Experimental Logic as a Model of Developement of Mathematical Knowledge

Michal Tomasz Godziszewski [1]

Abstract. We answer the question of computational reasons for epistemic hardness of certain class of philosophically interesting mathematical concepts. We justify the statement that mathematical knowability may be identified with algorithmic learnability. We present framework of experimental logics equivalent to the notion of learnability. Then we prove the main result. By adjoining the minimal possible set of undecidable sentences to recursive axiomatization of arithmetics and closing it under logical consequence, we obtain a non-learnable theory. This gives an explanation to the fact that undecidable arithmetical sentences are cognitively difficult. We conclude that cognitively accessible mathematical concepts are exactly within the scope of learnablity.

1 Knowability as algorithmic learnability

Emergence and development of recursion theory and computer science enable us to rigorously address the question of characterising the class of mathematical concepts that are cognitively accessible to computational devices such as human minds. The answer to this question would give reasons for which some concepts are epistemically easy (e.g. provable within first-order theories or possessing certain combinatorial properties) and the others are cognitively hard for the human mind.

Our explication is based on the assumption that human mind is a computing device. What we mean by this is that functions computable by human beings are exactly Turing-computable. In fact, we assume that human mind does not exhibit any non-recursive behaviour. A short reflection should convince us that people can perform any computation, provided sufficient amount of time and space. This leads us to the second assumption that cognitively accessible world is potentially infinite. In general, computations are unbounded with respect to required resources, like time and space. The latter are provided by the actual world. We can think of the world as if it was finite. However, we can always somehow finitely extend the actual world to fulfill our computational requirements.

Suppose we want to cope with the problem $P = \{x : \exists y R(x, y)\}$, where R is recursive. We may approach any instance „$a \in P$?” in the following way. Set answer to *no*. Start generating elements from the universe. For each generated element b check whether $R(a, b)$ and if so set answer to *yes* and stop; otherwise continue. This algorithm ensures that positive answers establish with certainty. Negative answers are subjected to uncertainty - there is no guarantee that there is no witness for the existential quantifier. Nevertheless, it is still a good cognitive strategy to rely on such algorithm. Justification comes from the work of mathematicians. Axiomatic method has been successfully used since Euclid of Alexandria. Nowadays, it is well recognized that the set of theorems of an axiomatic system with recursive set of axioms is recursively enumerable. Therefore, the algorithm described above applies to the work of „determining” the set of theorems of the particular axiomatic system. Intuitively, knowledge obtained through axiomatic systems by mathematicians seems fully legitimate. The reason *why* it is fully legitimate is that finding a proof may be difficult and take a long time, but once a proof is found, the answer is recursively conclusive. Hence, it seems theoretically justified why, for instance, theorems of axiomatic number theory or axiomatic set theory are cognitively accessible.

Observe that the above algorithm for „determining” P has the property that at some unspecified time of the computation the answer may change from *no* to *yes*. In general, it is impossible to recursively predict the moment after which the answer will change. Since if it was possible, we would easily construct a decision procedure for P. It turns out we accepted as cognitively sound a method that allows one mind-change and captures Σ^0_1 sets. This clearly shows, that decidability is too narrow concept to fit our purposes.

Consider the following method for „determining” the consistency of the theory of axiomatic system with recursive set of axioms: $T := \emptyset$. Set answer to *yes*. Inside infinite loop do the following. If T contains contradiction, answer *no* and stop. Otherwise generate next proof, add proved sentence to T and continue the loop. In this situation we can eventually arrive at conclusive answer if axiomatic system is inconsistent. But if the system is consistent, we are left uncertain. Similar procedure is easily applicable to any problem of the form $\{x : \forall y R(x, y)\}$, where R is recursive.

If we accepted as cognitively sound a method that proceeds by one mind change from *no* to *yes*, then we should also accept a method that allows one mind change from *yes* to *no*. However, the sets captured by the former kind of method are Σ^0_1, whereas the sets caputed by the latter kind of method are Π^0_1. Since $\Sigma^0_1 - \Pi^0_1$ and $\Pi^0_1 - \Sigma^0_1$ are both non-empty, neither of these two kinds of methods is adequate for explaining knowability. We would need something stronger to capture both these classes.

We can see, that the common property of these two kind of methods is that on every input after some finite time they level off on the right answer. Going further, it seems justified to accept as cognitively sound any method that proceeds by mind-changes and on every input stabilizes. In this way we arrive at the concept of algorithmic learnability

Definition 1 Let $A \subseteq N$. Say that A is algorithmically learnable iff there is a total computable function $g : N^2 \to \{0, 1\}$ such that for all $x \in N$: $\lim_{n\to\infty} g(x, t) = 1 \Leftrightarrow x \in A$ and $\lim_{n\to\infty} g(x, t) = 0 \Leftrightarrow x \notin A$.

The notion of algorithmic learnability is one of the equivalent formu-

[1] University of Warsaw, Poland, email: mtgodziszewski@gmail.com

lations of the concept of methods proceeding by mind-changes and stabilizing on every input.

2 Experimental logics

Following Jeroslow from [7], we identify the mechanistic conception of a theory which proceeds by trial-and-error with a recursive predicate $H(t, x, y)$ of three variables interpreted intuitively as follows: *At time t, the finite configuration with Gödel number y is accepted as a justification of the formula with Gödel number x.*

Definition 2 Given an experimental logic $H = H(t, x, y)$ we identify the **theorems** of H with **recurring formulae** defined by:

$$Rec_H(x) \equiv \forall t \exists s \geq t \exists y H(s, x, y).$$

Stable formulae of H are defined as follows:

$$Stbl_H(x) \equiv \exists t \exists y \forall s \geq t H(s, x, y).$$

We say that H is **convergent** if all recurring formulae are stable. Since the implication for the other direction is obvious by the predicate calculus, we may identify H being convergent with the following equivalence: $\forall x\, (Rec_H(x) \equiv Stbl_H(x))$.

Theorem 1 *The sets of theorems of convergent, experimental logics are precisely the Δ^0_2 sets.*

By the Shoenfield's Limit Lemma a set A is Δ^0_2 if and only if there exists a function $f : N^2 \to \{0, 1\}$ such that:

$$\forall x(x \in A \Leftrightarrow \exists t \forall s \geq t \;\; f(x, s) = 1),$$

$$\forall x(x \notin A \Leftrightarrow \exists t \forall s \geq t \;\; f(x, s) = 0).$$

Therefore we may take $H(t, x, y) := f(x, t) = y \wedge y = 1$. Then we have $Stbl_H(x) \equiv \exists y \exists t \forall s \geq t\, (f(x, t) = y \wedge y = 1)$.

This result is crucial, since it means that theorems of convergent, experimental logic are exactly algorithmically learnable (and simultaneously exactly meaningfully representable within finitstic mathematical means).

Our next theorem extends Gödel's incompleteness theorem in terms of intrinsic limitations of experimental logics. From now on, by PA, we denote only the set of axioms of PA which is not to be confused with the set of logical consequences of PA, from now on, denoted by $Cn(PA)$.

Theorem 2 (Jeroslow [7])
Let H be a consistent, convergent, experimental logic whose theorems contain those of first-order Peano arithmetic and whose theorems are closed under first-order predicate reasoning. Then there is a true Π^0_1 sentence that is not provable in H.

First of all, let us notice that if $\exists x \forall y \psi(x, y)$ is a true, but unprovable Σ^0_2 sentence, then for some n we have that $\forall y \psi(n, y)$ is true but unprovable Π^0_1 sentence.

By the diagonal lemma, we can easily obtain a formula φ such that:

$$\vdash Rec(\varphi) \equiv \neg\varphi. \tag{1}$$

We can see that φ is equivalent to a Σ^0_2 sentence. There are now two possibilities:

1. $\vdash Rec(\varphi) \Rightarrow \varphi$.
2. $\nvdash Rec(\varphi) \Rightarrow \varphi$.

Let us consider case 1 first. Since by Equation 1 we obatained that $\vdash Rec(\varphi) \Rightarrow \neg\varphi$, by our assumption we get $\vdash \neg Rec(\varphi)$, and by Equation 1 again we get that $\vdash \varphi$. Therefore $Stbl(\varphi)$ is a true Σ^0_2 sentence. It suffices to show that $Stbl(\varphi)$ is not provable. For the sake of contradiction, suppose $\vdash Stbl(\varphi)$. This obviously means that $\vdash Rec(\varphi)$ and from this it follow that H is inconsistent, contrary to our general assumption.

Now let us proceed with case 2. It now suffices to show that $Rec(\varphi) \Rightarrow \varphi$ is true since by its construction and assumption of our case, it is an unprovable Σ^0_2 sentence. Suppose $Rec(\varphi)$ is true. Since H is convergent, $Stbl(\varphi)$ is then true as well. Hence, we have $\vdash \varphi$. But then obviously $\vdash Rec(\varphi) \Rightarrow \varphi$, contradicting the case. Thus, $Rec(\varphi)$ is false and by trivial propositional calculus $Rec(\varphi) \Rightarrow \varphi$ is true.

From this theorem we have an immediate, but extremely important corollary:

Corollary 1 *The deductive closure of $PA + \{\varphi \in \Pi^0_1 - Sent_{\mathcal{L}} : N \models \varphi\}$ is not Δ^0_2.*[2]

If such a theory was Δ^0_2, it would be a convergent experimental logic and as such it would not contain some true Π^0_1 sentence, but it does contain all of them by the definition, which would be inconsistent.

3 Main results - Learnability and arithmetical incompleteness

We are working under the assumption that the theories: PA, $PA + Con(PA)$ and $PA + \neg Con(PA)$ are consistent.

Definition 3 Let us define the following sets of (codes of, i.e. Gödel numbers of) arithmetical sentences:

1. $X := \{\varphi \in \Pi^0_1 : \; PA + Con(PA) \vdash \varphi \;\; and \;\; PA \nvdash \varphi\}$.
2. $Y := \{\varphi \in \Pi^0_1 : \; PA \nvdash \varphi \;\; and \;\; PA \nvdash \neg\varphi\}$.
3. $Z := \{\varphi \in \Pi^0_1 : \; N \models \varphi\}$.

For convenience, we will omit the corner notations - the Reader is asked only to remember that while speaking of X, Y and Z, we are dealing with sets of natural numbers.

Theorem 3 $X \subset Y \subset Z$.

$(X \subset Y)$
Let us take any $\varphi \in X$. By assumption, we have $PA \nvdash \varphi$. For the sake of contradiction suppose $PA \vdash \neg\varphi$. But then, obviously $PA + Con(PA) \vdash \neg\varphi$. But this means that $PA + Con(PA)$ is inconsistent, which is inconsistent with out general assumption. Now we will show that the inclusion $X \subseteq Y$ is proper. By the diagonal lemma, there is a sentence $\psi \in Sent_{\mathcal{L}}$ such that:

$$PA + Con(PA) \vdash \; \psi \equiv \neg Pr_{PA+Con(PA)}(\overline{\psi}).$$

Obviously $\psi \in \Pi^0_1$. Therefore by the proof of Gödel's theorem we have that $PA + Con(PA) \nvdash \psi$. Then, obviously $PA \nvdash \psi$. On the other hand ψ is true, i.e. $N \models \psi$, therefore $PA \nvdash \neg\psi$. This means $\psi \in Y$ and $\psi \notin X$.

$(Y \subset Z)$
Let us take any $\varphi \in Y$. For the sake of contradiction, suppose $N \not\models \varphi$. Then, by the definition of satisfactian (Tarskian classical

[2] instead of this we can denote it more easily: $Cn(PA + \{\varphi \in \Pi^0_1 : N \models \varphi\})$ is not Δ^0_2

semantics) $N \models \neg\varphi$. However, $\neg\varphi \in \Sigma^0_1$. By Σ^0_1-completeness of PA we then obtain $PA \vdash \neg\varphi$ which is inconsistent with $\varphi \in Y$. The inclusion is proper, since every Π^0_1-sentence φ such that $PA \vdash \varphi$ is in Z, but not in Y, by the definition of both of them.

Lemma 1 *$PA + \neg Con(PA) \vdash \varphi$ is equivalent to $PA + \neg\varphi \vdash Con(PA)$.*

The statement of the lemma follows easily form the following sequence of equivalent statements:

1. $PA + \neg Con(PA) \vdash \varphi$.
2. For any model $\mathcal{M}$ if $\mathcal{M} \models (PA + \neg Con(PA))$, then $\mathcal{M} \models \varphi$.
3. For any model $\mathcal{M}$ if $\mathcal{M} \not\models \varphi$, then $\mathcal{M} \not\models (PA + \neg Con(PA))$.
4. For any model $\mathcal{M}$ if $\mathcal{M} \models \neg\varphi$, then $\mathcal{M} \not\models PA$ or $\mathcal{M} \models Con(PA))$.
5. For any model $\mathcal{M}$ if $\mathcal{M} \models \neg\varphi$ and $\mathcal{M} \models PA$, then $\mathcal{M} \models Con(PA))$.
6. $PA + \neg\varphi \vdash Con(PA)$.

Lemma 2 *$PA + \neg Con(PA)$ is Π^0_1-conservative over PA, i.e. for any arithmetical sentence $\varphi \in \Pi^0_1$ $PA + \neg Con(PA) \vdash \varphi$ if and only if $PA \vdash \varphi$.*

($\Leftarrow$) - obvious.

($\Rightarrow$) Let us assume that $PA + \neg Con(PA) \vdash \varphi$. From the previousl lemma $PA + \neg Con(PA) \vdash \varphi$ is equivalent to $PA + \neg\varphi \vdash Con(PA)$. But $\neg\varphi \in \Sigma^0_1$, and for any recursive extension of PA we have provable Σ^0_1-completeness, i.e. for any $\psi \in \Sigma^0_1$ and any T - recursive extension of PA we have: $T \vdash \psi \Rightarrow Pr_{PA}(\overline{\psi})$. We therefore have:

$$PA + \neg\varphi \vdash \neg\varphi \Rightarrow Pr_{PA}(\overline{\neg\varphi}).$$

But of course $PA + \neg\varphi \vdash \neg\varphi$. Hence,

$$PA + \neg\varphi \vdash Pr_{PA}(\overline{\neg\varphi}).$$

This and the fact that $PA + \neg\varphi \vdash Con(PA)$ give us $PA + \neg\varphi \vdash Con(PA + \neg\varphi)$. From the second Gödel's incompleteness theorem we obtain that $\neg Con(PA + \neg\varphi)$ which is equivalent to $PA \vdash \varphi$, which ends the proof.

Theorem 4 *The set of all Π^0_1-sentences which are unprovable in PA is many-one reducible to the set X.*

Let us define an arithmetical function $f : \omega \to \omega$ such that

$$f(\varphi) = Con(PA) \vee \varphi.$$

We will show that

$$f(\varphi) \in X \Longleftrightarrow PA \nvdash \varphi.$$

Obviously, for any sentence φ we have $PA + Con(PA) \vdash Con(PA) \vee \varphi$. Hence, by the definition of X, $f(\varphi) \in X$ if and only if $PA \nvdash Con(PA) \vee \varphi$, which is equivalent to $PA + \neg Con(PA) \nvdash \varphi$. By the previous lemma this is equivalent to $PA \nvdash \varphi$. This ends the proof.

Corollary 2 *The set X is Π^0_1-hard.*

Let $W = \{\varphi \in \Pi^0_1 : \; PA \nvdash \varphi\}$. From the theorem above we know that $W \leq_m X$. But the set W is Π^0_1-complete - it is defined by the Π^0_1-relation, i.e.

$$\forall x \in \omega \, (x \in W \Leftrightarrow (x \in \Pi^0_1 \wedge \forall y \neg Prov(y, x))).$$

This is a Π^0_1-relation since the set of Π^0_1-sentences has its own truth definition, as we proved. It is Π^0_1-complete because its complement - the set of sentences not being Π^0_1 or provable in PA is trivially Σ^0_1-complete.[3]

Theorem 5 $Cn(PA + Con(PA)) = Cn(PA + X)$

($\subseteq$) Let φ be such that $PA + Con(PA) \vdash \varphi$. Obviously $Con(PA) \in X$, therefore trivially $PA + X \vdash \varphi$.

($\supseteq$) Let φ be such that $PA+X \vdash \varphi$. Since this is a first-order theory, by completeness and compactness we can infer that in the proof of φ from $PA+X$ we use finitely many formulae, namely: $\phi_1, \phi_2, \ldots \phi_n$. All of them either belong to PA or belong to X or can be inferred from $PA + X$. In particular they are implied by $PA + Con(PA)$. If so, they can be used in the proof of φ form $PA + Con(PA)$, so $PA + Con(PA) \vdash \varphi$.

Corollary 3 *The set $Cn(PA + X)$ is Δ^0_2 (and as such: algorithmically learnable).*

Since $(PA + Con(PA))$ is a recursive extension of PA, it is a recursively enumerable set, i.e. Σ^0_1. By the fact that it is identical with the set $Cn(PA + X)$, the latter one also must be recursively enumerable, and in particular: algorithmically learnable.

High complexity of X comes from excluding certain sentences - namely those sentences that are provable in PA. But adding PA and then closing under consequence restores those sentences. That is why the complexity decreases. It is not very suprising that Cn operator can decrease the complexity of a set of sentences - we can always add a negation of a sentence of any given set to obtain an inconsistent theory which will be (primitive) recursive. The above is however a very nice example of how Cn can decrease the complexity of a given theory to something higher than just a set whose characteristic function is primitive recursive.

We have shown that although the complexity of the set X of the (Gödel numbers of) Π^0_1-sentences unprovable in PA but provable in

[3] Another way to see that X is Π^0_1-hard - explicitly using diagonalization - would be as follows (the argument below is a quotation of E. Jerabek - a proof given in the communication via Internet, see: www.mathoverflow.net/questions/63690):
Let $\sigma(x) = \exists v \, \theta(x, v)$ be a complete Σ^0_1-formula (such that it is not equivalent to any Δ^0_0-formula, where $\theta \in \Delta^0_0$, and find a formula $\pi(x)$ such that PA proves

$$\pi(x) \equiv \forall w \, (\mathrm{Prov_{PA}}(w, \pi(\dot{x})) \Rightarrow \exists v \leq w \, \theta(x, v))$$

by the diagonal lemma. Let $n \in \omega$. Since $\neg\pi(\bar{n})$ is equivalent to a Σ^0_1 sentence, PA proves $\neg\pi(\bar{n}) \Rightarrow \mathrm{Pr_{PA}}(\neg\pi(\bar{n}))$. By definition, $\neg\pi(\bar{n}) \Rightarrow \mathrm{Pr_{PA}}(\pi(\bar{n}))$, hence PA proves $\mathrm{Con_{PA}} \Rightarrow \pi(\bar{n})$. We claim that

$$(*) \quad N \models \sigma(n) \Longleftrightarrow \mathrm{PA} \vdash \pi(\bar{n}),$$

which means that $n \mapsto \pi(\bar{n})$ is a reduction of the Π^0_1-complete set $\{n : N \models \neg\sigma(n)\}$ to X.
To show $(*)$, assume first that $\mathcal{M} \models \mathrm{PA} + \neg\pi(\bar{n})$. Then there is no standard PA-proof of $\pi(\bar{n})$, hence the witness $w \in \mathcal{M}$ to the leading existential quantifier of $\neg\pi(\bar{n})$ must be nonstandard. Then $\neg\theta(n, v)$ holds for all $v \leq w$, and in particular, for all standard v, hence $N \models \neg\sigma(\bar{n})$.
On the other hand, assume that PA proves $\pi(\bar{n})$, and let k be the code of its proof. Since PA is sound, $N \models \pi(\bar{n})$, hence there exists $v \leq k$ witnessing $\theta(\bar{n}, v)$, i.e. $N \models \sigma(\bar{n})$, which ends the proof.

$PA + Con(PA)$ is high, the set $Cn(PA + X)$ is learnable, i.e. *easy* in terms of computational cognitive capacties. Jeroslow showed that the set $Cn(PA + Z)$ is not learnable. However, the set Z of all true Π^0_1-sentences seems to be very *big* - it contains a very large number of sentences and adjoining it to PA and closing under consequence also results in a complicated theory not very surprisingly. So a question rises: is there a way to improve Jeroslow's result by adjoining a *smaller* set to axioms of PA? The answer is YES and the set adjoined to the axioms of PA that results in a non-learnable theory after closing it under logical consequence is of particular epistemological interest - we can achieve epistemically hard, non-learnable theory by enriching PA with the set of Π^0_1-sentences undecidable in PA, namely the set: Y defined above.

Theorem 6 *Cn(PA+Y) = Cn(PA+Z)*

($\subseteq$) Let $\varphi \in Cn(PA + Y)$. Without loss of generality, assume $PA \nvdash \varphi$. Then, in the proof of φ from $PA + Y$ there occurs a finite number of sentences that are consequences of PA and a finite number of undecidable Π^0_1-sentences. But any undecidable Π^0_1-sentence is in Z, since if it was not, it would have to be a false Π^0_1-sentence, yet its negation would be a true Σ^0_1-sentence. By Σ^0_1-completeness of PA the latter would be provable and the theory would be inconsistent, contrary to our assumption. Therefore φ is also provable from $PA + Z$, which means $\varphi \in Cn(PA + Z)$.
($\supseteq$) Let $\varphi \in Cn(PA + Z)$. Without loss of generality, assume $PA \nvdash \varphi$. Then, in the proof of φ from $PA + Y$ there occurs a finite number of sentences that are consequences of PA and a finite number of true, but unprovable Π^0_1-sentences. But such sentences are in Y, therefore φ is also provable from $PA + Y$, which means $\varphi \in Cn(PA + Y)$.

Corollary 4 *The set $Cn(PA + Y)$ is not Δ^0_2.*

Immediate, by the fact that $Cn(PA + Z)$ is not Δ^0_2.
We may sum up this result in more philosophically plausible terms:

Corollary 5 Undecidable sentences of arithmetical theories (recursively) extending PA are not algorithmically learnable.

4 Conclusions and Final Remarks

Experimental logics framework, being in accordance with the trial-and-error learning concept, seems to be a good explication of the process of acquiring the content of mathematical concepts by the computational mind. While learning mathematical concepts, we conjecture some of its properties and search for justifications of our statements about them. If we accept some sequence of expressions as the justification for a given mathematical proposition in a given moment of time - e.g. a convincing example, it may happen that in view of new, empirical data we change our mind and abandon the justification we have. The situation in which we search for justifications of given conjectures and even sometimes adjust the notions we formalize (as it was convincingly shown by I. Lakatos in [9]) is formalized by the notion of recurring formula. Finding a correct notion, on the other hand, namely finding a proof, seems to be formalized by the notion of stable formula. Therefore, convergent logic is an idealization of a deductive apparatus such that justifications for our mathematical statements we find within the apparatus are always the proofs of those statements.

Within a computational view on mathematics presented in this paper, it is easily explainable, why some sentences in the language of our arithmetical theory are left independent of the theory and undecidable on its grounds - by the complexity of provability relations, adjoining the unprovable sentences to our arithmetics would provide us with a non-learnable theory. Such a theory would not be credible as set of epistemically accessible mathematical truths, since by the character of mathematical cognition we are not able to computationally *handle* such complicated sets.

ACKNOWLEDGEMENTS

We would like to thank the referees for their comments which helped improve this paper.

REFERENCES

[1] W. Craig, *On axiomatisability within a system*, **Journal of Symbolic Logic** 18 (1953), 30-32.

[2] S. Feferman, *Degrees of unsolvability associated with classes of formalized theories*, **Journal of Symbolic Logic** 22 (1957), 161-175.

[3] S. Feferman, *Arithmetization of metamathematics in a general setting*, **Fundamenta Mathematicae** 49 (1960), 35-92.

[4] M.E. Gold, *Language Identification in the Limit* **Information and Control** 10 (1967), 447-474.

[5] M.E. Gold, *Limiting Recursion*, **Journal of Symbolic Logic** 30 (1965), 28-48.

[6] K. Gödel, *On formally undecidable propositions of Principia Mathematica and related systems I*, in: J. van Heijenoort (ed.), **From Frege to Gödel. A source book in mathematical logic 1879-1931**, Harvard University Press, Cambridge MA, 1967, pp. 596-616.

[7] R. G. Jeroslow, *Experimental logics and Δ^0_2-theories*, **Journal of Philosophical Logic** 4 (1975), 253-267.

[8] R. Kaye, *Models of Peano Arithmetic*, Oxford University Press, Oxford, 1991.

[9] I. Lakatos, *Proofs and Refutations*, Cambridge University Press, Cambridge, 1976.

[10] H. Putnam, *Trial and Error Predicates and the Solution to a Problem of Mostowski*, **Journal of Symbolic Logic** 30 (1965), 49-57.

[11] J. R. Shoenfield, *On degrees of unsolvability*, **Annals of mathematics**, vol. 69, 1959.