except those he has given them. So this alternative seems to imply that mathematical objects and facts (or at least something in them) exist objectively and independently of our mental acts and decisions, that is to say, [it seems to imply] some form or other of Platonism or "realism" as to the mathematical objects. (Gödel 1951)

Now, it is rarely the case in philosophy that claims are actually established beyond a shadow of a doubt, and this is especially true when those claims concern such large matters as the relationship between mechanism, mind, and mathematical truth.

But Gödel—who was generally quite cautious in his claims—went so far as to call the disjunction a "mathematically established fact."

Our first order of business will be to determine whether the disjunction is indeed a "mathematically established fact."

Others, most notably Lucas and Penrose, have argued that the incompleteness theorems actually imply the first disjunct—that "the mind cannot be mechanized."

Our second order of business will be to first determine whether (a) their particular arguments are correct, and, more generally, whether (b) there can be *any* argument for that conclusion within the framework within which they are working.

One of the difficulties with the discussion in the literature is that the basic assumptions governing the fundamental principles—"an idealized finite machine," "the idealized human mind," and "mathematical truth"—are seldom articulated, and so it is hard to assess the cogency of the arguments that are put forward.

Our strategy will be to sharpen the fundamental notions and articulate the basic principles governing them. This will enable us to pull the entire debate into a setting where we can establish definitive results of the form: *If the fundamental notions and the principles governing them are such and such, then there can be no argument for or against the claim that "the mind can be mechanized."*

Gödel—The Axiomatic Method

In reasoning about mathematics—or any domain for that matter—the question of justification generally gets pushed back further and further until one reaches the bedrock of principles that admit no further justification. These fundamental principles are called *axioms*.

On the bedrock of axioms, one then proceeds in the opposite direction, to construct *proofs*, which, by dint of pure reason, settle undecided questions and establish *theorems*. This whole procedure is organized in a tidy system called a *formal system*.

- ▷ For example, on the basis of the axioms of number theory, Euclid proved that there are infinitely many prime numbers.

For any formal system F there are two basic questions one can ask about it:

(1) Is F *consistent*? That is, is it non-trivial in the sense that it doesn't prove *everything*.
   - Inconsistent systems are powerful, but undiscerning.

(2) Is F *complete*? That is, is it capable of settling *every* question about its subject matter.
   - Consistent and complete theories are both powerful and discerning. They provide us with the most we could hope for.

In *some* areas of mathematics we have consistent and complete formal systems.

- For example, the axioms of Euclidean Geometry are consistent and complete.

If we had consistent and complete formal systems for *all* areas of mathematics, then mathematics would, in principle, be mechanizable. We would have it all. We would have a "grand unified theory"!

Alas, it was not to be ...

The standard axiom system for number theory is known as *Peano Arithmetic* (PA).

Theorem (Gödel)

Assume that PA *is consistent. Then there is a statement* φ *such that* $\mathsf{PA} \nvdash \varphi$ *and* $\mathsf{PA} \nvdash \neg\varphi$*.*

- Strictly speaking, this is Rosser's strengthening of the first incompleteness theorem.

The second incompleteness theorem provides us with a limitation that is particularly close to home.

Theorem (Gödel)

Assume that PA *is consistent. Then* PA *cannot prove* Con(PA).

- To show that PA cannot prove ¬Con(PA) one needs to assume slightly more than the consistency of PA. One needs to assume that PA is Σ^0_1-sound.

It is important to note that these theorems are perfectly general—they apply to any sufficiently strong formal system T meeting minimal conditions.

In short: *Incompleteness is ubiquitous!*

We are interested in the question of what the incompleteness theorems tell us about the interrelations between three fundamental notions:

- F: relative provability ($\sim$what can be produced by an idealized finite machine)
- K: absolute provability ($\sim$what can be produced by the idealized human mind)
- T: truth

Gödel makes three central claims concerning the relationship between F, K, and T.

First Claim

Claim 1

For any formal system F,

$$F \subseteq T \rightarrow F \subsetneq T.$$

- If $F \subseteq T$, then $\mathrm{Con}(F) \in T$. But $\mathrm{Con}(F) \notin F$. So $F \subsetneq T$.

Second Claim

Claim 2

For any formal system F,

$$K(F \subseteq T) \rightarrow F \subsetneq K.$$

- Suppose $K(F \subseteq T)$. We saw above that if $F \subseteq T$ then $\mathrm{Con}(F) \in T$ but $\mathrm{Con}(F) \notin F$. Now, if we also have $K(F \subseteq T)$ then we have $K(\mathrm{Con}(F))$. So $\mathrm{Con}(F) \in K$ but $\mathrm{Con}(F) \notin F$. Thus, $F \subsetneq K$.

It is of interest to note that Gödel did *not* conclude that for any F, $F \subseteq T \rightarrow F \subsetneq K$.

In fact, he was careful to say that no such conclusion was warranted: He explicitly said that for all we have shown, it may indeed be the case that there is a "master system" F^* such that $F^* = K$. We have only shown that if there *is* such an F^* then we can't have $K(F^* \subseteq T)$.

Furthermore, he noted that *if* there were such a "master system" F^*, then it would have important implications.

Third Claim

Claim 3 (Gödel's Disjunction)

Either $(\neg\exists F\,(F = K))$ *or* $(\exists\varphi\,(T(\varphi)\wedge\neg K(\varphi)\wedge\neg K(\neg\varphi)))$.

- Suppose that there were an F^* such that $F^* = K$. Then we have $F^* \subseteq T$ (since $K \subseteq T$). So, by the incompleteness theorems, we have $F^* \subsetneq T$. Thus, there is a $\varphi \in T$ such that $\varphi \notin F^*$, and for any such φ we also have $\neg\varphi \notin F^*$ (since if $\neg\varphi \in F^*$ then $\neg\varphi \in T$ (as $F^* \subseteq T$), which is impossible since $\varphi \in T$). But $F^* = K$. So this φ is such that $\varphi \in T$ and yet neither $\varphi \in K$ nor $\neg\varphi \in K$.

Sharpening the Notions

The above arguments are informal. To render them precise we need to spell out the background assumptions on F, K and T.

It is straightforward to do this for F and T.

- Turing gave us a substantive analysis of F.
- Tarski gave us a structural analysis of T.

But what about K?

In the case of K there is no hope of giving a *substantive* analysis—the most that one could hope for is a *structural* analysis. The trouble is that there is little agreement on the element of idealization involved in the notion of "absolute provability" ($\sim$ "the idealized human mind").

Aside: I have to confess: I am skeptical of the very idealization involved. In particular, I do not understand the lines along which we are idealizing when we target "the idealized human mind." The reason that I do not understand this, is that in contrast to the case of successful idealizations—like frictionless planes and Turing machines—we presently have no good model of how the mind actually works. So as far as I can see, we might as well be speaking of "the angelic mind."

But here I would like to lay those worries aside, and rest my case on mathematical results rather than on philosophical misgivings.

I would like to stress that the idealizing assumptions that I will be making are made on behalf of my opponent. Furthermore, I wish to be as charitable as possible. For the more I grant concerning K the easier it will be for my opponent to show that K must outstrip any F.

"The strength of a criticism is proportional to the degree to which it is charitable."

The System $\mathsf{EA_T}$

The axioms governing F are simply the standard axioms of arithmetic, namely, PA.

The axioms of K are (the universal closures of):

(E_1) $K\varphi$, where φ is valid.

(E_2) $K(\varphi \to \psi) \to K\varphi \to K\psi$.

(E_3) $K\varphi \to \varphi$.

(E_4) $K\varphi \to KK\varphi$.

The resulting system is called EA (Epistemic Arithmetic).

- Notice that EA embodies a highly idealized notion of K.

The axioms governing T are the standard Tarskian axioms. They include such axioms as:

(T_1) $(\forall x)[\mathsf{Sent}(x) \rightarrow (T(\neg x) \leftrightarrow \neg T(x))]$ and

(T_2) $(\forall x)(\forall y)[\mathsf{Sent}(x, y) \rightarrow (T(x \vee y) \leftrightarrow T(x) \vee T(y))]$.

The resulting system is called $\mathsf{EA_T}$ (Epistemic Arithmetic with Truth).

Gödel Revisited

Recall Gödel's three central claims:

(1) *For any formal system* F, $F \subseteq T \rightarrow F \subsetneq T$.

(2) *For any formal system* F, $K(F \subseteq T) \rightarrow F \subsetneq K$.

(3) *Either* $\neg\exists F\,(F = K)$ *or* $\exists\varphi\,\big(T(\varphi) \wedge \neg K(\varphi) \wedge \neg K(\neg\varphi)\big)$.

Theorem (Reinhardt)

Each of Gödel's three claims is provable in $\mathsf{EA_T}$*; in particular, Gödel's Disjunction is provable in* $\mathsf{EA_T}$*.*

In this sense, the disjunction is indeed "a mathematically established fact."

So . . . Which is it?

Is it the case that the "mind cannot be mechanized" or is it the case that "there are absolutely undecidable sentences"?

Lucas-Penrose—The Particular Argument

Lucas and Penrose argued that the incompleteness theorems actually implied the first disjunct—that "the mind cannot be mechanized."

The central claim is that:

> *"Human mathematicians are not using a knowably sound algorithm in order to ascertain mathematical truth"—Penrose, 1994*

We saw that this is actually provable in $\mathsf{EA_T}$: It is just Gödel's second claim: $K(F \subseteq T) \to F \subsetneq K$.

But notice that this is a *conditional* statment. To discharge the antecedent one must be able to determine whether or not F is correct.

This is no small task. For example, let F be the system $\mathsf{PA} + R$ where 'R' stands for Riemann Hypothesis.

Fact

$\mathsf{PA} + R$ *is consistent if and only if* R *is true.*

- It follows that no one knows whether or not $\mathsf{PA} + R$ is consistent.

The point is much stronger: It is not about just this one sentence, R.

- To have an oracle for the question of whether or not an arbitrary F is consistent is to have an oracle for Π^0_1-truth. But that is not something one can assume *at the start* of an argument for the claim that "the mind cannot be mechanized" since it *trivially* implies the conclusion.

So, this *particular* argument for the first disjunct fails ...

The General Situation

But perhaps there is another argument …

Let's distinguish between three grades of mechanism:

(1) (WMT) $\exists e\,(K = F_e)$
- "there is a machine such that I am that machine"

(2) (SMT) $K\,\exists e\,(K = F_e)$
- "I know that there there is a machine such that I am that machine"

(3) (SSMT) $\exists e\,K(K = F_e)$
- "There is a machine such that I know that I am *that* machine"

Theorem (Reinhardt)

"$\mathsf{EA_T} + \mathsf{WMT}$" is consistent.

- So $\mathsf{EA_T}$ cannot rule out that I am a machine.

Theorem (Carlson)

"$\mathsf{EA_T} + \mathsf{SMT}$" is consistent.

- So $\mathsf{EA_T}$ cannot even rule out that I *know* that I am a machine.

What, then, could have led people astray?

Theorem (Reinhardt)

"$\mathsf{EA}_\mathsf{T} + \mathsf{SSMT}$" *is inconsistent.*

- So EA_T *can* rule out that there is a machine such that I know that I am *that* machine.

Perhaps what led people astray was an unjustified slide from

$$\neg\exists e(K(K = F_e))$$

$$\rightarrow \neg K(\exists e(K = F_e))$$

$$\rightarrow \neg\exists e(K = F_e)$$

In any case, regardless of what may have led people astray, the particular argument given turns on an error, and, in fact, there can be no argument for the conclusion within EA_T, a system that would seem to embody every assumption that the proponents of the first disjunct would be willing to make.

I hope that this puts to rest the first generation of arguments for the first disjunct, and that all participants in the debate can agree.

Penrose's New Argument

The above results take place in EA_T where one avoids the paradoxes of K by treating it as an *operator* and one avoids the paradoxes of T treating it as *typed*.

Gödel did not think that the typed approach was an adequate approach to either K or T. Moreover, he held that:

> *If one could clear up the intensional paradoxes somehow, one would get a clear proof that mind is not [a] machine. (Wang 1996, p. 187)*

Curiously, Penrose's new argument requires a type-free truth predicate. So perhaps Penrose has fulfilled Gödel's hope.

Though I don't know that I necessarily am F, I conclude that if I were, the system F would have to be sound and, more to the point, F' would be sound, where F' is F supplemented by the further assertion "I am F." I perceive that it follows from the assumption that I am F that the Gödel statement $G(F')$ would have to be true and, furthermore, that it would not be a consequence of F'. But I have just perceived that "if I happened to be F, then $G(F')$ would have to be true", and perceptions of this nature would be precisely what F is supposed to achieve. Since I am therefore capable of perceiving something beyond the powers of F, I deduce that I cannot be F after all.

Formal

(1) $K = F_e \rightarrow F_e \subseteq T$

[Uses "$K \rightarrow T$".]

(2) $K = F_e \rightarrow F_{e^+} \subseteq T$

[Uses "T-In". (Here n^+ is such that $F_{e^+} = (F_e + (K = F_e))$.)]

(3) $K = F_e \rightarrow G(F_{e^+})$

[Follows from (2) by the incompleteness theorem. (Here '$G(F_{e^+})$' is the Gödel sentence for F_{e^+}.)]

(4) $K = F_e \rightarrow F_{e^+} \nvdash G(F_{e^+})$

[By the incompleteness theorem.]

(5) $K(K = F_e \rightarrow G(F_{e^+}))$

[Uses "K-Intro" (on (3)).]

(6) $K = F_e \rightarrow (F_e \nvdash (K = F_e \rightarrow G(F_{e^+}))$

[By (4).]

(7) $K = F_e \rightarrow K \neq F_e$

[By (5) and (6).]

(8) $K \neq F_e$

The key principles used are:

(a) $K \to T$

(b) T-In

(c) K-Intro

Notice that in (2) T is applied to K and with K-Intro K is applied to T, for example, with K-Intro applied to line (3). So T is applied to T and K is applied to K.
We thus need a type-free theory of truth and knowledge to formalize this argument.

Sharpening the Notions: DTK

The type-free system of truth I shall employ is Feferman's system DT.

The base language is $\mathscr{L}_{\mathsf{PA}}$ and the base system PA. The language is extended by adding a (type-free) truth predicate 'T', allowing it to figure in induction, and adding axioms of *determinateness* and *truth*. Here determinateness is symbolized by letting '$D(x)$' be short for '$T(x) \vee T(\neg x)$'. The primitive connectives are '$\neg$', '$\vee$', and '$\rightarrow$'.

The axioms governing 'D' include

(D_2) $(\forall x)[\mathsf{Sent}(x) \rightarrow (D(\dot{\neg} x) \leftrightarrow D(x))]$

(D_3) $(\forall x)(\forall y)[\mathsf{Sent}(x) \wedge \mathsf{Sent}(y) \rightarrow (D(x \dot{\vee} y) \leftrightarrow D(x) \wedge D(y))]$

(D_4) $(\forall x)(\forall y)[\mathsf{Sent}(x) \wedge \mathsf{Sent}(y) \rightarrow (D(x \dot{\rightarrow} y) \leftrightarrow D(x) \wedge (T(x) \rightarrow D(y))]$

The axioms governing 'T' include

(T_2) $(\forall x)[\mathsf{Sent}(x) \wedge D(x) \rightarrow (T(\dot{\neg} x) \leftrightarrow \neg T(x))]$

(T_3) $(\forall x)(\forall y)[\mathsf{Sent}(x) \wedge \mathsf{Sent}(y) \wedge D(x \dot{\vee} y) \rightarrow (T(x \dot{\vee} y) \leftrightarrow T(x) \vee T(y))]$

(T_4) $(\forall x)(\forall y)[\mathsf{Sent}(x) \wedge \mathsf{Sent}(y) \wedge D(x \dot{\rightarrow} y) \rightarrow (T(x \dot{\rightarrow} y) \leftrightarrow T(x) \rightarrow T(y))]$

Theorem (Feferman)

For all $\varphi \in L_{\mathsf{DT}}$

$$\mathsf{DT} \vdash D(\ulcorner\varphi\urcorner) \rightarrow (T(\ulcorner\varphi\urcorner) \leftrightarrow \varphi).$$

Now, it would be natural to merge this theory of truth (DT) with our theory of absolute provability (EA), adding the $K \rightarrow T$ principle. Let $\mathsf{EA}_{\mathsf{DT}}$ be the resulting system.

Theorem

$\mathsf{EA}_{\mathsf{DT}}$ *is inconsistent.*

Perhaps this is not so surprising in hindsight. For once we enter the realm of indeterminate sentences we should condition the axioms of absolute provability on determinateness.

Surprisingly, when one does this one can circumvent the limitative results of Gödel, Montague, and Thompson and treat 'K' as a predicate.

The System DTK

The axioms governing 'K' are:

(K_1) $(\forall x)[\mathsf{Sent}(x) \rightarrow (K(x) \rightarrow T(x))]$

(K_2) $(\forall x)(\forall y)[\mathsf{Sent}(x) \wedge \mathsf{Sent}(y) \rightarrow (K(x \dot{\rightarrow} y) \rightarrow (K(x) \rightarrow K(y)))]$

(K_3) $(\forall x)[\mathsf{Sent}(x) \rightarrow (K(x) \rightarrow K(\dot{K}(\dot{x})))]$

We also have two rules: DK-Intro and DT-Intro:

$$\frac{\varphi \wedge D(\ulcorner \varphi \urcorner)}{K(\ulcorner \varphi \urcorner)}$$

$$\frac{\varphi \wedge D(\ulcorner \varphi \urcorner)}{T(\ulcorner \varphi \urcorner)}$$

Theorem

DTK *is consistent.*

Gödel Revisited

Recall Gödel's three central claims:

(1) *For any formal system* F, $F \subseteq T \rightarrow F \subsetneq T$.

(2) *For any formal system* F, $K(F \subseteq T) \rightarrow F \subsetneq K$.

(3) *Either* $\neg\exists F\,(F = K)$ *or* $\exists \varphi\,(\varphi \in T \wedge \varphi \notin K \wedge \neg\varphi \notin K)$.

The first claim is easily provable in DTK.

For the second claim, once again we have:

Theorem

Assume that T includes DTK. *Suppose $F(x)$ is a formula in L_{PA} such that*

$$T \vdash K(\ulcorner F(\ulcorner \varphi \urcorner) \rightarrow \varphi \urcorner)$$

for each $\varphi \in L_{\mathsf{PA}}$. Then there is a sentence $\bar{\varphi}$ such that

$$T \vdash K(\ulcorner \bar{\varphi} \urcorner) \wedge K(\ulcorner \neg F(\ulcorner \bar{\varphi} \urcorner) \urcorner).$$

For the third claim, once again we have:

Theorem

DTK $\vdash$ GD

So . . . Which is it?

Is it the case that the "mind cannot be mechanized" or is it the case that "there are absolutely undecidable sentences"?

Penrose Revisited—The Particular Argument

Theorem

For all $\varphi \in L_{\mathsf{DTK}}$

$$\mathsf{DTK} \vdash D(\ulcorner \varphi \urcorner) \rightarrow (T(\ulcorner \varphi \urcorner) \leftrightarrow \varphi).$$

So we have T-Out and DT-In.

Recall that Penrose's new argument uses:

(a) $K \rightarrow T$

(b) T-In

(c) K-Intro

We have (a) but in place of (b) and (c) to skirt inconsistency we now only have the restricted forms DT-In and DK-Intro.

To apply these rules we need to know that the sentences in question—namely, "$K = F_e \rightarrow F_{e^+} \subseteq T$" and "$K(K = F_e \rightarrow G(F_{e^+}))$" are determinate.

Theorem

Assume DTK. *The statements*

(1) $K = F_e \rightarrow F_{e^+} \subseteq T$

(2) $K(K = F_e \rightarrow G(F_{e^+}))$

are indeterminate.

So this *particular* argument for ¬WMT fails in DTK.

But perhaps there is *another* argument ....

The General Situation

In fact:

Theorem

$\mathsf{DTK} \vdash \neg D(\mathsf{GD}) \wedge \neg D(\neg \mathsf{WMT}) \wedge \neg D(\mathsf{AU})$.

There is a particular irony here. For to have a hope of proving the first disjunct in its full generality we have had to switch to a system in which truth is type-free. But when we make this shift we have to be on the lookout for sentences which (like the liar sentence) are indeterminate. The irony is that the very statement we set out to prove—the first disjunct, in the full version—is provably indeterminate.

But perhaps if we restrict to determinate sublanguages then we secure determinateness.

In fact, we do:

Theorem

$\mathsf{DTK} \vdash D(\mathsf{GD}_{\mathsf{PA}}) \wedge D(\neg \mathsf{WMT}_{\mathsf{PA}}) \wedge D(\mathsf{AU}_{\mathsf{PA}})$.

So in this case there is at least a hope of getting started.

Theorem

Assume that DTK *is correct for arithmetical statements. Then* DTK *can neither prove nor refute either* $\neg\mathsf{WMT}_{\mathsf{PA}}$ *or* $\mathsf{AU}_{\mathsf{PA}}$*.*

In other words, in the restricted case we have determinateness, the disjunction is provable, but neither disjunct is provable or refutable in DTK.

In particular, there is *no* argument for the first disjunct in DTK.

One might try to salvage the situation by bolstering the theory of knowledge.

But the result is robust.

For example, let DTK' be the result of adding

(K_4) $(\forall x)[\mathsf{Sent}(x) \wedge K(\ulcorner D(x)\urcorner) \rightarrow K(\ulcorner K(\dot{x}) \rightarrow T(\dot{x})\urcorner)]$

(K_5) The scheme: For each formula φ

$$K(\ulcorner(\forall x)\varphi(x)\urcorner) \rightarrow (\forall x)\, K(\ulcorner \varphi(\dot{x})\urcorner).$$

Theorem

Assume that DTK' *is correct for arithmetical statements. Then* DTK *can neither prove nor refute either* $\neg\mathsf{WMT}_{\mathsf{PA}}$ *or* $\mathsf{AU}_{\mathsf{PA}}$*.*

I hope that this puts to rest the second generation of arguments for the first disjunct.

I have really only scratched the surface in the last part of this talk. There is much more to be said. The results are rather robust: One can change the truth theory and one can supplment the principles of knowledge with further principles of knowledge, and independence persists.

In the end, I think that we can draw a disjunctive conclusion of our own:

> *Either the statements that "the mind can be mechanized" and "there are absolutely undecidable statements" are indefinite (as the philosophical critique maintains) or they are definite and the above results provide evidence that they are about as good examples of "absolutely undecidable" propositions as one might find.*

Thank You