

Information and Diversity

◇ Klaus Frovin Jørgensen,

Section for Philosophy and Science Studies, Roskilde University

The principle of the excluded middle and the axiom of choice are two ingredients of mathematics which during the last one hundred years have attracted much attention. It is often claimed that they are strongly non-constructive and problematic, also when they are considered separately. It is nevertheless an established mere fact that when they are viewed separately both of them can be given interpretations which validate them. Classical logic, thus including the disputed principle, can be interpreted into intuitionistic logic by a negative translation as developed independently by K. Gödel and G. Gentzen in the 1930s. This shows that if intuitionistic logic is consistent, then so is the classical one. On the other hand, the informal interpretation given in 1920s by A.N. Kolmogorov and A. Heyting which was the guide line when Heyting formalised intuitionistic logic validates the axiom of choice. This insight has been utilised in several more technical interpretations showing that the axiom of choice over intuitionistic

logic is fully constructive. It is, however, the case that the *combination* of the axiom of choice and the principle of the excluded middle is constructively highly problematic since they imply, for instance, full comprehension. Thus we have a situation where two *prima facie* non-constructive principles viewed separately can be given interpretations, but the principles together are non-constructive.

In this essay I will try to show that this situation is not uncommon when one tries to ascribe constructive meaning to abstract principles and laws of mathematics which at first glance seem or in a general context are non-constructive. On the other hand, it will also be shown how constructive results can be obtained *without* restricting the methodology to intuitionistic logic, which often (wrongly) is identified as *the* constructive logic.

On Foundations

A fundamental assumption behind the following is that there is a part of the mathematical corpus which is completely safe, stable and unproblematic. Much of the research within the philosophy of mathematics, the history of mathematics and the foundations of mathematics has shown this. I will not try to argue for this assumption here. Let me just note, that when mathematics is viewed over time, beginning from the very early ages, it is quite clear that a substantial body of mathematical knowledge is hardly changed. This part includes basic reasoning such as applying modus ponens and similar arguments. It also includes, at least, reasoning about the natural numbers using only restricted quantifiers; it includes simple reasoning about finite graphs and geometrical figures and many other elements of basic mathematics. It is, of course, a difficult problem to give a characterisation of what that part or kernel precisely consists of. Much energy and effort have gone into the philosophy of mathematics in arguing against skepticism in favour of the possibility of basic mathematical knowledge. Wittgenstein, for instance, was very concerned with this. But here we want to move on—we

simply assume that there exists such a part of mathematics and quickly go on to the more ambitious parts of mathematics.

Examples of such ambitious parts are the axiom of choice and the principle of the excluded middle. But examples of abstract elements and methods can be found scattered all over the mathematical landscape. One would be infinitesimals which came into mathematics with the 18th century formulation of the differential calculus. Then by the end of the 19th century they were ruled out by the ϵ - δ understanding of continuity. But then, ironically, in 1961 A. Robinson found a rigorous foundation for infinitesimals, and they were back again. Another example is the idea of using generating functions which are to be found in areas like combinatorics, number theory, probability theory, Fourier transformation, etc. The way this typically works is that one has a rather simple problem. This could be a question concerning partitions of natural numbers or how to obtain a closed expression from a recursive one, as when one wants a direct way to calculate, say, a Fibonacci number. Such problem is then embedded in a much more complex one which could be an expression within

complex analysis involving a formal series which may or may not converge. Heavy methods from analysis are then used in order to manipulate the expression and, magically, the simple problem is solved and all the complex stuff is thrown away. Such methods are magnificent, elegant and efficient. Often it even seems impossible or at least awkward to solve the problem in a straight forward and ‘simple’ way. One example of this is the proof of the prime number theorem and another rather extreme example is Wiles’ proof of Fermat’s last theorem.

Such ideal parts of mathematics really give mathematics its power and beauty; some would say that they are the essence of the mathematical method. Nevertheless, a critical mind must raise questions pertaining to their use.

In the following we will try in a more logical framework to single out several of these ideal principles of mathematics. These principles are generally non-constructive, but we will see that it all very much depends on the context the principles are considered in. We will see which of the principles and which combinations can be given a constructive interpretation relative to some specific

part of mathematics which is considered to be constructive. Here we will take Gödel’s system T as what we find reasonable to assume as ‘secure’ mathematical knowledge. System T will be defined, explained and discussed below, but let us note that it is a natural generalisation of ordinary primitive recursive arithmetic.

As there is a foundation for our analysis, namely T, there are also more abstract elements. The philosopher of mathematics adopting this view has many ideal elements within mathematics which he must take into consideration. One way to be more specific about it is to analyse mathematics in more logical terms. Phrased in these terms examples of ideal objects and methods of, perhaps, disputable nature are besides the already mentioned principle of the excluded middle and the axiom of choice, also extensionality, König’s lemma, bar induction, double negation shift – which has the form $\forall x \neg \neg A(x) \rightarrow \neg \neg \forall x A(x)$ – and other non-intuitionistic principles such as Markov’s principle/rule, the principle of independence-of-premise and so on.

We will primarily investigate – locally in the framework of typed arithmetic – the following elements of ideal mathematics:

1. *Extensionality*.

2. *Markov's principle*: If $A(x)$ is decidable for any x and if $\neg\neg\exists xA(x)$ holds then $\exists xA(x)$, i.e.

$$\forall x(A(x) \vee \neg A(x)) \wedge \neg\neg\exists xA(x) \rightarrow \exists xA(x).$$

3. *Independence-of-premise* (for certain classes of formulas). If $A \rightarrow \exists yB(y)$ holds and y is not a free variable of A then $\exists y(A \rightarrow B(y))$ holds, i.e.

$$(A \rightarrow \exists yB(y)) \rightarrow \exists y(A \rightarrow B(y)).$$

To see why, say, 3 is constructively doubtful let the following be noted. The constructive reading (which is the interpretation given by Kolmogorov and Heyting) of $A \rightarrow \exists yB(y)$ is that given a proof of A we can construct an object t and proof of the fact that t has property B , i.e. $B(t)$. The principle says that in case $A \rightarrow \exists yB(y)$ holds then in fact $\exists y(A \rightarrow B(y))$. But the constructive reading of the conclusion is that we can construct an object t such that given any proof of A we can prove t to have property B . Thus the principle says that t can be constructed independently of the proof of A , and this is not constructively evident given $A \rightarrow \exists yB(y)$.

One of our main purposes here is to sketch an analysis of 1–3 in

order to provide coherent and consistent systems that partly contain these *prima facie* non-constructive principles. Systems that nevertheless can be seen as constructively meaningful. A system is constructively meaningful when it has several important properties, say the existence property, in a crucial way. In the case at hand this means, for instance, that when an existential statement is proved in one of the theories then there is a number or a function from Gödel's system T actually realising the existential quantifier. And this realiser can be found given any proof of the statement. Thus existence in the disputed theories is shown to be based on objects coming from T which we take to be constructive.

T as Point of Departure

T is a quantifier free type theory developed by Gödel in the the 1930s. The first publication which includes a formulation of T was, however, given some 20 years later [1]. As T is typed it has a type structure. The ground type of T is type o which represent the natural numbers. Now, the types of T are generated inductively: if σ and τ are types then $\sigma \rightarrow \tau$ is also a type. Intuitively each type represents a class of objects, and for o

this is the natural numbers. Going to higher types, $\sigma \rightarrow \tau$ is the type containing functions from σ to τ . If F is a function of type $\sigma \rightarrow \tau$, then we write this as is standard in mathematics $F : \sigma \rightarrow \tau$ or sometimes $F^{\sigma \rightarrow \tau}$. Thus $F : (o \rightarrow o) \rightarrow o$ is a functional taking an arithmetical function as an argument and gives a natural number. The question now is of course: Which objects inhabit T?

The language of T is multi-sorted, meaning that each symbol is assigned a certain type. It includes a symbol for the number zero, $0 : o$ and a symbol for the successor function $S : o \rightarrow o$.¹ T has axioms concerning these objects stating for instance that $0 : o$ is the first number and that $S : o \rightarrow o$ is injective. T also has the standard axioms from the propositional logic, but in order to do, among other things, primitive recursive arithmetic it has combinators k , s and recursor R. The axioms concerning these are²

$$kxy = x \quad sxyz = xz(yz),$$

and

$$\begin{aligned} Rxy0 &= x \\ Rxy(Sz) &= y(Rxyz)z \end{aligned}$$

One can think of the equations as reduction rules: The terms on the left 'reduce' to the terms on the right,

thus defining rules for calculations. As such it is seen that the prescribed operations are step-by-step calculable and that they are not conflicting with our notion of constructivity.

By using projector k and combinator s one can introduce the λ -operator as a *defined* notion. The operator behaves in the following way: if $\lambda x.t[x]$ is a term of type $\sigma \rightarrow \tau$ and s is a term of type σ then $(\lambda x.t[x])s = t[s]$ where $t[s]$ is of type τ . With the lambda-notation it is simple to define basic arithmetical operations as for instance $+$ which is of type $o \rightarrow (o \rightarrow o)$. A definition of $+xy$, or in more usual notation $x + y$, can be given by $Rx(\lambda w.u.Sw)y$.

Actually, one can define fairly complicated functionals in T. But the more complicated functionals are always defined inductively by a chain of definitions where each step defines a new functional in terms of previously defined ones. Now, the single steps describe simple calculations. However, given a concrete well-formed closed term we cannot directly read off how many calculations we have to perform before the overall calculation is done. But, by (non-constructive) mathematical reasoning we can, of course, give bounds for any such term, but this is in our case not that inter-

esting since it is the same (non-constructive) mathematics we want to justify. However, we are given inductive rules in order to do the computation—intuitively we just do not know how many computations we will have to perform.

Therefore, when we define new procedures (or functionals) out of previously defined ones by the schemes we get a calculable functional, since the equations defining k , s and R prescribe constructive operations. Epistemologically, we consider it as a general human ability to carry out such inductive operations and consequently, we find it justified to consider Gödel's functionals as constructive.

Reductions

Given we have accepted T as constructively unproblematic the next task is to consider ideal elements which can be given an interpretation relative to T . The general idea is that given the conceptual acceptance of a certain system, what can (or maybe should) we then accept on the basis of this; in other words, Given the acceptance of some system S_0 (which in our case is T), if S_1 reduces to (is interpretable in) S_0 , then we also accept S_1 .

The concept of *reduction* or *in-*

terpretation that we work with here is the following:

S_1 reduces to S_0 , in symbols $S_1 \preceq S_0$ if

- (i) S_1 has the existence property and the disjunction property.³
- (ii) The terms realising (i) live in S_0 .

If we accept S_0 and have $S_1 \preceq S_0$ then we consequently accept S_1 on the basis of S_0 , since S_1 has various nice properties and the realisers which make this possible come from S_0 , found to be constructively unproblematic.

Specific Reductions

We will now take a look at some specific ideal elements. First extensionality. In any of the following formal theories we have equality only between objects of the lowest type—that is, equality between numbers. Functions and functionals are equal if they are equal on equal arguments. Equality in higher types is thus *defined* in terms of lower types and is therefore extensional. As the types can be fairly complicated it is not obvious that our functionals respect this defined notion of equality. However, we can introduce axioms which

claim all functionals to respect it. Let σ be type $\sigma_1 \rightarrow (\dots \rightarrow (\sigma_n \rightarrow o) \dots)$. Then z of type σ is extensional, i.e. respects equality iff

$$\forall x_1^{\sigma_1}, y_1^{\sigma_1}, \dots, x_n^{\sigma_n}, y_n^{\sigma_n} \left(\bigwedge_{i=1}^n x_i = y_i \rightarrow z\mathbf{x} = z\mathbf{y} \right),$$

where \mathbf{x} denotes x_1, \dots, x_n . Claiming full extensionality is to claim that all functionals of the theory are extensional. However, it turns out that we cannot always interpret full extensionality. In certain situations we can interpret only *weak* extensionality. This is not introduced by axioms but rather as a rule which within natural deduction has the following form:

$$\frac{\begin{array}{c} \Delta \\ \vdots \\ s = t \end{array}}{r[s] = r[t]}$$

where Δ consists only of quantifier free formulas, and s, t and $r[x]$ are terms.⁴

In the following there will be two different theories on top of which we will put ideal principles. They are two different versions of intuitionistic number theory generalised to finite types; one will have full extensionality, E-HA^ω, and the other one will only have weak extensionality, WE-HA^ω. Different elements which

are generally non-constructive will then be added to these theories. First of all, two different versions of the principle of independence-of-premise. The principle in its general form was defined on page 5. Here we will work with the case where the formula A (see page 5) does not contain any existential quantifier, nor any disjunction—this will be denoted IP. As it turns out, this principle is constructively in conflict with Markov's principle. Thus, we will have to consider a more restricted version, namely the case where the formula A is purely universal. This principle will be denoted IP_∀. Markov's principle is found on page 5. However, in our case the formulas which are decidable are the quantifier free formulas. Consequently, the form of Markov's principle (MP) which we will look at is

$$\neg\neg\exists x A_{\text{qf}}(x) \rightarrow \exists x A_{\text{qf}}(x),$$

where A_{qf} is quantifier free.

Finally, the axiom of choice, denoted AC, is also added.

Let Γ be any arbitrary but fixed set of true existence and disjunction free sentences and likewise, let Γ_{\forall} be any arbitrary but fixed set of true purely universal sentences. Then we define the following theories:

$$\begin{aligned} T_1 &:= \text{E-HA}^\omega + \text{IP} + \text{AC} + \Gamma, \\ T_2 &:= \text{WE-HA}^\omega + \text{IP}_\forall + \text{MP} \\ &\quad + \text{AC} + \Gamma_\forall. \end{aligned}$$

By modified realisability as defined by G. Kreisel [3] T_1 is interpretable in Gödel's system T , that is $T_1 \preceq T$. On the other hand we have that Gödel's functional interpretation [1] interprets T_2 in T and so we have both $T_1 \preceq T$ and $T_2 \preceq T$. Therefore we accept both T_1 and T_2 as truly constructive theories.⁵ A crucial point here, which I will return to, is that these two theories are constructively incompatible.

An objection against this way of getting confidence in complex theories is found by asking: How do we know that the realisers actually do what is required? If we look at the details of the reductions then we see that this fact is actually proved using the exact principles in question. Using only basic mathematical reasoning we know for instance that if $T_1 \vdash \exists x C(x)$, for any formula C then there exists a term in T which we can find such that $T_1 \vdash C(t)$. The thing is that we really need the full theory T_1 in order to prove that t has property C . However, we do not find this problematic. The whole thing is, first of all, computationally meaningful, and the

properties that the theories possess are certainly not enjoyed by theories based on classical logic. But one could still argue that the theories incorporate principles which are non-intuitionistic and one may *therefore* be skeptical towards the theories. Now, such an attitude is dogmatic: Why should precisely intuitionistic logic monopolise constructivity? The interpretations referred to here show that the respective principles are locally constructive—not that, say, Markov's principle generally is constructive. But it is relative to typed arithmetic with weak extensionality together with a restricted form of the independence-of-premise and the axiom of choice. Note that Markov's principle is not only validated for numbers (type o) but is also validated in higher types. It is easy to attach some understanding to the principle for numbers, but this is surely not straight forward in higher types.

The really interesting thing is that many of these principles are in several *combinations* constructively problematic. As shown in [2] the combination of IP and MP is demonstratively non-constructive (relative to arithmetic) and also full extensionality together with MP is likewise problematic. In this case, if one

wants to work with Markov's principle and an extensional notion of equality, then one will have to work with weak extensionality only.

Conclusions

It is clear that there is no unique global characterisation of constructivity. Surely, the proof interpretation by Kolmogorov and Heyting can be seen as a global interpretation trying to characterise constructivity. But this should not be taken as more than a guideline or a rule of thumb. For as we have seen above, locally in the context of typed arithmetic one can constructively accept more. But then things turn out in a very complex manner: It is indeed a *very* subtle issue to combine the different principles which treated separately can be given an interpretation. One can obtain constructive methodologies in various different ways which are not in harmony with one and each other, since the constructivity of mathematical methods depends (also) on the context in which the methods are applied.

It is also clear that the finer details and the more information one seeks, the more complex and diverse things get. More information but also more diversity follow if one wants to do mathematics con-

structively. This extreme variation and diversity clearly causes pain among classical mathematicians and is, probably, a primary reason why doing constructive mathematics is not a popular occupation.

Here we will not try to answer why diversity occurs when one does constructive mathematics. But it is clear that when one tries to ascribe constructive meaning to different ideal elements then it gets really difficult to obtain coherent systems.

There are several conclusions to draw from these investigations. When one does mathematics and wants constructive results, an analysis of what is actually required seems to be a good idea. For a certain concrete purpose one chooses the most optimal system. If one tries to develop everything within one single system, then it is clear that many results cannot be obtained, since they may need principles which together with those accepted at the outset are non-constructive. Markov's principle is, for instance, very useful—but maybe it is not such a good idea always to have it in your tool box.

Today larger and larger constructive systems are being build, but as the above has shown this is indeed a difficult task. First of all, one easily messes things up and gets non-

constructive systems; second of all, one loses much information.

References

- [1] Kurt Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica*, 12:280–287, 1958.
- [2] Klaus Frovin Jørgensen. Finite Type Arithmetic: Computable Existence Analysed by Modified Realisability and Functional Interpretation. Master's thesis, Roskilde, March 2001.
- [3] Georg Kreisel. On weak completeness of intuitionistic predicate logic. *Journal of Symbolic Logic*, 27:139–158, 1962.

Notes

¹The intended meaning of $S : o \rightarrow o$ is of course the mathematical function $f : \mathbf{N} \rightarrow \mathbf{N}$ defined by $n \mapsto n + 1$.

²With respect to terms, parentheses are associated to the left; thus $t_1 t_2 \dots t_n$ is short for $(\dots((t_1 t_2) t_3) \dots t_n)$.

³A theory S is said to have the existence property in case the following rule holds: $S \vdash \exists x A(x) \Rightarrow S \vdash A(t)$ for some term t . Likewise, S has the disjunction property if we have that for any provable closed formula $A \vee B$ there is a term deciding whether A or B holds.

⁴Full extensionality could also be formulated as a rule instead of using axioms. Then it has the same form as weak extensionality, just without any restrictions on the assumptions.

⁵Note, however, that T_1 is not closed under Markov's rule as shown in [2].