Logic and Logical Philosophy
Online First Articles (2024)
DOI: 10.12775/LLP.2024.025

Antonella Corradini
Sergio Galvan

A Note on Gödel's First Disjunct Formalised in DTK System

Abstract. This note clarifies the significance of the proof of Gödel's first disjunct obtained through the formalisation of Penrose's second argument within the **DTK** system. It analyses two formulations of the first disjunct — one general and the other restricted — and dwells on the demonstration of the restricted version, showing that it yields the following result: if by F we denote the set of propositions derivable from any formalism and by K the set of mathematical propositions humanly knowable, then, given certain conditions, F necessarily differs from K. Thus it is possible that K surpasses F but also, on the contrary, that F surpasses K. In the latter case, however, the consistency of F is humanly undecidable.

Keywords: Penrose's second argument; Gödel's disjunction; **DT** system; **DTK** system; computational model of mind; arguments in favour of the first horn of Gödel's disjunction

Introduction

In this note we aim to illustrate some consequences that can be derived from Penrose's second argument formalised within the **DTK** system. These consequences also have a strong impact on the interpretation of the first disjunct of Godel's disjunction (see Gödel, 1995, p. 310). Penrose's argument (see 1994; 1996) aims at the proof of the first disjunct, i.e., the proving of the essential difference between mind and machine. Now the greatest difficulties in Penrose's argumentation come from claims involving the use of a type-free predicate of truth, which exposes the argument to incorrect paradoxical passages. Koellner (2016, 2018a,b) (see

Received November 7, 2022. *Accepted* September 27, 2024. *Published online* October 23, 2024

also Corradini and Galvan, 2022) proposed to formalise the argument within the **DTK** system precisely to avoid the incorrectness attributable to truth paradoxes. **DTK** is, in fact, an extension of Feferman's system **DT**, (see 2008), which is characterised by the knowability predicate K (or absolute provability, as it is called by Koellner), whereas the **DT** system is itself characterised by a type-free predicate of truth T and the determinateness predicate D. The predicate of determinateness serves, precisely, to contain the destructive effects of paradoxical situations generated by the untyped character of the truth predicate. By means of predicate D it is possible to guard against incorrect passages attributable to the use of indeterminate sentences, i.e., those that are incapable of being true or false, such as paradoxical sentences. Regarding the formal structure of the system, we refer to (Koellner, 2016) and (Corradini and Galvan, 2022). Here we focus only on certain — also notational — aspects of particular importance.

A. An aspect of particular importance is the fact that the syntax of **DTK** is completely arithmetisable. Thus, while $\vdash_{\mathbf{DTK}} \varphi$ is the usual syntactic expression of derivability in **DTK** of the formula φ, $\mathrm{PR}_{\mathbf{DTK}}(\ulcorner\varphi\urcorner)$ is the result of the arithmetisation of $\vdash_{\mathbf{DTK}} \varphi$, where $\mathrm{PR}_{\mathbf{DTK}}$ is a recursively enumerable predicate and $\ulcorner\varphi\urcorner$ represents the code of the proposition φ. Clearly to L(**DTK**) also belongs the expression $\mathrm{PR}_F(\ulcorner\varphi\urcorner)$, which arithmetises the concept of derivability of φ in the system F and is usually denoted by $\vdash_F \varphi$ and is shortened to $F(\ulcorner\varphi\urcorner)$. Also the predicates D, K and T have codes as arguments. Thus $D(\ulcorner\varphi\urcorner)$ means that proposition φ is determinate, $K(\ulcorner\varphi\urcorner)$ means that proposition φ is known and $T(\ulcorner\varphi\urcorner)$ means that proposition φ is true. In what follows, however, we will use a metatheoretical simplified language, omitting the code function. Hence, $F(\ulcorner\varphi\urcorner)$ is simplified to $F\varphi$, $K(\ulcorner\varphi\urcorner)$ to $K\varphi$, $D(\ulcorner\varphi\urcorner)$ to $D\varphi$ and $T(\ulcorner\varphi\urcorner)$ to $T\varphi$.

Last, it is worth noting that F serves both as a symbol for a formal system, the concept of formal derivability in the system, and as a symbol for the set of theorems of the system itself. Moreover predicate F corresponds to the Gödelian idea of formalised mathematics. K, on the other hand, is the predicate intended to translate the Gödelian idea of subjective mathematics, i.e., mathematics that is accessible with certainty to the human mind, while T is that of objective mathematics. Therefore, F, K and T may also be understood as sets: F as the set of theorems of the corresponding F-system, K as the set of humanly knowable propositions and T as the set of true mathematical propositions.

B. Referring back to the abovementioned place for a full discussion of the three predicates, we can limit ourselves to emphasising three fundamental principles of the predicate K.

1. K-correctness principle: $\vdash_{\mathbf{DTK}} K\varphi \rightarrow \varphi$

Correctness is a universally recognised characteristic of knowledge. One cannot consider oneself to know if what one knows is not true.

2. K-intro rule: $\vdash_{\mathbf{DTK}} \varphi \wedge D\varphi \;\Rightarrow\; \vdash_{\mathbf{DTK}} K\varphi$

It should be noted that the K-intro rule is correct only if the premise satisfies the condition of determinateness. Therefore, to derive $\vdash_{\mathbf{DTK}} K\varphi$ it is not sufficient that $\vdash_{\mathbf{DTK}} \varphi$, but it is also necessary that the determinateness of φ be derivable, i.e., $\vdash_{\mathbf{DTK}} D\varphi$. The consequences of this are significant. Tarski's biconditional $T\varphi \leftrightarrow \varphi$ does not apply unconditionally, but only under the condition of determinateness of the proposition concerned. That is, $\vdash_{\mathbf{DTK}} D\varphi \rightarrow (T\varphi \leftrightarrow \varphi)$. What applies unconditionally is $\vdash_{\mathbf{DTK}} T\varphi \rightarrow \varphi$ but not $\vdash_{\mathbf{DTK}} \varphi \rightarrow T\varphi$.

With the K-intro rule, the function of the predicate of determinateness D becomes clear. Only determinate propositions can be known and only these are known as true.

3. Rule of mathematical knowledge: $\vdash_{\mathbf{PA}} \varphi \Rightarrow \vdash_{\mathbf{DTK}} K\varphi$ i.e., $\mathrm{PR}_{\mathbf{PA}}\varphi \vdash_{\mathbf{DTK}} K\varphi$

Arithmetical propositions are determinate and, consequently, if they are theorems, they are also knowable and therefore true.

1. DTK system and the Gödel's disjunction

How can Gödel's disjunction be obtained in the **DTK** system? The disjunction can be obtained in two different formulations: a general and a restricted. The former concerns all propositions that can be formulated in the language of **DTK**, while the latter is restricted to arithmetical propositions only — i.e. belonging to L(**PA**), which, as we know, is a part of L(**DTK**). In particular this difference can be seen in the formulation of the first disjunct that includes the expression of the identity between "Mind" and "Machine". This identity can be formulated in two ways:

Definition 1. $K = F := (\forall\varphi \in \mathrm{L}(\mathbf{DTK}))(K\varphi \leftrightarrow F\varphi)$.

That is, the set of propositions belonging to the language of **DTK** and derivable in F coincides with the set of propositions belonging to the same language that are humanly knowable, i.e., the set of propositions belonging to the language of **DTK** and derivable in F coincides with the set of propositions belonging to the same language that are humanly knowable.

DEFINITION 2. $K =_{\mathbf{PA}} F := (\forall\varphi \in \mathrm{L}(\mathbf{PA}))(K\varphi \leftrightarrow F\varphi)$.

That is, the set of propositions belonging to the language of **PA** and derivable in F coincides with the set of propositions belonging to the same language that are humanly knowable.

Depending on these two formulations, one can have two forms of Gödel's disjunction in **DTK**:

FORMULATION 1. $\vdash_{\mathbf{DTK}} \neg\exists F(K = F) \lor \exists\varphi(T\varphi \land \neg K\varphi \land \neg K\neg\varphi)$.

That is, either there is no formalism F capable of deriving all propositions belonging to the language of **DTK** that are humanly knowable — i.e. belong to K — or there is a proposition of that language that is true and humanly undecidable.

FORMULATION 2. $\vdash_{\mathbf{DTK}} \neg\exists F(K =_{\mathbf{PA}} F) \lor (\exists\varphi \in \mathrm{L}(\mathbf{PA}))(T\varphi \land \neg K\varphi \land \neg K\neg\varphi)$.

That is, either there is no formalism F capable of deriving all propositions belonging to the language of **PA** that are humanly knowable — i.e., belong to K — or there is a proposition of that language that is true and humanly undecidable.

Both formulations are derivable in **DTK** (see Koellner, 2016, pp. 174–176). However, there is an important difference: only the restricted formulation is determinate, so that only it is provably knowable and consequently true. It alone can therefore be declared fully acceptable. What is the reason for this difference? As we have already mentioned, **DTK** is characterised by the possibility of containing indeterminate propositions, i.e., neither true nor false. The predicate of determinateness is defined in such a way as to ensure the conditions of formation and derivation of determinate propositions. Among these conditions, there is also the condition that all mathematical propositions (belonging to the language of **PA**) are determinate; this is why the restriction of the assertion of equivalence to mathematical propositions ensures the determinateness of equivalence. This difference also plays an important role in the proof of the first disjunct. For it is one thing to prove $\neg\exists F(K = F)$ and another to prove $\neg\exists F(K =_{\mathbf{PA}} F)$. While it is possible to prove the first formulation, it is not possible to prove the second.

2. DTK and Gödel's first disjunct

Koellner (2016, Theorem 7.16.2(1), pp. 177–179) first achieved the following result on the first disjunct:

THEOREM 1. $\nvdash_{\mathbf{DTK}} \neg\exists F(K =_{\mathbf{PA}} F)$, *i.e., it is impossible to prove in* **DTK** *that no formalism exists that is capable of proving all mathematical propositions (belonging to the language of Peano's arithmetic) knowable by the human mind.*

This is an important result of independence. It makes it clear that it is not ruled out that there exists a formalism capable of proving all mathematical propositions knowable by the human mind. The restriction to mathematical propositions is essential. We already know why the assertion of equivalence between K and F must be restricted to mathematical propositions. The restriction ensures the determinateness — and thus the possibility of being true — of the proposition whose nonderivability is asserted in **DTK**. Therefore, while $\neg\exists F(K =_{\mathbf{PA}} F)$ is a determinate proposition — and thus susceptible of being true — the same unrestricted proposition $\neg\exists F(K = F)$ is not. On the basis of this result — and of the further theorem 7.16.2(2) concerning the second disjunct, which we do not consider here — Koellner states that it is impossible to decide on any disjuncts in **DTK**.[1]

In reality, it is possible to go a few steps further. In fact, in **DTK** it is possible to obtain the following partial result (see Corradini and Galvan, 2022, step 8′, p. 493):

THEOREM 2. $K =_{\mathbf{PA}} F \vdash_{\mathbf{DTK}} (\exists\varphi \notin \mathrm{L}(\mathbf{PA}))(D\varphi \wedge K\varphi \wedge \neg F\varphi)$, *i.e., if there is a formalism F which coincides with K with respect to mathematical propositions, there is at least one not purely mathematical determinate proposition that is humanly knowable and not derivable in F.*

[1] He says in (Koellner, 2018b, p. 468): "Assume that **DTK** is correct for arithmetical statements. Then [. . .] from the point of view of **DTK**, it is in principle impossible to prove or refute either disjunct."

Clearly, both theorems are restricted to determinate formulas. In particular, the assumption of identity between K and F that appears in the antecedent in Theorem 2 is explicitly restricted to arithmetical formulas, but the non-arithmetical formula whose existence is declared in the consequent is also determinate, and is proved to be so. The meaning of the theorem is subtle: it declares that, even assuming that there is a formalism F which coincides with K with respect to mathematical propositions — i.e., which is able to match the human mind in this respect — there is at least one *not purely* mathematical proposition that is not derivable in F — and therefore inaccessible to it. In other words, it is impossible for a formalism to equal the human mind in both mathematical and non-mathematical knowledge. It may equal it mathematically, but in that case it cannot equal it in other respects. In this sense, then, the difference between mind and machine is demonstrated.

Theorem 2 seems to contradict Theorem 1. But this is a false impression. Theorem 1 rejects the possibility of proving that the human mind cannot be equalled by a formalism as far as mathematical propositions are concerned; that is, it excludes the provability of the first restricted disjunct. Theorem 2, instead, states that, if there exists a formalism F equivalent to the human mind from a mathematical point of view, then there exists at least one non-mathematical proposition that is humanly knowable but not derivable in F. This means that formalism cannot match the human mind in both respects: if it equals it from a mathematical point of view, it cannot equal it from a non-purely mathematical point of view. Therefore, Theorem 2 excludes that there can be a formalism equivalent to the mind and consequently seems to prove the first disjunct in contrast to Theorem 1.

However, things are not like that. It is true that Theorem 2 excludes that the mind is equivalent to a machine, but it excludes it because equivalence is understood in a broader sense than that of Koellner. The equivalence between mind and machine is understood in terms of their ability to derive *both* mathematical and non-purely mathematical determinate propositions. That the first disjunct can be understood according to this meaning of the notion of equivalence is not senseless, since the equivalence between mind and machine must not be exclusively ascribed to the ability to derive mathematical propositions. But then, if the first disjunct is understood in the sense that there is no formalism capable of deriving all mathematical *or* non-purely mathematical determinate propositions knowable by the human mind, it is reasonable to propose expressing it formally and proving it in DTK. The following steps serve precisely this purpose.

The first idea is to exploit the following two facts:

1. $K = F \vdash_{\mathbf{DTK}} K =_{\mathbf{PA}} F$
2. $(\exists\varphi \notin \mathrm{L}(\mathbf{PA}))(K\varphi \wedge \neg F\varphi) \vdash_{\mathbf{DTK}} \exists\varphi(K\varphi \wedge \neg F\varphi)$

Starting from Theorem 2. we can then obtain:

$K = F \vdash_{\mathbf{DTK}} (\exists\varphi \notin \mathrm{L}(\mathbf{PA}))(K\varphi \wedge \neg F\varphi)$	by chain
$K = F \vdash_{\mathbf{DTK}} K \neq F$	from def. of $K = F$
$\vdash_{\mathbf{DTK}} \forall F(K \neq F)$	by contradiction
$\vdash_{\mathbf{DTK}} \neg\exists F(K = F)$	logic

Clearly, we went thus far as to falsify that the mind can be a machine, i.e., the first unrestricted disjunct. Unfortunately, however, we cannot prove in **DTK** that this result is true. The unrestricted proposition is in fact indeterminate and, as a consequence, it is not possible to apply the K-intro rule to it and thereby derive the truth.

However there is another way to give a different but correct form to the statement that the mind is not a machine. That road is to stick to the formulation restricted to mathematical propositions and work on the double assumption $K =_{\mathbf{PA}} F$ and $K \neq_{\mathbf{PA}} F$. This is the content of the following theorem:

THEOREM 3. $\vdash_{\mathbf{DTK}} (\exists\varphi \notin \mathrm{L}(\mathbf{PA}))(D\varphi\wedge K\varphi\wedge\neg F\varphi)\vee(\exists\varphi \in \mathrm{L}(\mathbf{PA}))(K\varphi\wedge \neg F\varphi) \vee (\exists\varphi \in \mathrm{L}(\mathbf{PA})(\neg K\varphi \wedge F\varphi)$.

PROOF. 1. Part: $K =_{\mathbf{PA}} F$

$K =_{\mathbf{PA}} F \vdash_{\mathbf{DTK}} (\exists\varphi \notin \mathrm{L}(\mathbf{PA}))(D\varphi\wedge K\varphi\wedge\neg F\varphi)\vee(\exists\varphi \in \mathrm{L}(\mathbf{PA}))(K\varphi\wedge \neg F\varphi) \vee (\exists\varphi \in \mathrm{L}(\mathbf{PA})) \wedge \neg K\varphi \wedge F\varphi)$ (from Theorem 2).

2. Part: $K \neq_{\mathbf{PA}} F$

$K \neq_{\mathbf{PA}} F \vdash_{\mathbf{DTK}} (\exists\varphi \in \mathrm{L}(\mathbf{PA}))(K\varphi \wedge \neg F\varphi) \vee (\exists\varphi(\varphi \in \mathrm{L}(\mathbf{PA}))(\neg K\varphi \wedge F\varphi)$ (def. $K =_{\mathbf{PA}} F$),
$K \neq_{\mathbf{PA}} F \vdash_{\mathbf{DTK}} (\exists\varphi \notin \mathrm{L}(\mathbf{PA}))(D\varphi\wedge K\varphi\wedge\neg F\varphi)\vee(\exists\varphi \in \mathrm{L}(\mathbf{PA}))(K\varphi\wedge \neg F\varphi) \vee (\exists\varphi \in \mathrm{L}(\mathbf{PA}))(\neg K\varphi \wedge F\varphi)$ (logic).

Conclusion:

$\vdash_{\mathbf{DTK}} (\exists\varphi \notin \mathrm{L}(\mathbf{PA}))(D\varphi \wedge K\varphi \wedge \neg F\varphi) \vee (\exists\varphi \in \mathrm{L}(\mathbf{PA}))(K\varphi \wedge \neg F\varphi) \vee (\exists\varphi \in \mathrm{L}(\mathbf{PA}))(\neg K\varphi \wedge F\varphi)$ (logic). □

Manifestly, the three admitted possibilities exclude the possibility of the mind being a machine. If mind and machine do not differ for mathematical propositions, they do differ for propositions that are not purely mathematical. Furthermore, the formulation is restricted to determinate formulas and the entire disjunction is itself determinate. In this sense, we have gained the formulation of the first disjunct we wanted.

3. No Identity $\neq$ Superiority

The explication of the result achieved, however, makes it clear that the difference between mind and machine does not nevertheless imply the mathematical superiority of the mind over the machine.[2] In fact, Theorem 3 contemplates the possibility that the machine is superior to the mind in mathematical propositions. It is possible that there exists a machine capable of deriving mathematical propositions inaccessible to the human mind. Far from disproving the difference result, this possibility reaffirms it. Indeed, as the following corollary shows, if there is a formalism that overcomes the human mind in mathematical knowledge, it is impossible to know whether this formalism is consistent. Let us see the proof of this important corollary.

COROLLARY. $(\exists \varphi \in \mathrm{L}(\mathbf{PA}))(\neg K\varphi \wedge F\varphi) \vdash_{\mathbf{DTK}} \neg K(\mathrm{Cons}_F)$

PROOF.

	$\vdash_{\mathbf{DTK}} F\varphi \to \mathrm{PR}_{\mathbf{PA}}(F\varphi)$	by Σ_1−Completeness
	$\vdash_{\mathbf{DTK}} \mathrm{PR}_{\mathbf{PA}}(F\varphi) \to K(F\varphi)$	from Rule of mathematical knowledge
(*)	$\vdash_{\mathbf{DTK}} F\varphi \to K(F\varphi)$	logic
(**)	$\vdash_{\mathbf{DTK}} \varphi \in \mathrm{L}(\mathbf{PA}) \to K(\varphi \in \mathrm{L}(\mathbf{PA}))$	idem
	$\vdash_{\mathbf{DTK}} \mathrm{Cons}_F \to (\forall \varphi \in \mathrm{L}(\mathbf{PA}))(F\varphi \to \varphi)$	Cons implies Sound

Now, $\mathrm{Cons}_F \to (\forall \varphi \in \mathrm{L}(\mathbf{PA}))(F\varphi \to \varphi)$ is determinate. In fact Cons_F is determinate since it is an arithmetical proposition. Moreover, $F\varphi \to \varphi$ is itself determinate since $F\varphi$ and φ are both arithmetical and therefore determinate. That is why it is legitimate to apply K-intro.

$\vdash_{\mathbf{DTK}} K(\mathrm{Cons}_F \to (\forall \varphi \in \mathrm{L}(\mathbf{PA}))(F\varphi \to \varphi))$ K-intro

[2] The notion of *superiority of the mind over the machine* is very complex. For a more detailed discussion see (Shapiro, 2016, pp. 192–200). For us, the notion of *superiority* simply stems from the negation of the introduced notion of identity between mind and machine.

$\vdash_{\mathbf{DTK}} K(\mathrm{Cons}_F) \to K((\forall\varphi \in \mathrm{L}(\mathbf{PA})(F\varphi \to \varphi)))$ Distributivity of K

$\vdash_{\mathbf{DTK}} K(\mathrm{Cons}_F) \to \forall\varphi(K(\varphi \in \mathrm{L}(\mathbf{PA})) \to K(F\varphi \to \varphi))$ Distributivity of K

$\vdash_{\mathbf{DTK}} K(\mathrm{Cons}_F) \to \forall\varphi(K(\varphi \in \mathrm{L}(\mathbf{PA})) \to (KF\varphi \to K\varphi))$ Distributivity of K

$\vdash_{\mathbf{DTK}} K(\mathrm{Cons}_F) \to (\forall\varphi \in \mathrm{L}(\mathbf{PA}))(F\varphi \to K\varphi)$ by $(*)$ and $(**)$

$(\exists\varphi \in \mathrm{L}(\mathbf{PA}))(\neg K\varphi \wedge F\varphi) \vdash_{\mathbf{DTK}} \neg K(\mathrm{Cons}_F)$ by contraposition □

Conclusion

In conclusion we have that for all F it is impossible that $F = K$; that is, if $F = K$ as far as mathematical formulas are concerned, then F differs from K in some propositions that are not purely mathematical. It is possible, however, that K surpasses F as far as mathematical propositions are concerned but also, on the contrary, that F surpasses K in the derivation of mathematical propositions not accessible to human knowledge. In the latter case, however, the consistency of F is humanly undecidable. This fact is important from two points of view.

First of all, it is important since it cannot be ruled out that the unknowability of the consistency of F is due to the fact that F is in fact inconsistent. Therefore, this proposition — as well as the consistency of F — does not belong to the set of propositions mentioned in Gödel's second disjunct. These are indeed humanly undecidable, but true. Of the proposition provable in F but not humanly knowable, we cannot say that it is true, because we do not know that F is consistent. In any case, the fact that the first disjunct is — according to Theorem 3 — true affects the whole disjunction. It implies that the second disjunct can be true or false, but not necessarily true. If it is false, then there is no true proposition that escapes human knowledge. In that case the above formalism F would be inconsistent. For, if it were consistent — i.e., if it were true that it is consistent — then it would be known and consequently there could be no proposition derivable in F that is not knowable — i.e., F could not exceed the capacity of the human mind. If, on the other hand, it is true — and so there would exist propositions that are true but humanly absolutely undecidable — then the case of F being superior to K could arise, because among the humanly unknowable true propositions could also be the proposition declaring the consistency of F.

Second, that the machine cannot equal the human mind in the knowledge of non-purely mathematical truths — although it can equal it in specifically mathematical knowledge — is a clear evidence of the difference between mind and machine. Further evidence of this difference is the fact that, if the machine surpasses the human mind in mathematical knowledge — a case equally possible on the basis of Theorem 3 — its consistency is not knowable. Indeed, diversity is to be understood not only in terms of the greater/smaller amount of mathematical truth attainable by the machine or the mind, but also in terms of the different way in which mathematical truth is accessed. Even if, as follows from Theorem 3, it is not excluded that the mathematical goals attainable by the human mind can be surpassed by a machine, it is important that the mind's way of proceeding is different from that of the machine. For, even assuming for the sake of argument that humans, unlike machines, have the ability to understand, manipulate and work with abstract objects — as Gödel and Kreisel believe according to Shapiro (2016, p. 205) — the mind still does not have the capacity to access the consistency of the machine that surpasses it in mathematical knowledge. This means that the machine is so different from the mind that the latter cannot recognise the way the former identifies the axioms. The mind knows the recursive way in which the machine proceeds, but it does not recognise the way in which it establishes the starting axiomatic points. The way in which the machine determines the axioms is inaccessible to the mind. The corollary to Theorem 3 thus reinforces the meaning of Theorem 3. While Theorem 3 explains the difference between the mind and the machine in terms of the quantity and type of mathematical or non-mathematical propositions that can be known by the mind or derived from the formalism, the corollary explains a further element of difference in the fact that the ways that the mind and the machine operate are profoundly different.

References

Corradini A., and S. Galvan, 2022, "Analysis of Penrose's second argument formalised in DTK system", *Logic and Logical Philosophy*, 31: 471-500. DOI: 10.12775/LLP.2021.019

Feferman, S., 2008, "Axioms for determinatess and truth", *The Journal of Symbolic Logic*, 1(2), 204–217. DOI: 10.1017/S1755020308080209

Gödel, K., 1995, "Some basic theorems on the foundations of mathematics and theirimplications" (1951), pages 304–323 in S. Feferman et al. (eds.), *Collected Works, Volume III: Unpublished Essays and Lectures*, New York: Oxford University Press.

Koellner, P., 2016, "Gödel's disjunction", pages 148–188 in L. Horsten and P. Welch (eds.), *Gödel's Disjunction: The Scope and Limits of Mathematical Knowledge*. New York: Oxford University Press. DOI: 10.1093/acprof:oso/9780198759591.003.0007

Koellner, P., 2018a, "On the question of whether the mind can be mechanized, I: From Gödel to Penrose", *The Journal of Philosophy*, CXV(7): 337–360. DOI: 10.5840/jphil2018115721

Koellner, P., 2018b, "On the question of whether the mind can be mechanized, II: Penrose's New Argument", *The Journal of Philosophy*, CXV(7): 453–484. DOI: 10.5840/jphil2018115926

Penrose, R., 1994, *Shadows of the Mind: A Search for the Missing Science of Consciousness*, Oxford University Press. DOI: 10.1093/oso/9780195106466.001.0001

Penrose, R., 1996, "Beyond the doubting of a shadow. A reply to commentaries on *Shadows of the Mind*". https://calculemus.org/MathUniversalis/NS/10/01penrose.html

Shapiro, S., 2016, "Idealization, mechanism, and knowability", pages 189–207 in L. Horsten and P. Welch (eds.), *Gödel's Disjunction: The Scope and Limits of Mathematical Knowledge*, New York: Oxford University Press. DOI: 10.1093/acprof:oso/9780198759591.003.0008

ANTONELLA CORRADINI
Department of Psychology
Università Cattolica del Sacro Cuore Milano, Italy
antonella.corradini@unicatt.it
https://orcid.org/0000-0001-8227-2226

SERGIO GALVAN
Department of Philosophy
Università Cattolica del Sacro Cuore Milano, Italy
sergio.galvan@unicatt.it
https://orcid.org/0000-0001-8552-1099