Hille, E. and R. S. Phillips. 1957. *Functional Analysis and Semi-groups*. Providence, R.I., American Mathematical Society.

Hindley, J. and J. Sheldon, eds. 1980. *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*. London, Academic Press.

Howard, W. 1980. The formulae-as-types notion of construction. In Hindley and Seldin (1980), 107–124.

Jacobson, N. 1953. *Lectures in Abstract Algebra II*. New York, Van Nostrand.

———. 1964. *Lectures in Abstract Algebra III*. New York, Van Nostrand.

Kelley, J. L. 1955. *General Topology*. New York, Van Nostrand.

Lang, S. 1965. *Algebra*. Reading, MA, Addison-Wesley.

Lebesgue, H. 1902. Intégrale, longueur, aire. *Annali di matematica pura ed applicata* 7(3), 231–259.

Martin-Löf, P. 1982. Constructive mathematics and computer programming. In Cohen et al. (1982), 153–178.

Moore, G. 1982. *Zermelo's Axiom of Choice: Its Origins, Development and Influence*. New York, Springer-Verlag.

Prawitz, D. 1965. *Natural Deduction: A Proof-Theoretic Study*. Stockholm, Almqvist & Wiksell.

Rowe, D. E. and J. McCleary, eds. 1989. *The History of Modern Mathematics*, Vol. I: *Ideas and Their Reception*. San Diego, Academic Press.

Rudin, W. 1953. *Principles of Mathematical Analysis*. New York, McGraw-Hill.

Tait, W. 1983. Against intuitionism: Constructive mathematics is part of classical mathematics. *Journal of Philosophical Logic* 12, 173–195.

———. 1986. Proof and truth: The Platonism of mathematics. *Synthese* 69, 341–70.

van der Waerden, B. L. 1937. *Moderne Algebra*, 2nd ed. Berlin, Julius Springer.

van Heijenoort, J. 1967. *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*. Cambridge, MA, Harvard University Press.

Zermelo, E. 1904. Beweise, das jede menge wohlgeordet werden kann (Aus einem an Herrn Hilbert gerichten Briefe). *Mathematische Annalen* 59, 514–516; translated in van Heijenoort (1967).

———. 1908. Neuer Beweise für die Möglichkeit einer Wohlordnung. *Mathematische Annalen* 65, 107–128.

From: Mathematics and Mind, ed. by A. George
Oxford University Press, 1994

4

Mechanical Procedures and Mathematical Experience

Wilfried Sieg

> Turing's "Machines." These machines are *humans* who calculate.
> —Ludwig Wittgenstein

Wittgenstein's terse remark[1] captures the feature of Alan Turing's analysis of calculability that makes it epistemologically relevant. Focusing on the epistemology of mathematics, I will contrast this feature with two striking aspects of mathematical experience implicit in repeated remarks of Kurt Gödel. The first, the *conceptional* aspect, is connected to the notion of mechanical computability through his assertion that "with this concept one has for the first time succeeded in giving an absolute definition of an interesting epistemological notion"; the second, the *quasi-constructive* one, is related to axiomatic set theory through his claim that its axioms "can be supplemented without arbitrariness by new axioms which are only the natural continuation of the series of those set up so far." Gödel speculated on how the second aspect might give rise to a humanly effective procedure that cannot be mechanically calculated and thus provide a reason for his belief that the class of mental procedures is not exhausted by mechanical ones. Leaving this latter speculation aside, Gödel's remarks point to data that underlie the two aspects and "challenge," in the words of Charles Parsons (in press, 19), "any theory of meaning and evidence in mathematics."[2]

Not that I will present a theory accounting for these data; rather, I will mainly clarify the first datum by reflecting on the question that is at the root of Turing's analysis and central for mathematical logic, as well as for cognitive psychology and artificial intelligence. In its sober mathematical form the question simply asks, "What is an effectively calculable function?" The equivalent answers given in the mid-1930s are widely taken to be of fundamental significance also for the less sober question, "Are we (reducible to) machines?" After all, Turing's answer to the mathematical question used the concept of an idealized computing machine. Turing presented his characterization in 1936 to give a negative solution to Hilbert's *Entscheidungsproblem*, and his characterization is generally accepted as correct or, at least, as more convincing than others. But what are the reasons for such a judgment? It seems to me that this issue has yet to be treated adequately.[3]

When approaching the original question, it is important to note its emergence from work in the foundations of mathematics. The first part of my essay provides this background by presenting epistemological concerns that motivated the use of effectively decidable notions in mathematics as well as in logic and by summarizing (meta-) mathematical issues that required an analysis of effective calculability. The second part starts out with a discussion of general recursive functions as introduced by (Herbrand and) Gödel, but it focuses on Church's main argument for the proposal to identify recursiveness with the informal notion of effective calculability. Thus, *Church's Thesis* is at the center of the second part. I will point out unsatisfactory aspects of Church's argument, but also the centrality of the concept *calculability in a logic* for this early discussion. That prepares the ground for *Turing's Analysis* of mechanical processes carried out by a human computor. The third part refines and generalizes that analysis, isolates *Turing's Thesis* as asserting that a human computor satisfies certain finiteness conditions, and argues for the pertinency and correctness of this thesis.[4]

The generalized form of the analysis allows me to connect Turing's considerations in a most informative way with Church's argument and Gödel's proposal. These systematic connections reinforce the conceptual core of the early investigations and weaken, if not undermine, the "argument from confluence of different notions" in favor of Church's Thesis. Turing's analysis was in perfect accord with Church's views, as can be gathered from the 1937 review Church wrote of Turing's paper. In his review Church asserted it is "*immediately clear*" that the notion of Turing computability "can be identified with ... the notion of effectiveness as it appears in certain mathematical problems (various forms of the Entscheidungsproblem ... and in general any problem which concerns the discovery of an algorithm)." Gödel was also convinced by Turing's analysis of the correctness of Church's Thesis and used the adequacy of Turing's notion to establish rigorous consequences for the mind and machine problem. The fourth part of this essay presents these Gödelian consequences and two closely related, but more general, aspects of mathematical experience alluded to in Gödel's remarks that were quoted previously. But note that these features are separable from Gödel's Platonism; they are more subtly attuned, I will argue, to the practice of mathematics.

1. Background

The precise connection between the informal notion of effective calculability and the mathematical notion of computability is brought to light, as section 3 will show in detail, by Turing's analysis of effectively calculable functions. According to Turing (1939), "A function is said to be 'effectively calculable' if its values can be found by some purely mechanical process,"[5] and mechanical

process—with a specifically human touch—was characterized by Turing through axiomatic conditions and was shown to be equivalent to Turing machine computation. For a critical appreciation of the analysis and its remarkable pertinency, it is crucial to be clear about the mathematical and philosophical context in which it arose; indeed, in section 3.2 I argue that the general "problematic" *required* an analysis of the kind Turing offered.

1.1. *Effectiveness in Mathematics and Logic*

The problematic of effective calculability emerged within two traditions in logic and mathematics where proper symbolic representations of problems and their algorithmic solution were sought.[6] These traditions met briefly in Leibniz; he viewed algorithmic solutions of mathematical and logical problems as paradigms of problem solving in general. Remember that he recommended to disputants in *any* field to sit down at a table, take pens in their hands, and say "Calculemus!" His recommendation was clearly based on high hopes for his *lingua characterica* and *calculus ratiocinator*. This is relevant pre-history; relevant history begins in the second half of the nineteenth century with detailed work in the foundations of mathematics, in particular, with the so-called arithmetization of analysis and the axiomatic characterization of the real numbers. Dirichlet had demanded that a systematic arithmetization should show that *any* theorem of algebra and higher analysis could be formulated as a theorem about natural numbers. In this way, I assume, he hoped to clarify the role of analytic methods in number theory; recall that it was he who had introduced such methods in the proof of his famous theorem on arithmetic progressions. Dedekind and Kronecker, both deeply influenced by Dirichlet, sought to give an arithmetization satisfying Dirichlet's demand, but they proceeded in radically different ways. Their pertinent essays brought out conflicting philosophical positions that have influenced, directly or indirectly, the subsequent foundational discussion. But—and I would like to emphasize this very strongly—these positions evolved from and influenced their closely related mathematical work in algebraic number theory. (The background and the evolution of their work should be the focus of a case study concerned with the revolutionary changes in mathematics during the nineteenth century.)

Kronecker admitted only natural numbers as objects of analysis outright; from them he constructed, in now familiar ways, integers and rationals. Even algebraic reals were introduced, because they could be isolated effectively as roots of algebraic equations. The general notion of irrational number, however, was rejected in consequence of two restrictive methodological conditions to which mathematical considerations have to conform: (1) concepts must be decidable in finitely many steps, and (2) existence proofs must be carried out in such a way that they present objects of the required kind. Consequently, for Kronecker, infinite mathematical objects could not exist. All of this added up to a strictly arithmetic procedure, and Kronecker thought that by following it analysis could be re-obtained. More than 100 years later, we know that such a redevelopment is not as chimerical as people in the 1920s, for example Hilbert, believed.[7]

Dedekind opposed Kronecker's methodological restrictions. With respect to the previous decidability condition he maintained that it is determined independently of our knowledge, whether an object does or does not fall under a concept. He also used infinite sets of natural numbers as respectable mathematical objects, for example, in his definition of real numbers by cuts. But how, you may ask, was the existence of such mathematical objects to be secured? Dedekind proposed to give purely logical proofs for the existence of models of axiomatically characterized notions, not of individual mathematical objects. Thus the consistency of the notions would be guaranteed. With regard to the development in his 1888 booklet, he wrote to Keferstein in a letter dated February 27, 1890:

> After the essential nature of the simply infinite system, whose abstract type is the number sequence N, had been recognized in my analysis, ... the question arose: does such a system *exist* at all in the realm of our ideas? Without a logical proof of existence it would always remain doubtful whether the notion of such a system might not perhaps contain internal contradictions. Hence the need for such a proof.[8]

Dedekind viewed his considerations not as specific for foundational systems, but rather as paradigmatic for a general mathematical procedure intended to secure the coherence of axiomatically given notions. In sum, Dedekind tried to safeguard his axiomatic approach by consistency proofs relative to logic broadly conceived, whereas Kronecker insisted on a radical restriction of mathematical objects and methods.

Dedekind recognized that Frege's logical foundation for natural numbers agreed with his own, not only in the details of justifying induction, but also in assuming the unrestricted comprehension schema as a logical principle. Dedekind's development of his theory was uncompromisingly rigorous but mathematically informal; Frege, by contrast, insisted on giving arguments in his Begriffsschrift. With this *formula language* Frege had realized some of Leibniz's hopes and for the first time provided the means necessary to formalize mathematical proofs. His booklet *Begriffsschrift* (1879) offered not only a rich language with relations and quantifiers, but its logical calculus also required that all assumptions be explicitly listed and that each step in a proof be taken in accord with one of the antecedently specified rules. Frege correctly considered this last requirement as a sharpening of the axiomatic method he traced back to Euclid's *Elements*. With this sharpening Frege pursued the aim of recognizing the "epistemological nature" of theorems. In the introduction to *Grundgesetze der Arithmetik* (1893) he wrote:

> By insisting that the chains of inference do not have any gaps we succeed in bringing to light every axiom, assumption, hypothesis or whatever else you want to call it on which a proof rests; in this way we obtain a basis for judging the epistemological nature of the theorem.

But such a basis can be obtained, Frege realized, only if inferences do not require contentual knowledge: their applications have to be recognizable as correct on account of the form of the sentences occurring in them. Indeed, Frege claimed that in his logical system "inference is conducted like a calculation," and he continued:

> I do not mean this in a narrow sense, as if it were subject to an algorithm the same as ... ordinary addition and multiplication, but only in the sense that there is an algorithm at all, i.e., a totality of rules which governs the transition from one sentence or from two sentences to a new one in such a way that nothing happens except in conformity with these rules.[9]

Almost fifty years later, in 1933, Gödel pointed back to Frege and Peano when he formulated "the outstanding feature of the rules of inference" in a formal mathematical system. The rules, Gödel said, "refer only to the outward structure of the formulas, not to their meaning, so that they can be applied by someone who knew nothing about mathematics, or by a machine" (Gödel 1933, 1).[10] To Frege, formulas represented (abstract) propositions in a concrete way; their concrete character provided the basis for the algorithmic transitions reflecting logical inferences. For this to be really useful the representation has to be adequate, and Frege asserted, "Proper use can be made of this only if the content is not just indicated, but if it is built up from its components by means of the very same logical signs, that serve for the computation." (This continues the quote in note 9.) Frege believed that his Begriffsschrift provided the means to represent content adequately.

1.2. Finitist Mathematics

It is all too well known that Frege's precise formal (re-)presentation did not prevent Russell from deducing a contradiction from the basic laws. A contradiction could also be obtained from the principles for Dedekind's notion of system. How this problem in Dedekind's foundational work already stirred Hilbert's concerned interest in the last few years of the nineteenth century is detailed in Sieg (1990a). But it was only in his paper of 1904 that Hilbert proposed a radically new, although still vague, approach to the consistency problem for mathematical theories. He suggested using the finiteness of mathematical proofs in order to establish directly, not through models, that contradictions could not be derived within particular mathematical theories. During the early 1920s he turned the issue into an elementary arithmetical problem and strategically joined the developments arising out of Frege's formal logical work with Kronecker's requirements for "genuine" mathematics (in order to save Dedekind's conception of the subject).[11]

The possibility of mechanically drawing inferences and of algorithmically solving some problems was not considered by Frege to be among the *logically* significant achievements of his *Begriffsschrift*. But Hilbert grasped the potential of this formal aspect, radicalized it, and exploited it for programmatic purposes, namely, to justify finitistically the use of classical theories **T** for establishing finitist statements without taking into account the problematic content of **T**.[12] That amounted to giving a finitist proof of the *reflection principle*

$$\Pr_T(x, `\phi') \rightarrow \phi$$

where Pr_T is the finitist proof predicate for T, ϕ a finitist statement, and 'ϕ' its translation in the language of T. This is directly related to the consistency problem, because the reflection principle is equivalent to the consistency statement for T under well-known conditions. It is perhaps worthwhile to mention that the connection to the nineteenth-century issues in the foundations of analysis was emphasized also by the independently minded Herbrand (1929b), who described a special case of the general consequence to be drawn from a finitist consistency proof for the system of *Principia Mathematica*:

> If an arithmetical theorem has been proved by using incommensurable numbers or analytic functions, then it can also be proved by using only purely arithmetic elements (integers and functions defined by recursion). Examples of this are Dedekind's theorem of prime numbers, and class field theory. (p. 43)[13]

It seemed that proof theoretic investigations would resolve the earlier methodological problems in a most satisfactory way because of the restricted character of finitist mathematics. Hilbert took finitist mathematics to be an elementary part of arithmetic, and it was assumed to coincide with the part of mathematics accepted by Kronecker and Brouwer.[14] Thus, for his metamathematical investigations, Hilbert joined the constructivist tradition in mathematics, crucial requirements of which had been formulated by Kronecker. The latter's views had influenced the French discussion surrounding the validity of the axiom of choice and of set theoretic methods in general at the very beginning of the twentieth century, and they were alive and well in Germany even during the 1920s.[15] The epistemological motivation for the restrictions was quite explicit in those discussions. As far as the still-evolving program of Hilbert and its direction were concerned, it was clearly formulated by Bernays in a talk at the 1921 meeting of the Deutsche Mathematiker Vereinigung in Jena:

> The assumption of such a system with particular connection properties [i.e., the assumption of the existence of a set of objects that satisfies certain axioms, W.S.] contains something as it were transcendent for mathematics, and thus the question arises, which principled position one should take with respect to it. . . . It would be quite hasty to deny from the very beginning any farther-reaching kind of intuitive evidence; nevertheless, we certainly want to take into account the tendency of the exact sciences to eliminate the more subtle organs of knowledge and use only the most primitive means of [acquiring] knowledge. From this perspective we are going to try [to determine] whether or not it is possible to justify those transcendent assumptions in such a way that only *primitive intuitive knowledge is being applied.* (1922, 11)[16]

Bernays goes on to discuss how Hilbert's approach addresses this problem and how it combines what is "positively fruitful" in the attempts of the intuitionists and logicists to provide a foundation for mathematics. A methodological point, similar to the main point in the previous quotation, was made by Bernays (1923), where he emphasized:

> The possibility of a philosophical position that recognizes [natural] numbers as existent, non-sensory objects is not excluded by Hilbert's theory—but then, logically speaking, the same kind of ideal existence would have to be granted

> also to transfinite numbers, in particular, to the numbers of the so-called second number class. It is its [the theory's] goal, however, to make such a position dispensable for the foundation of the exact sciences. (p. 163)

Within the finitist frame this ultimate goal of Hilbert's program could not be achieved, due to Gödel's incompleteness theorems; the latter forced a re-evaluation of the epistemological perspective that had been underlying Hilbert's program.[17]

There is one point I would like to emphasize, namely: Hilbert's metamathematical way of precisely describing formalisms and of investigating them by finitist means opened the way to the rigorous treatment of fascinating issues that are still being pursued. This novel approach, going radically beyond Frege, and its parallel to ordinary mathematical investigations were lucidly expressed in Hilbert and Ackermann (1928): "Mathematical logic achieves more than a sharpening of language by a symbolic representation of inferences. Once the logical formalism is fixed, we can expect that a systematic, so-to-speak calculatory treatment [rechnerische Behandlung] of logical formulas is possible that corresponds roughly to the theory of equations in algebra. . . ." Herbrand, as well as other young and quite brilliant mathematicians, was attracted by Hilbert's approach and viewed mathematical logic as a new branch of mathematics. He emphasized that it was independent of Hilbert's philosophical opinions, but that the novel questions opened "a scarcely explored domain of arithmetical investigations of the greatest interest, which may well contain surprises" (Herbrand 1931b, 276). In particular, metamathematics allowed a mathematical treatment of what Herbrand viewed "in a sense" as "the most general problem of mathematics." Already at the end of his thesis he had emphasized: "The solution of this problem would yield a general method in mathematics and would enable mathematical logic to play with respect to classical mathematics the role that analytic geometry plays with respect to ordinary geometry."[18] We turn now to this problem.

1.3. *Entscheidungsproblem*

The problem Herbrand alluded to is closely related to the consistency problem; it is the so-called *Entscheidungsproblem*, or decision problem, and was to be subjected to a *rechnerische Behandlung*, that is, a calculatory treatment. Its classical formulation is found in Hilbert and Ackermann (1928, 72–73): "The Entscheidungsproblem is solved if one knows a procedure that allows one to decide the validity (respectively, satisfiability) of a given logical expression by a finite number of operations." Hilbert and Ackermann italicized this paragraph and emphasized the fundamental importance of a solution to the decision problem. Indeed, Herbrand viewed the decision problem as another route to establishing consistency. Assume that T is a theory with finitely many axioms $H_1, \ldots, H_n$[19]; if $\neg\phi$ is a theorem of T, then the validity (for Herbrand that meant provability in predicate logic) of the formula $(H_1 \& \cdots \& H_n) \rightarrow \phi$ is equivalent to the inconsistency of T. This connection was explained by

Herbrand (1930b, 213) after having described the decision problem most interestingly as follows:

> However, there is another viewpoint from which work can be done and in which encouraging results have already been obtained: the study of what the Germans call the *Entscheidungsproblem*, which consists of seeking a method allowing us to recognize with certainty (at the end of a number of operations which can be determined beforehand) whether or not a given proposition is an identity, and if it is to find a proof of the proposition.

Thus, a solution to the decision problem must consist of a method that effectively yields an answer (i.e., after finitely many steps) and of a finitist proof that establishes the termination of the method; such a proof would guarantee certainty and would also provide a bound on the number of required steps.[20] These additional requirements make it understandable why Herbrand reproved and extended partial results that had been obtained by Löwenheim and Behmann.[21] He emphasized that only his proof, not Löwenheim's, satisfied the stringent metamathematical or finitist requirements: "We could say that Löwenheim's proof was sufficient in mathematics; but, in the present work, we had to make it 'metamathematical', . . so that it would be of some use to us."[22]

Researchers in the Hilbert school realized full well that a positive solution for predicate logic—together with the assumption of finite axiomatizability of theories and the quasi-empirical completeness of *Principia Mathematica*[23]—would allow the decision concerning the provability (truth) of any mathematical statement. For some that was a sufficient reason to expect a negative solution; von Neumann, for example, expressed his views (1927, 11–12) as follows:

> . . . it appears that there is no way of finding the general criterion for deciding whether or not a well-formed formula *a* is provable. (We cannot at the moment establish this. Indeed, we have no clue as to how such a proof of undecidability would go.) . . . the undecidability is even the *conditio sine qua non* for the contemporary practice of mathematics, using as it does heuristic methods, to make any sense. The very day on which the undecidability does not obtain any more, mathematics as we now understand it would cease to exist; it would be replaced by an absolutely mechanical prescription (eine absolut mechanische Vorschrift), by means of which anyone could decide the provability or unprovability of any given sentence.
>
> Thus we have to take the position: it is generally undecidable, whether a given well-formed formula is provable or not.

When claiming that we have no clue as to how a proof of undecidability would go, von Neumann pointed to the underlying conceptual problem. After all, there were well-known proofs for the unsolvability of certain mathematical problems; but these impossibility results were given relative to a determinate class of admissible means, for example, doubling the cube by using only straightedge and compass. Exactly here lies the problem: a negative solution to the *Entscheidungsproblem* required a mathematically precise answer to the question "What are *absolut mechanische Vorschriften*?"

It is almost a platitude to say that particular aspects of mathematical experience informed broad philosophical views on the nature of human knowledge; we just need to remind ourselves of Plato, Leibniz, Kant, or—closer to our own days—Frege and Husserl. On the other hand, epistemologically motivated concerns evolved, as I have shown, into normative requirements for the presentation of axiomatic mathematical theories. The resulting formal development of parts of mathematics seemed to give substance to the Hobbesian claim that mathematical reasoning is nothing but mechanical computation. This view came to the fore through the formalist and polemical side of Hilbert's Program: The whole "thought-content" of mathematics, so it was claimed, can be expressed in a comprehensive formal theory; mathematical activity can be reduced to the manipulation of symbolic expressions, and mathematics itself can be viewed as a formula game. Hilbert defended this playful view of classical mathematics against the intuitionists by remarking that

> The formula game that Brouwer so deprecates has, besides its mathematical value, an important general philosophical significance. For this formula game is carried out according to certain definite rules, in which the *technique of our thinking* is expressed. These rules form a closed system that can be discovered and definitively stated. The fundamental idea of my proof theory is none other than to describe the activity of our understanding, to make a protocol of the rules according to which our thinking actually proceeds. Thinking, it so happens, parallels speaking and writing: we form statements and place them one behind another. If any totality of observations and phenomena deserves to be made the object of serious and thorough investigation, it is this one— . . .[24]

Hilbert's last remark is undoubtedly correct. However, if we take the possibility of developing mathematics formally as a significant datum for reflection, we must keep in mind that the formality requirement expressed a philosophically motivated *restriction* on human cognitive capacities for particular purposes.[25] By addressing von Neumann's conceptual problem, we also will lay the basis for a characterization of these restricted cognitive capacities that are presupposed in formal presentations.

2. Church's Thesis

The background I just described—with its interweaving of mathematical, logical, and philosophical questions—should be kept in mind when we turn our attention to the central conceptual issue. I want to emphasize that in depicting the decision problem as *the* immediate context in which an analysis of effective calculability was needed I do not intend to neglect two other significant and closely related issues; namely, the *general formulation* (and thus applicability) of the incompleteness theorems and the *general characterization* of effective solvability for mathematical problems.[26] Indeed, it was the detailed examination of the incompleteness theorems and the notion of *Entscheidungsdefinitheit*, so pivotal for their proofs, that led the way to the (informal) understanding of effective calculability as rule-governed evaluation of number-theoretic functions in something like a formal calculus. This understanding underlies Gödel's proposal, is crucial for Church's early considerations and for his main argument (analyzed on pp. 85–87), and leads to a specially important class of Post's finitary processes.[27] It is this notion that is recognized by Gödel as *absolute* and was generalized, later on, by Hilbert and Bernays to their notion of *regelrechte Auswertbarkeit* (i.e., evaluation according to rules) in deductive formalisms. Technically, this understanding found its distinctive expression in Kleene's Normal Form Theorem. *Here we have a conceptual core that is associated, however, with a major stumbling block.* After all, this core does not provide a convincing analysis: steps taken in a calculus must be of a restricted character and they are assumed, for example by Church, without argument to be recursive. A related assumption is made by Hilbert and Bernays; the proof predicate of their deductive formalisms has to be primitive recursive. Finally, Post offers as a *working hypothesis* only that the primitive acts (steps) of his "formulation 1" are sufficient for a reduction of ever wider formulations. As to Gödel's dissatisfaction with his proposal, see the discussion in sections 2.1 and 2.4. In section 3, I will show that Turing's analysis removes exactly this stumbling block.

2.1. *Gödel's General Recursion*

Examples of effectively calculable functions were given by primitive recursive functions; they had been used in mathematical practice for a long time. The standard arithmetic operations like addition, multiplication, and exponentiation, but also the sequence of prime numbers and the Fibonacci numbers, are all primitive recursive. The schema of primitive recursion leads from primitive recursive functions g and h to a new primitive recursive function f satisfying the equations

$$f(x_1, \ldots, x_n, 0) = g(x_1, \ldots, x_n)$$

$$f(x_1, \ldots, x_n, y') = h(x_1, \ldots, x_n, y, f(x_1, \ldots, x_n, y))$$

The defining equations for f can be used as rules for determining the value of f for any particular set of arguments. Clearly, in order to recognize that this is a well-defined procedure one appeals to the buildup of the structure N. Dedekind gave a set-theoretic foundation of these functions,[28] whereas Skolem used them directly with their naive number-theoretic meaning in his development of elementary arithmetic through the recursive mode of thought. Hilbert and Bernays, finally, sharpened Skolem's mathematical frame to their Primitive Recursive Arithmetic (PRA). And it is most plausible that finitist mathematics, as intended by them, coincides with PRA—up to an elementary and unproblematic coding of finite mathematical objects as numbers.[29]

Primitive recursive functions and predicates were used in Gödel's classical paper (1931) to describe a simplified system of *Principia Mathematica*; obviously, syntactic structures had to be coded as numbers. From a finitist standpoint it was perfectly sensible to restrict the means for describing syntactic structures to primitive recursive functions; from a broader perspective, however, there was no reason to exclude other effective procedures in presenting "formal" theories.

Ackermann (1928) gave an effectively calculable, nonprimitive recursive function; that result had been mentioned already in 1925 by Hilbert. Thus, it made extremely good sense that in his 1934 Princeton lectures Gödel strove, as indicated by the title "On undecidable propositions of formal mathematical systems," to make his incompleteness results less dependent on particular formalisms. In the introductory section 1 he discussed the notion of "a formal mathematical system" in some generality and required that

> the rules of inference, and the definitions of meaningful formulas and axioms, be constructive; that is, for each rule of inference there shall be a finite procedure for determining whether a given formula B is an immediate consequence (by that rule) of given formulas $A_1, \ldots, A_n$, and there shall be a finite procedure for determining whether a given formula A is a meaningful formula or an axiom. (p. 346)

Again, he used primitive recursive functions and relations to present syntax, viewing the primitive recursive definability of formulas and proofs as a "precise condition which in practice suffices as a substitute for the unprecise requirement of section 1 that the class of axioms and the relation of immediate consequence be constructive."[30] But a notion that would suffice *in principle* was really needed, and Gödel attempted to arrive at a more general notion. He considered the fact that the value of a primitive recursive function can be computed by a "finite procedure" for each set of arguments as an "important property" and in footnote 3 added that "The converse seems to be true if, besides recursions according to the scheme (2) [i.e., primitive recursion as given above], recursions of other forms (e.g., with respect to two variables simultaneously) are admitted. This cannot be proved, since the notion of finite computation is not defined, but it can serve as a heuristic principle" (1986, 348).[31]

Other recursions that might be admitted are discussed in the last section of the lecture notes under the heading "General Recursive Functions." In it Gödel described a proposal for the definition of a general notion of recursive function that (he thought) had been suggested to him by Herbrand in a private communication of April 7, 1931: "If ϕ denotes an unknown function, and $\psi_1, \ldots, \psi_k$ are known functions, and if the ψ's and ϕ are substituted in one another in the most general fashions and certain pairs of resulting expressions are equated, then, if the resulting set of functional equations has one and only one solution for ϕ, ϕ is a recursive function" (Gödel 1986, 368).[32] He went on to make two restrictions on this definition and required, first of all, that the left-hand sides of the functional equations be in a standard form, with ϕ being the outermost symbol and, secondly, that "for each set of natural numbers $k_1, \ldots, k_l$ there shall be exactly one and only one m such that $\phi(k_1, \ldots, k_l) = m$ is a derived equation." The rules that were allowed in giving derivations are of a very simple character: variables in any derived equation can be replaced by numerals, and if the equation $\phi(k_1, \ldots, k_l) = m$ has been obtained, then occurrences of $\phi(k_1, \ldots, k_l)$ on the right-hand side of a derived equation can be replaced by m. So much about this proposal; it was taken up for systematic development by Kleene (1936).

What was important about Gödel's modifications? For Gödel himself the crucial point was the precise specification of *mechanical* rules for deriving equations or, to put it differently, for carrying out computations. That point of view was also expressed by Kleene, who wrote with respect to the definition of "general recursive function of natural numbers" that "it consists in specifying the form of the equations and the nature of the steps admissible in the computation of the values, and in requiring that for each given set of arguments the computation yield a unique number as value" (1936, 727). In a letter to van Heijenoort, dated August 14, 1964, Gödel asserted that "it was exactly by specifying the rules of computation that a mathematically workable and fruitful concept was obtained" (van Heijenoort 1985a, 115). When making this claim, Gödel took for granted what he had expressed in an earlier letter to van Heijenoort, namely, that Herbrand's suggestion had been "formulated *exactly* as on page 26 of my lecture notes, i.e., without reference to computability."[33] But Gödel had been unable to find Herbrand's letter among his papers and had to rely on his recollection, which, he said, "is very distinct and was still very fresh in 1934." However, Herbrand's letter was found by John W. Dawson in Gödel's *Nachlass*, and it reads like a preliminary version of parts of Herbrand (1931d). On the evidence of that letter it is clear that Gödel misremembered. As a matter of fact, describing a system of arithmetic and the introduction of recursively defined functions *into that system* with intuitionistic, that is, finitist, justification, Herbrand wrote:

> In arithmetic we have other functions as well, for example functions defined by recursion, which I will define by means of the following axioms. Let us assume that we want to define all the functions $f_n(x_1, x_2, \ldots, xp_n)$ of a certain finite or infinite set F. Each $f_n(x_1, \ldots)$ will have certain defining axioms; I will call these axioms (3F). These axioms will satisfy the following conditions:
> (i) The defining axioms for f_n contain, besides f_n, only functions of lesser index.
> (ii) These axioms contain only constants and free variables.
> (iii) We must be able to show, by means of intuitionistic proofs, that with these axioms it is possible to compute the value of the functions univocally for each specified system of values of their arguments.

It is most plausible that, in addition to the (intuitionistically interpreted) axioms, Herbrand admitted substitution rules of the sort formulated by Gödel as rules of computation. Indeed, he asserted (1931d)—as he had done in his letter to Gödel—that all intuitionistic computations can be carried out, for example, in the formal system P of *Principia Mathematica*. This is not to suggest that Gödel was wrong in his assessment, but rather to point to the most important step he had taken, namely, to *disassociate recursive functions from an epistemologically restricted notion of proof.* Later on, Gödel even dropped the regularity condition that guaranteed the totality of calculable functions. He emphasized then "that the precise notion of mechanical procedures is brought out clearly by Turing machines producing partial rather than general recursive functions" (Wang 1974, 84).[34] At this earlier historical juncture, however, the explicit introduction of an equational calculus with purely formal, mechanical

rules for computing was important for the mathematical development of recursion theory and also for the conceptual analysis. After all, it brought out clearly what, according to Gödel, Herbrand had failed to see, namely, "that the computation (for all computable functions) proceeds by exactly the same rules" (van Heijenoort 1985a, 115).

2.2. *Herbrand's Provably Recursive Functions*

Before moving on to the further development, I want to make some additional remarks concerning Herbrand's proposals, emphasizing, in particular, the restrictive provability conditions he imposed. These remarks complement the discussion of the last subsection, but do constitute a digression: the main considerations are taken up again in section 2.3. A careful description and thoughtful interpretation of the proposals can be found in van Heijenoort's essay (1985a) on Herbrand's work. It should be noted, however, that this paper was written before Dawson's discovery of the Gödel–Herbrand correspondence. Van Heijenoort thus had to rely on Gödel's reports concerning the details of Herbrand's suggestion to him and its very framing, namely, that Herbrand was concerned with a general characterization of effective calculability. In any event, van Heijenoort distinguished three different occasions in 1931 on which Herbrand "proposed . . . to introduce a class of computable functions that would be more general than that of primitive recursive functions" (1985a, 114).[35] The first proposal is found in Herbrand's note (1931b, 273), where he described the restricted means allowed in metamathematical arguments and required, in particular, that "all the functions introduced must be actually calculable for all values of their arguments by means of operations described wholly beforehand." The second proposal is the one reported in Gödel's lectures (without reference to computability), and the third suggestion was made in Herbrand's paper (1931d, 290, 291). It is formulated as follows, again in the context of a system for arithmetic:

> We can also introduce any number of functions $f_i(x_1, x_2, \ldots, xn_i)$ together with hypotheses such that
> (a) the hypotheses contain no apparent variables;
> (b) considered intuitionistically, they make the actual computation of the $f_n(x_1, x_2, \ldots, xp_n)$ possible for every given set of numbers, and it is possible to prove intuitionistically that we obtain a well-determined result.

To the first occurrence of "intuitionistically" in this quotation Herbrand attached the following footnote: "This expression means: when they are translated into ordinary language, considered as a property of integers and not as mere symbols." With van Heijenoort I assume that Herbrand used "intuitionistic" also here as synonymous with "finitist." (A more detailed description of intuitionistic arguments is given in Herbrand 1931d, note 5, 288–289.[36]) This third proposal is identical with the one made by Herbrand in his letter to Gödel that I quoted previously, except for clause (i) from the earlier definition; that clause is implicitly assumed, as is clear from the examples Herbrand discusses.

[On] the one hand, I view the first formulation as a preliminary, not fully elaborated, version of the second and third formulations; on the other hand, I view it as a more explicit description of the Kroneckerian elements in metamathematics that were pointed out in section 1. Thus, we can see the evolution of essentially one formulation!

Between these proposals there is certainly no conflict of the sort Gödel considered and that is reported by van Heijenoort (1985a, 115–117); that is, that Herbrand envisioned "unformalized and perhaps unformalizable computation methods" and indeed refused "to confine himself to formal rules of computation." This should become clear from the following discussion. Two features of Herbrand's schema have to be distinguished, namely, (1) that the defining axioms (plus suitable rules) must make the actual intuitionistic computation of the function value possible, and (2) that the termination of the computation with a unique value has to be provable intuitionistically. That is, in modern terminology, we are dealing with "intuitionistically provably total (or provably recursive) functions," where provability is, however, not a formal notion.

A connection to a formal notion is given by Herbrand in the fourth section of 1931d. Gödel's incompleteness theorems for the system P of *Principia Mathematica* are discussed there, and Herbrand asserts that any intuitionistic computation can be carried out in P and that any intuitionistic argument can be formalized in P. He concludes, after sketching Gödel's proof, that P's consistency is not provable by arguments formalizable in P and thus is not provable intuitionistically. What is most interesting is his remark that Gödel's argument does not apply to the system of arithmetic that includes the above schema for introducing functions: the functions that are introducible cannot be described intuitionistically, because we could easily diagonalize and obtain additional functions. In two side remarks I want to mention that (1) Herbrand's last observation can be turned around so as to imply that the class of provably total functions of a formal theory cannot be enumerated by an element of that class, and (2) the aim of precisely characterizing the class of provably total functions for formal theories has been taken up in proof-theoretic research, starting with Georg Kreisel (1952); see also Gandy (1988, 74–75) and my "Herbrand analyses" (1991).

What is the extension of Herbrand's class of functions? According to Herbrand's discussion noted in the last three paragraphs, it properly includes the primitive recursive functions and is included in the class of provably recursive functions of P. Indeed, at the end of 1931d, Herbrand asserts that in the previous claims concerning the formalizability of intuitionistic computations and arguments ordinary analysis (I assume that means full second-order arithmetic) can take the place of P. Indeed, he conjectures that full first-order arithmetic, with recursion equations for addition and multiplication only, might be sufficient. If the latter conjecture were true, Herbrand's class would be included in the class of provably recursive functions of Peano arithmetic. Basic to this discussion is Herbrand's conviction that the system of arithmetic described in 1931d (even without the infinitary rule D) allows one to carry out all intuitionistic proofs. That paper was dated Göttingen, July 14, 1931; in the

letter to Gödel sent from Berlin on April 7, 1931, the claim concerning intuitionistic proofs is explicitly stated for the much weaker system with only quantifier-free induction. As a matter of fact, Herbrand claims there also that "... each proof in this arithmetic, which has no bound variables, is intuitionistic—this fact rests on the definition of our functions and can be seen directly." If that were true, Herbrand's class would consist of only the primitive recursive functions.[37] In conclusion, it seems that Gödel was right—for stronger reasons than he put forward—when he cautioned that Herbrand had *foreshadowed*, but not *introduced*, the notion of general recursive function.[38]

2.3. Church's Main Argument

The concept introduced by Gödel characterized a wide class of calculable functions, a class that contained all known effectively calculable functions. Indeed, footnote 3 of the Princeton lectures that I quoted earlier seems to express a form of Church's Thesis. But in a letter to Martin Davis dated February 15, 1965, Gödel emphasized that no formulation of Church's Thesis is implicit in that footnote. He wrote:

> The conjecture stated there only refers to the equivalence of "finite (computation) procedure" and "recursive procedure." However, I was, at the time of these lectures, not at all convinced that my concept of recursion comprises all possible recursions; and in fact the equivalence between my definition and Kleene's ... is not quite trivial. (Davis 1982, 8)[39]

At the time Gödel was equally unconvinced by Church's proposal that effective calculability should be identified with λ-definability. In a conversation with Church early in 1934, he called the proposal "thoroughly unsatisfactory."[40] In spite of Gödel's not exactly encouraging reaction, Church announced his "thesis" in a talk given at the meeting of the American Mathematical Society in New York City on April 19, 1935. It was formulated in terms of recursiveness, not λ-definability.[41] I quote the abstract of the talk in full (Church 1935b):

> Following a suggestion of Herbrand, but modifying it in an important respect, Gödel has proposed (in a set of lectures at Princeton, N.J., 1934) a definition of the term *recursive function*, in a very general sense. In this paper a definition of *recursive function of positive integers* which is essentially Gödel's is adopted. And it is maintained that the notion of an effectively calculable function of positive integers should be identified with that of a recursive function, since other plausible definitions of effective calculability turn out to yield notions that are either equivalent to or weaker than recursiveness. There are many problems of elementary number theory in which it is required to find an effectively calculable function of positive integers satisfying certain conditions, as well as a large number of problems in other fields which are known to be reducible to problems in number theory of this type. A problem of this class is the problem to find a complete set of invariants of formulas under the operation of conversion (see abstract 41.5.204). It is proved that this problem is unsolvable, in the sense that there is no complete set of effectively calculable invariants.

2.4. *Absoluteness*

The concept used in Church's argument is an extremely natural and fruitful one and is, of course, directly related to *Entscheidungsdefinitheit* for relations and classes introduced by Gödel in his 1931 paper and to representability of functions as used in his 1934 Princeton lectures.[50] Clearly, the equational calculus and the λ-calculus are two particular "logics" allowing the formal, mechanical computation of calculable functions in ways that are motivated by special circumstances. Gödel himself used the general notion "f is computable in a formal system S" in a brief note of 1936 entitled "On the Length of Proofs." He considered a hierarchy of systems S_i (of order i, $1 \leq i$), and observed in a "Remark" added to the note in proof that this notion of computability is independent of i in the following sense: If a function is computable in any of the systems S_i, possibly of transfinite order, then it is already computable in S_1. "Thus," Gödel (1986, 397) concluded, "the notion 'computable' is in a certain sense 'absolute', while almost all metamathematical notions otherwise known (for example, provable, definable, and so on) quite essentially depend upon the system adopted."[51] For someone who stressed the type-relativity of provability as strongly as Gödel did, this must have been a very surprising insight indeed. In his contribution to the Princeton Bicentennial Conference (1946), Gödel re-emphasized this absoluteness and took it as the main reason for the special importance of general recursiveness or Turing computability: Here, Gödel thought, we have the first interesting epistemological notion whose definition is not dependent on the chosen formalism. The significance of his discovery was described by Gödel to Kreisel in a letter of May 1, 1968: "That my [incompleteness] results were valid for all possible formal systems began to be plausible for me (that is since 1935[52]) only because of the *Remark* printed on p. 83 of 'The Undecidable' . . . But I was completely convinced only by Turing's paper" (Odifreddi 1990, 65).[53] And there was good reason not to be completely convinced. After all, the absoluteness was achieved, ironically, only relative to the description of the "formal" systems S_i; the stumbling block shows up exactly here.

Remark. If Gödel had been completely convinced of the adequacy of this notion, he could have established most easily the unsolvability of the decision problem for first-order logic: Given that mechanical procedures are exactly those that can be computed in the system S_1 (or any other system to which Gödel's incompleteness theorem applies) the unsolvability follows from Theorem IX of Gödel (1931). The theorem states that there are formally undecidable problems of predicate logic; it rests on the observation (made by Theorem X) that every sentence of the form $(\forall x)F(x)$, with F primitive recursive, can be shown in S_1 to be equivalent to the question of satisfiability for a formula of predicate logic. Historically, Theorem IX made a positive solution of the decision problem very unlikely. But, for the appendix to his paper on the fundamental problem of mathematical logic, Herbrand wrote in April 1931, when he already knew Gödel's results quite well (1931a, 259): "Note finally that, although at present it seems unlikely that the decision problem can be solved, it has not yet been proved that it is impossible to do so."

The absoluteness of the notion of computability in Gödel's sense follows from a marvelous and detailed example of conceptual analysis due to Hilbert and Bernays. They established independence from formalisms in an even stronger sense in *Grundlagen der Mathematik*; supplement 2 was entitled "Eine Präzisierung des Begriffs der berechenbaren Funktion und der Satz von Church über das Entscheidungsproblem." They made the core notion of *calculability in a logic* directly explicit and defined a number-theoretic function to be *regelrecht auswertbar* when it is computable (in the above sense) in *some* deductive formalism. Deductive formalisms must satisfy three *Rekursivitätsbedingungen* (recursiveness conditions). I will discuss only the crucial condition here, an analogue to Church's Central Thesis. That condition requires that the theorems of the formalism be enumerated by a primitive recursive function or, equivalently, that the proof predicate be primitive recursive. Then it is shown that (1) a special number-theoretic formalism (included in Gödel's S_1) suffices to compute the functions that are *regelrecht auswertbar*, and (2) the functions computable in this particular formalism are exactly the general recursive ones. Hilbert and Bernays' analysis is, in my view, a natural and most satisfactory capping of the development from *Entscheidungsdefinitheit* to an "absolute" notion of computability, because it captures directly the informal notion of rule-governed evaluation of effectively calculable number-theoretic functions and isolates the necessary restrictive conditions. But this analysis does *not* overcome the major stumbling block; rather, it puts it in plain view. Let me emphasize the main point of this subsection: the notion of rule-governed computation (in something like a logical calculus) provides *a conceptual core for the attempts to characterize effective calculability of number-theoretic functions; the core is associated, however, with a major stumbling block.* It is Turing's analysis, taking processes underlying computations in a "calculus" as a starting-point, that removes the stumbling block.

3. Turing's Analysis

Turing's notion of machine computability turned out to be equivalent to recursiveness and λ-definability, but it was hailed by Gödel as providing "a precise and unquestionably adequate definition of the general concept of formal system." In his review of Turing's paper, Church (1937a, 43) claimed when comparing Turing computability, recursiveness, and λ-definability: "Of these, the first has the advantage of making the identification with effectiveness in the ordinary (not explicitly defined) sense evident immediately—i.e., without the necessity of proving preliminary theorems."[54] What distinguished Turing's proposal so dramatically from Church's, at least for Gödel and Church? One has to find an answer to this question, because the naive examination of Turing machines hardly produces the conviction formulated by Gödel and hardly carries the immediate evidence asserted by Church. (The key to the answer to the question is offered at the end of section 3.1; the answer is formulated in sections 3.2 and 3.3.)

3.1. Turing's Machines and Post's Workers

Let me describe Turing machines very briefly, following Davis (1958), rather than Turing's original presentation.[55] A Turing machine consists of a finite but potentially infinite tape; the tape is divided into squares, and each square may carry a symbol from a finite alphabet, say, just the two-letter alphabet consisting of 0 and 1, or B(lank) and |. The machine is able to scan one square at a time and perform, depending on the content of the observed square and its own internal state, one of four operations: print 0, print 1, or shift attention to one of the two immediately adjacent squares. The operation of the machine is given by a finite list of commands in the form of quadruples $q_i s_k c_l q_m$ that express the following: If the machine is in internal state q_i and finds symbol s_k on the square it is scanning, then it is to carry out operation c_l and change its state to q_m. The deterministic character of the machine operation is guaranteed by the requirement that a program must not contain two different quadruples with the same first two components. Gandy (1988, 88) gave a lucid description of a Turing machine computation in very general terms without using internal states or, as Turing called them, states of mind: "The computation proceeds by discrete steps and produces a record consisting of a finite (but unbounded) number of cells, each of which is either blank or contains a symbol from a finite alphabet. At each step the action is local and is locally determined, according to a finite table of instructions." How the reference to internal states can be avoided should be clear from the following discussion of Post's worker, and why such a general formulation is appropriate will be seen in section 3.2.

For the moment, however, let me consider the special Turing machines I just described. Taking for granted a representation of natural numbers in the two-letter alphabet and a straightforward definition of when to call a number-theoretic function Turing computable, I turn the earlier remark into a question: Does this notion provide (via some Gödel numbering) "an unquestionably adequate definition of the general concept of formal system?" Is it at all plausible that every effectively calculable function is Turing computable? It seems to me that a naive inspection of the seemingly very restricted notion of Turing computability should lead to "no" as a tentative answer to the second question and, thus, to the first one. However, a systematic development of the theory of Turing computability quickly convinces one that it is indeed a powerful notion. One almost immediately goes beyond the examination of particular functions and the writing of programs for machines computing them; instead, one considers machines that correspond to operations that yield, when applied to computable functions, other functions that are again computable.[56] Two such functional operations are crucial, namely, composition and minimalization. Given those operations and the Turing computability of a few simple initial functions, the computability of *all* recursive functions follows. (This claim takes for granted Kleene's 1936 proof of the equivalence between general recursiveness in Gödel's sense and μ-recursiveness.) Because Turing computable functions are readily shown to be among the recursive ones, it seems that we are now in exactly the same position as before with respect to the evidence for

Church's Thesis. This remark holds also for Post's model of computation, which is so strikingly similar to Turing's.

In 1936, the very year in which Turing's paper appeared, Post published a brief note in *The Journal of Symbolic Logic* with the title "Finite Combinatory Processes—Formulation 1." Here we have a worker who operates in a *symbol space* consisting of

> a two-way infinite sequence of spaces or boxes, i.e., ordinally similar to the series of integers.... The problem solver or worker is to move and work in this symbol space, being capable of being in, and operating in but one box at a time. And apart from the presence of the worker, a box is to admit of but two possible conditions, i.e., being empty or unmarked, and having a single mark in it, say a vertical stroke.[57]

The worker can perform a number of *primitive acts*, namely, make a vertical stroke [V], erase a vertical stroke [E], move to the box immediately to the right [M_r] or to the left [M_l] (of the box he is in), and determine whether that box is marked or not [D]. In carrying out a particular combinatory process, the worker begins in a special box (the *starting point*) and then follows directions from a finite, numbered sequence of instructions. The ith direction, i between 1 and n, is in one of the following forms: (1) carry out act V, E, M_r, or M_l and then follow direction j_i; (2) carry out act D and then, depending on whether the answer was positive or negative, follow direction j_i' or j_i''. (Post has a special stop instruction, but that can be replaced by stopping, conventionally, in case the number of the next direction is greater than n.) Are there intrinsic reasons for choosing formulation 1, except for its simplicity and Post's expectation that it will turn out to be equivalent to recursiveness? An answer to this question is not clear (from Post's paper), and the claim that psychological fidelity is the aim seems quite opaque. At the very end of his paper Post wrote:

> The writer expects the present formulation to turn out to be equivalent to recursiveness in the sense of the Gödel–Church development. Its purpose, however, is not only to present a system of a certain logical potency but also, in its restricted field, of psychological fidelity. In the latter sense wider and wider formulations are contemplated. On the other hand, our aim will be to show that all such are logically reducible to formulation 1. We offer this conclusion at the present moment as a *working hypothesis*. And to our mind such is Church's identification of effective calculability with recursiveness. (Davis 1965, 291)[58]

For Post, investigating wider and wider formulations and reducing them to formulation 1 would change this "hypothesis not so much to a definition or to an axiom but to a *natural law*."[59]

It is methodologically remarkable that Turing proceeded in exactly the opposite way when trying to support the claim that all computable numbers are machine computable or, in our way of speaking, that all effectively calculable functions are Turing computable: He did not try to extend a narrow notion reducibly and obtain in this way additional quasi-empirical support; rather, he analyzed the intended broad concept and reduced it to a narrow one—once and

...l. (I would like to emphasize this, as it is claimed over and over that Post provided in his 1936 paper "much the same analysis as Turing."[60]) By examining Turing's analysis and reduction we can find the key to answering the question I raised about the difference between Church's and Turing's proposals. Very briefly put, it is this: Turing deepened Church's step-by-step argument by focusing on the mechanical operations underlying the steps and by formulating finiteness conditions that *guarantee* their recursiveness. Let me now present Turing's considerations in systematic detail, with simplifications and added structure.

3.2. Mechanical Computor

Turing's classical paper (1936) opens with a brief description of what is ostensibly its subject, namely, "computable numbers" or "the real numbers whose expressions as a decimal are calculable by finite means" (Davis 1965, 116). Turing is quick to point out that the fundamental problem of explicating "calculable by finite means" is the same when considering computable functions of an integral variable, computable predicates, and so forth. So it is sufficient to address the question: What does it mean for a real number to be calculable by finite means? Turing admits that "this requires rather more explicit definition. No real attempt will be made to justify the definitions given until we reach §9. For the present I shall only say that *the justification lies in the fact that the human memory is necessarily limited*" (Davis 1965, 117).[61] In section 9 he argues that the operations of his machines "include all those which are used in the computation of a number." (Clearly, the operations need not be available as basic ones; it suffices that they can be mimicked by suitably complex subroutines.) He does not try to establish the claim directly; rather, he attempts to answer "the real question at issue," that is, "What are the possible processes which can be carried out [implicitly: by a human computor[62]] in computing a number?" Given the systematic context that reaches back to Leibniz's "Calculemus!", this is exactly the pertinent question to ask, because the general problematic *requires* an analysis of the possibilities of a mechanical computor. Gandy (1988, 83–84) emphasizes, absolutely correctly, as we will see, that "Turing's analysis makes no reference whatsoever to calculating machines. Turing machines appear as a result, as a codification, of his analysis of calculations by humans."

Turing imagines a mechanical computor writing symbols on paper that is divided into squares "like a child's arithmetic book." Since the two-dimensional character of this computing space is taken not to be an "essential of computation,"[63] Turing takes a one-dimensional tape divided into squares as the basic computing space.[64] What determines the steps of the computor? And what kind of elementary operations can he carry out? Before turning to these questions, let me formulate one important restriction. It is motivated by definite limits of our sensory apparatus to distinguish—*at one glance*—between symbolic configurations of sufficient complexity; it states that only finitely many distinct symbols can be written on a square. This restriction will be part of condition

(1.1). Turing suggests a reason for this restriction by remarking,[65] "If we were to allow an infinity of symbols, then there would be symbols differing to an arbitrarily small extent." There is a second but closely related way of arguing for this restriction: If, for example, Arabic numerals like 17 or 9999999 are considered as one symbol, then it is not possible for us to determine at one glance whether or not 9889995496789998769 is identical with 98899954967899998769.

The behavior of a computor is determined *uniquely* at any moment by two factors: (1) the symbols or symbolic configuration he observes, and (2) his "state of mind" or "internal state." This uniqueness requirement may be called the *determinacy condition* (D) and guarantees that computations are deterministic. Internal states are introduced in order to have the computor's behavior depend possibly on earlier observations, that is, to reflect the computor's experience.[66] Turing wants to isolate operations of the computor that are "so elementary that it is not easy to imagine them further divided" (Davis 1965, 136). Thus, it is crucial that symbolic configurations relevant for fixing the circumstances for the actions of a computor are *immediately recognizable*, and we are led to postulate that a computor has to satisfy two *finiteness conditions*:

(1.1) There is a fixed finite number of symbolic configurations a computor can immediately recognize.

(1.2) There is a fixed finite number of states of mind that need to be taken into account.

For a given computor there are only finitely many different relevant combinations of symbolic configurations and internal states. Because the computor's behavior is—according to (D)—uniquely determined by such combinations and associated operations, the computor can carry out at most finitely many different operations and, *consequently*, his behavior is fixed by a finite list of commands. The operations a mechanical computor can carry out are restricted as follows:

(2.1) Only elements of observed symbolic configurations can be changed.[67]

(2.2) The distribution of observed squares can be changed, but each of the new observed squares must be within a bounded distance *L* of an immediately previously observed square.[68]

Turing emphasizes that "the new observed squares must be immediately recognizable by the computor," and that means the distributions of the new observed squares arising from changes according to (2.2) must be among the finitely many ones of (1.1). Clearly, the same must hold for the symbolic configurations resulting from changes according to (2.1). Because some of the operations may involve a change of state of mind, Turing concludes:

> The most general single operation must therefore be taken to be one of the following: (A) A possible change (a) of symbol [as in (2.1)] together with a possible change of state of mind. (B) A possible change (b) of observed squares [as in (2.2)] together with a possible change of state of mind.[69]

With this restrictive analysis of the steps a mechanical computor can take, the proposition that his computations can be carried out by a Turing machine is established rather easily. Indeed, Turing first "constructs" machines that mimic the work of the computor directly and then observes:

> The machines just described do not differ very essentially from computing machines as defined in §2, and corresponding to any machine of this type a computing machine can be constructed to compute the same sequence, that is to say the sequence computed by the computer [in my terminology: comput*or*].[70]

Thus, shifting back to computations of number-theoretic functions, we have *Turing's Theorem*: Any number-theoretic function F that can be computed by a computor, satisfying the determinacy condition (D) and the conditions (1.1)–(2.2), can be computed by a Turing machine.

Both Gödel and Church state they were convinced by Turing's analysis that the identification of effective calculability with Turing computability (thus also with recursiveness and λ-definability) is correct. Church expressed his views in the 1937 review of Turing's paper, from which I quoted in the introduction: On account of Turing's work the identification is "immediately evident." As to Gödel, I have not been able to find in his papers any reference to Turing's analysis before 1946; that paper was discussed in section 2.4. But what did convince them? Gödel gave some indication in the postscriptum to his Princeton lectures, where he is perfectly clear about the *structure* of Turing's argument. "Turing's work," he writes, "gives an analysis of the concept 'mechanical procedure' (alias 'algorithm' or 'computation procedure' or 'finite combinatorial procedure'). This concept is *shown* [my emphasis] to be equivalent with that of a 'Turing machine'." In a footnote attached to this observation he calls "previous equivalent definitions of computability"—referring to λ-definability and recursiveness—"much less suitable for our purpose." What is not elucidated by any remark of Gödel, as far as I know, is the *result* of Turing's analysis, that is, the axiomatic formulation of restrictive conditions. And there is consequently no discussion of the *reasons* for the correctness of these conditions or, for that matter, of the analysis.

Church was very much on target in his review, though there is a misunderstanding as to the relative role of the human computor and machine computability in Turing's argument. For Church, computability by a machine "occupying a finite space and with working parts of finite size" is analyzed by Turing; one then can observe that "in particular, a human calculator, provided with pencil and paper and explicit instructions, can be regarded as a kind of Turing machine" (Church 1937a). On account of the analysis and this observation for Church it is then "immediately clear" that (Turing) machine computability can be identified with effectiveness. This is reemphasized in the rather critical review of Post's 1936 paper in which Church points to the essential finiteness requirements in Turing's analysis: "To define effectiveness as computability by an arbitrary machine, subject to restrictions of finiteness, would seem to be an adequate representation of the ordinary notion, and if this is done the need for a working hypothesis disappears" (1937b). This is right,

as far as emphasis on finiteness restrictions is concerned. But Turing analyzed, as we saw, a mechanical computor, and *that* provides the basis for judging the correctness of the finiteness conditions.

Church's apparent misunderstanding is rather common; cf. note 4 on Mendelson (1990). So it is worthwhile to point out that machine computability was analyzed only much later, by Gandy (1980). Turing's three-step procedure of analysis, axiomatic formulation of general principles, and proof of a "reduction theorem" is followed there, but for "discrete deterministic mechanical devices." Gandy showed that everything computable by a device satisfying the principles—a "Gandy machine"—can already be computed by a Turing machine; cf. Sieg (1989). To see clearly the difference between Turing's and Gandy's analysis, note that Gandy machines incorporate parallelism. They compute directly, for example, Conway's Game of Life, and thus violate the basic assumption that mechanical computors operate only on symbolic configurations of bounded size. Furthermore, the different boundedness conditions for Gandy machines (in particular, the principle of local causation) are motivated not by limitations of the human sensory apparatus, but by physical considerations.

3.3. *Turing's Thesis*

Turing's analysis and his theorem can be generalized by making an observation concerning the determinacy condition: (D) is not needed to guarantee the Turing computability of F in the theorem. More precisely, (D) was used in conjunction with (1.1) and (1.2) to argue that computors can carry out only finitely many operations; this claim follows already from conditions (1.1)–(2.2) *without* appealing to (D). Thus, the behavior of computors can still be fixed by a finite list of commands, but it may exhibit nondeterminism. Such computors can be mimicked by nondeterministic Turing machines and thus, exploiting the reducibility of nondeterministic to deterministic machines, by deterministic Turing machines.

This observation is by no means difficult, but it is intellectually pleasing that it allows one to connect in a straightforward way Turing's considerations with those of Church discussed in section 2.3. Consider an effectively calculable function F and a nondeterministic computor who calculates—in Church's sense—the value of F in a logic L. Using the generalized theorem and the fact that Turing computable functions are recursive, F is then recursive. This argument for F's recursiveness does no longer appeal to Church's Thesis, not even to the more restricted Central Thesis; rather, such an appeal is replaced by the assumption that the calculation in the logic is done by a computor satisfying conditions (1.1)–(2.2). Indeed, any system satisfying these axiomatic conditions would do. Turing's analysis thus leads to a result that is in line with Gödel's general methodological expectations expressed to Church in 1934 (and reported by Church to Kleene in 1935): "His [i.e. Gödel's] only idea at the time was that it might be possible, in terms of effective calculability as an undefined notion, to state a set of axioms which would embody the generally accepted properties of this notion, and to do something on that basis."[71]

If the assumption in the argument for the recursiveness of F is to be discharged, then a substantive thesis is needed. And it is this thesis that I want to call *Turing's Thesis*. It says that a mechanical computor satisfies the finiteness conditions (1.1) and (1.2), and that the elementary operations the computor can carry out are restricted as conditions (2.1) and (2.2) require. In short, if the clarification of effective calculability as meaning computability by a mechanical computor is accepted, then Turing's Thesis is the final piece to guarantee the equivalence of that notion and recursiveness. And *if* Turing's Thesis is correct, then the conceptual problem of von Neumann is resolved, because we have a precise and—via Turing's Theorem—mathematically handy characterization of *absolut mechanische Vorschriften*. Gödel, then, was also right when concluding that Turing computability captures the "essence" of formal systems, namely, "that reasoning is completely replaced by mechanical operations on formulas."[72]

The first section of this chapter had the explicit purpose of describing the context for the investigations of Herbrand, Gödel, Church, Kleene, and Turing. There is no doubt, it seems to me, that an analysis of human computability on finite (symbolic) configurations was called for, and that the epistemological restrictions were cast in "mechanical" terms; see the remarks of Frege and Gödel quoted in section 1.1 as striking examples. Thus, Turing's clarification of effective calculability as *calculability by a mechanical computor* should be accepted. Two related issues remain: first, the question whether the thesis is correct and, second, Turing's claim that its ultimate justification lies in the necessary limitation of human memory. As to the first issue, we have to ask ourselves whether the restrictive conditions do in fact apply to mechanical computors. According to Gandy, Turing arrives at the restrictions "by considering the limitations of our sensory and mental apparatus." However, only limitations of our sensory apparatus seem to be involved, unless "state of mind" is given a mental touch. That is technically unnecessary, as I pointed out in section 3.1, *and* is not central to Turing: In section 9(III) of his paper he describes a modified computor and avoids introducing "state of mind," considering instead "a more physical and definite counterpart of it."[73] Without discussing this modification, whose context is a little complex, the analysis appeals only to sensory limitations of the type I discussed at the beginning of section 3.2. Such limitations are operative when we work as purely mechanical computors.[74]

Turing sees memory limitations as ultimately justifying the restrictive conditions. But none of the conditions is motivated by such a limitation; so how are we to understand his claim? I suggest the following: If our memory were not subject to limitations of the same character as our sensory apparatus, we could scan (with the limited sensory apparatus) a symbolic configuration that is not immediately recognizable, read in sufficiently small parts so that their representations could be assembled in a unique way to a representation of the given symbolic configuration, and, finally, carry out (generalized) operations on that representation in memory. Is one driven to accept Turing's assertion as to the limitation of memory? I suppose so, if one thinks that information concerning symbolic structures is physically encoded and that there is a bound on the number of available codes.

Turing viewed his argument for the identification of effectively calculable functions with functions computable by his machines as being basically "a direct appeal to intuition." Indeed, he claimed more strongly, "All arguments which can be given [for this identification] are bound to be, fundamentally, appeals to intuition, and for that reason rather unsatisfactory mathematically."[75] If we look at his paper on ordinal logics (Turing 1939), the claim that such arguments are "unsatisfactory mathematically" becomes at first rather puzzling, because he observed that intuition is inextricable from mathematical reasoning. Turing's concept of intuition is much more general than that ordinarily used in the philosophy of mathematics. It was introduced in Turing's 1939 paper explicitly to address the general issues raised by Gödel's first incompleteness theorem in the context of work on ordinal logics or what was later called progressions of theories; the discussion is in Davis (1965, section 11, 208–210):

> Mathematical reasoning may be regarded rather schematically as the exercise of a combination of two faculties, which we may call *intuition* and *ingenuity*. The activity of the intuition consists in making spontaneous judgements which are not the result of conscious trains of reasoning. These judgements are often but by no means invariably correct (leaving aside the question of what is meant by "correct"). . . The exercise of ingenuity in mathematics consists in aiding the intuition through suitable arrangements of propositions, and perhaps geometrical figures or drawings. It is intended that when these are really well arranged the validity of the intuitive steps which are required cannot seriously be doubted.

It seems to me that the propositions in Turing's argument are arranged with sufficient ingenuity so that "the validity of the intuitive steps which are required cannot seriously be doubted"; or, at least, their arrangement allows us to point to the central conditions, with clearer, adjudicable content than Church's normative Central Thesis.

4. Aspects of Mathematical Experience

Let me move from the details of the conceptual analysis to its use for an interpretation of the incompleteness theorems. If their formulation and their interpretation are to be general, the relation of Turing computability to effective calculability, and the informal understanding of the latter notion, have to come to the fore. I argued that historically the insistence on formality was motivated by epistemological concerns; it is quite clear that a genuine restriction on our cognitive, more particularly, mathematical capacities was intended. Thus, it may be surprising that some of the pioneers interpreted these results, prima facie, in a quite dramatic way. For example, Post (1936) emphasized that the theorems I mentioned exemplify "a fundamental discovery in the limitations of the mathematizing power of Homo Sapiens." Later he remarked with respect to these results: "Like the classical unsolvability proofs, these proofs are of unsolvability by means of given instruments. What is new is that in the present case these instruments, in effect, seem to be the only instruments at man's disposal."[76]

The last part of this chapter will start out with reflections of Gödel on these issues. His considerations will lead me to look at two aspects of mathematical experience. The first, the *quasi-constructive* aspect, has to do with the recognition of laws for "accessible domains"; this includes, in particular, our recognition of set-theoretic axioms and their extendability by suitable axioms of infinity. The second, the *conceptional* aspect, deals with the uncovering of abstract, axiomatically characterized notions. Turing's analysis shows, it seems to me, that calculability exemplifies the second aspect.

4.1. Gödel's Consequences

Turing's work provides "a precise and unquestionably adequate definition of the general concept of formal system"; consequently, the incompleteness theorems hold for *arbitrary formal* systems (satisfying the usual conditions). But, in contrast to Post, for Gödel they do not establish "any bounds for the powers of human reason, but rather for the potentialities of pure formalism in mathematics" (Davis 1965, 72–73). In Gödel (1972b) there is a discussion of a "philosophical error in Turing's work" that can, according to Gödel, be regarded as a footnote to the word "mathematics" in this very quotation. Gödel claims that Turing gives an argument, on page 136 of Davis (1965), to show that "mental procedures cannot go beyond mechanical procedures". But Turing sketches there only an argument showing that "the number of states of mind that need be taken into account is finite." The context makes it crystal clear that mechanical procedures are being analyzed. Thus, I cannot see a philosophical error in Turing's work, but rather in Gödel's interpretation.[77] However, the further remarks in Gödel's note are of interest independent of Gödel's error; they summarize points he had made more extensively in his Gibbs Lecture of 1951.

In that lecture Gödel stated that if mathematics is viewed as a body of propositions that "hold in an absolute sense," then the incompleteness theorems express the fact that mathematics is not exhaustible by a mechanical enumeration of its theorems. After all, the first theorem yields, for any consistent formal system *S* containing a modicum of number theory, a simple arithmetic sentence that is independent of *S*. But Gödel (1951, 5–6) emphasized that it is the second theorem that makes this phenomenon of inexhaustibility particularly evident:

> For it makes it impossible that someone should set up a certain well-defined system of axioms and rules and consistently make the following assertion about it: All of these axioms and rules I perceive (with mathematical certitude) to be correct, and moreover I believe that they contain all of mathematics.

It someone claims this, he contradicts himself: Recognizing the correctness of all axioms and rules means recognizing the consistency of the system. Thus, a mathematical insight has been gained that does not follow from the axioms. To explain the meaning of this situation Gödel carefully distinguished between "objective" and "subjective" mathematics: objective mathematics is viewed as the body of true mathematical propositions; subjective mathematics is conceived of as consisting of all humanly provable mathematical propositions.

There clearly cannot be a complete formal system for objective mathematics, but it is not excluded that for mathematics in the subjective sense there might be a finite procedure yielding all of its evident axioms. Clearly, we could never be certain that all of these axioms are correct; but if there were such a procedure, then—at least as far as mathematics is concerned—the human mind would be equivalent to a Turing machine. Furthermore, there would be simple arithmetic problems that could not be decided by any mathematical proof intelligible to the human mind. If, according to Gödel (1951, 7), we call such a problem *absolutely undecidable* we have established with full mathematical rigor that *either* mathematics is inexhaustible in the sense that its evident axioms cannot be generated by a finite procedure *or* there are absolutely undecidable arithmetic problems.

This fact appears to Gödel to be of "great philosophical interest." That is not surprising, as he explicates the first alternative in the following way: "... that is to say, the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine." The further philosophical consequences Gödel tries to draw are concerned with his Platonism, familiar from some of his published writings. In 1933 Gödel already claimed that the axioms of set theory, "if interpreted as meaningful statements, necessarily presuppose a kind of Platonism." But at that time he added the relative clause "which cannot satisfy any critical mind and which does not even produce the conviction that they are consistent" (p. 7). I would go too far afield if I tried to present the reasons why I do not find Gödel's general considerations (in the Gibbs Lecture) convincing. My criticism would not start with his Platonism for set theory, but at the point where he contrasts the objects of finitist and intuitionistic mathematics in his *Dialectica* paper. There he tried to draw an extremely sharp distinction within constructive mathematics that seems to me to be mistaken (and to parallel his equally mistaken radical opposition of classical and constructive mathematics). According to Gödel (1958, 240), the specifically finitist character of mathematical objects requires them to be "finite space-time configurations whose nature is irrelevant except for equality and difference"; furthermore, in proofs of propositions concerning them, one uses only insights that derive from the combinatorial space-time properties of sign combinations representing them.[78] These remarks stand in conflict with Bernays' position, to which Gödel appealed in his *Dialectica* paper. Bernays stressed the uniform character of the generation of natural numbers, the local structure of the schematic "iteration figure," and the need to "*reflect* on the general features (allgemeine Charakterzüge) of intuitive objects" (1930). Indeed, our understanding of natural numbers as being generated in such a uniform way allows us to grasp laws concerning them. It seems to me that this observation is correct also for more general inductively generated classes, and it points to the first of two critical aspects of mathematical experience I want to describe now.

4.2. Accessibility and Conception

If one takes seriously the reformulation of the first alternative in Gödel's Main Theorem, then one certainly should try to see ways in which the human mind transcends" the limits of mechanical computors. Gödel (1972b) suggested that there may be (humanly) effective but nonmechanical procedures. But even the most specific of his proposals, Gödel admitted, "would require a substantial advance in our understanding of the basic concepts of mathematics." That proposal concerned the extension of systems of axiomatic set theory by axioms of infinity, that is, extending segments of the cumulative hierarchy. The problem of extending what I call *accessible domains* is not special to the case of set theory (and Platonism); rather, there are completely analogous issues for the theory of primitive recursive functionals (and finitism) and for the theory of constructive ordinals in the second number class (and intuitionism). This is the first of the two aspects of mathematical experience on which I want to focus; both are related to features of "mental procedures" Gödel discussed, but their interest is quite independent of Gödel's speculations.

Accessible domains, constituted by inductively generated elements, are most familiar from mathematics and logic. In proof theory, for example, inductively defined higher constructive number classes have been used in consistency proofs for impredicative subsystems of analysis. These and other classes provide special cases in which generating procedures allow us to grasp the intrinsic buildup of mathematical objects. Such an understanding is a fundamental source of our knowledge of mathematical principles for the domains constituted by them; for it is the case, I suppose, that the definition and proof principles for such domains follow directly from the comprehended buildup. A broad framework for the "inductive or rule-governed generation" of mathematical objects is described by Aczel (1977); it is indeed so general that it encompasses not only finitary i.d. classes, higher number classes, and models of a variety of constructive theories, but also segments of the cumulative hierarchy. It provides a uniform framework in which the difficulties (in our understanding) of generating procedures can be compared and explicated. If we understand the set-theoretic generation procedure for a segment of the cumulative hierarchy, then it is indeed the case that the axioms of ZF^- (i.e., ZF without the postulate for the existence of the first infinite ordinal), together with a suitable axiom of infinity, "force themselves upon us as being true" in Gödel's famous phrase; they simply formulate the principles underlying the "construction" of the objects in this segment.[79]

The sketch of this quasi-constructive aspect of mathematical experience is extremely schematic and yet, I think, helpful for further orientation. Recall that for Dedekind consistency proofs were intended to ensure that axiomatically characterized notions (like that of a complete ordered field) were free from "internal contradictions." Here we are dealing with *abstract* notions without an "intended model" constituted by *inductively generated* elements.[80] These notions are distilled from mathematical practice for the purpose of comprehending complex connections, of making analogies precise, and of obtaining a more profound understanding. It is in this way that the axiomatic method teaches us, as Bourbaki (1950) expressed it in Dedekind's spirit,

> to look for the deep-lying reasons for such a discovery [that two, or several, quite distinct theories lend each other "unexpected support"], to find the common ideas of these theories, . . . to bring these ideas forward and to put them in their proper light. (p. 223)

Notions like group, field, topological space, and differentiable manifold are abstract in this sense; They are properly and in full generality investigated in category theory. Another example of such a notion is that of Turing's mechanical computor. Although Gödel used "abstract" in a more inclusive way than I do here, it seems that the notion of computability exemplifies his broad claim (1972b, 306), "that we understand abstract terms more and more precisely as we go on using them, and that more and more abstract terms enter the sphere of our understanding."

This conceptional aspect of mathematical experience and its profound function in mathematics has been entirely neglected in the logico-philosophical literature on the foundations of mathematics—except in the writings of Paul Bernays. Indeed, detailed investigations of these two aspects of mathematical experience can be viewed as addressing the central problem expressed by Bernays in the quotation from his 1922 work given in section 1.2: Which principled position can we take regarding the "transcendent assumptions" of mathematics? Those assumptions are reflected through accessible domains, relative to which abstract notions can be shown to be consistent via structural reductions. Here we have a generalized and redirected Hilbert program that mediates between Richard Dedekind and a liberalized version of Leopold Kronecker! The traditional contrast between "Platonist" and "constructivist" tendencies in mathematics comes to light here in refined distinctions concerning the admissibility of operations, of their iteration, and of deductive principles. The considerations on the quasi-constructive aspect of mathematical experience cut across traditional "school" boundaries; so do those on its conceptional aspect.

5. Final Remarks

I argued that the sharpening of axiomatic theories to formal ones was motivated by epistemological concerns. A central point was the requirement that the checking of proofs ought to be done in a radically intersubjective way; it should involve only operations similar to those used by a computor when carrying out an arithmetic calculation. Turing analyzed the processes underlying such operations and formulated a notion of computability by means of his machines; that was in 1936. In a paper written about ten years later and entitled "Intelligent Machinery" Turing (1948, 21) stated what really is the central problem of cognitive psychology:

> If the untrained infant's mind is to become an intelligent one, it must acquire both discipline and initiative. So far we have been considering only discipline [via the universal machine, W.S.]. . . . But discipline is certainly not enough in itself to produce intelligence. That which is required in addition we call initiative. This statement will have to serve as a definition. Our task is to discover the nature of this residue as it occurs in man, and to try to copy it in machines.

The task of copying is difficult, some would argue impossible, in the case of mathematical thinking. But before we can start copying, we have to discover—at least partially—"the nature of the residue." As you may recall, Turing distinguished between ingenuity and intuition in his 1939 paper, and he argued that in formal logics their respective roles take on a greater definiteness: intuition is used for "setting down formal rules for inferences which are always intuitively valid"; ingenuity, to "determine which steps are the more profitable for the purpose of proving a particular proposition." He noted, "In pre-Gödel times it was thought by some that it would be possible to carry this programme to such a point that all the intuitive judgements of mathematics could be replaced by a finite number of these rules. The necessity for intuition would then be entirely eliminated" (1939, 209).

Proofs in a formal logic can be obtained uniformly by a (patient) search through an enumeration of all theorems, but additional nonmechanical, intuitive steps remain necessary because of the incompleteness theorems. Turing suggested particular kinds of intuitive steps in his ordinal logics; his arguments are utterly theoretical, but connect directly to the discussion of actual or projected computing devices that appear in his "Lecture to London Mathematical Society" (1986) and in "Intelligent Machinery" (1948). In these papers he calls for "intellectual searches" (i.e., heuristically guided searches) and "initiative" (that includes, in the context of mathematics, proposing intuitive steps). However, Turing (1986, 122) emphasizes that "As regards mathematical philosophy, since the machines will be doing more and more mathematics themselves, the centre of gravity of the human interest will be driven further and further into philosophical questions of what can in principle be done etc." Thus we are straightforwardly led back to the questions: What are essential aspects of mathematical experience? Are they mechanizable?

I have tried to give a very tentative and partial answer to the first question. As far as the second question is concerned, I don't have even a conjecture on how it will be answered. Is Gödel's search for humanly effective but non-mechanical procedures in mathematics more than searching for a "pie in the sky" (as Kleene thinks)? Or is Post, drawing on similar mathematical facts, right when making the observation:

> The creative germ . . . can be stated as consisting in constructing ever higher types. These are as transfinite ordinals and the creative process consists in continually transcending them by seeing previously unseen laws which give a sequence of such numbers. Now it seems that this complete seeing is a complicated process mostly subconscious. But it is not given till it is made completely conscious. But then it ought to be constructable [*sic*] purely mechanically.[81]

Whatever the right answers may be, mathematical experience represents an extremely important component of Turing's problem, and we should investigate crucial aspects vigorously: by historical case studies, theoretical analysis, psychological experimentation, and—quite in Turing's open spirit—by machine simulation.

Appendix

This appendix uses some new documents to further elucidate significant conceptual issues and to support conjectural remarks pertaining to the impact of (the proof of) Gödel's incompleteness theorems on Herbrand and Church. Incidentally, both men got to know Gödel's results through Johan von Neumann; Herbrand in November 1930 in Berlin, Church about a year later in Princeton. With respect to Herbrand I want to emphasize, as I did in section 2.2, that he was concerned with the notion of *provably* recursive function; as to Church, I want to stress that his belief in the correctness of his thesis hardly rested on any particular "motivation" for λ-definability, but rather on general facts concerning the notion of "calculability in a logic" and on his Central Thesis. In any event, there still is extremely interesting material to be uncovered and evaluated; there also remains a great deal of important analytical work to be done. Gödel's proof provided the seminal idea of representing number-theoretic functions in a formal system; his results provided the stimulus for investigations concerning their proper applicability and the precise extension of effectiveness. How surprising his results were (for logicians) is sometimes no longer appreciated; consider Herbrand's reaction, described in his letter of December 3, 1930, to his friend Claude Chevalley:

> Les mathématiciens sont une bizarre chose; voici une quinzaine de jours que chaque fois que je vois [von] Neumann nous causons d'un travail d'un certain Gödel, qui a fabriqué de bien curieuses fonctions; et tout cela détruit quelques notions solidement ancreés.

This sentence opens the letter; after having sketched Gödel's arguments and reflected on the results Herbrand concludes it with: "*Excuse ce long début; mais tout cela me poursuit, et de l'écrire m'en exorcise un peu.*"

1. If, as I described in sections 2.1 and 2.2, Gödel took off in a generalizing mood from Herbrand's schema for the introduction of "recursive" functions, and if Herbrand was not attempting to characterize a general notion of effectively calculable function, what did motivate Herbrand to formulate the schema? First recall the widely shared assumptions, namely: (1) the general notion of recursive function was captured by Herbrand's schema, and (2) the schema emerged from Herbrand's general reflections on intuitionistic methods. These assumptions are formulated, for example, by van Heijenoort (1971, 283), but also by Dawson (1991):

> The functions [characterized by the schema formulated in Herbrand (1931d), W.S.] are, in fact, (general) recursive functions, and here is the first appearance of the notion of recursive (as opposed to primitive recursive) function. It is interesting to see how, a few months earlier, Herbrand had been led to this notion by his conception of "intuitionism."

For the earlier discussion van Heijenoort refers to Herbrand's note (1931b), and, with respect to Herbrand (1931d), he writes:

Herbrand's consistency proof for a fragment of arithmetic still belongs to the period that preceded Gödel's famous result (1931). He probably started to write his paper before Gödel's paper reached him. But he had ample opportunity to examine Gödel's result and he wrote a last section dealing with it.

This scenario is incorrect: The notes (1931b and c) and the paper (1931d) were all written *after* Herbrand knew quite well about the incompleteness theorems. This seems to be clear from internal evidence, but Herbrand's letter to Chevalley puts it beyond any doubt. In it Herbrand tells us (1) that it was von Neumann from whom he learned of Gödel's theorems, and (2) that the encounters with von Neumann took place in the second half of November 1930. That new information also puts into sharper focus the remark in Herbrand (1931c, 279) that was submitted, according to the introduction by Goldfarb, to Hadamard "at the beginning of 1931":

> Recent results (not mine) show that we can hardly go any further: it has been shown that the problem of consistency of a theory containing all of arithmetic (for example, classical analysis) is a problem whose solution is impossible. [Herbrand is here alluding to Gödel 1931.] In fact, I am at the present time preparing an article in which I will explain the relationships between these results and mine [this article is 1931d].[82]

Thus, it is Herbrand's attempt to come to a thorough understanding of the relationship between Gödel's incompleteness theorems and his own work that seems to have prompted the specific details in his letter to Gödel and his paper (1931d). Indeed, I think that Herbrand's proposal for the introduction of functions is a natural generalization of the definition schemata for effectively calculable functions known to him and that it emerges quite directly from his way of proving consistency of (weak) systems of arithmetic already in his thesis. In the note to Bernays that accompanied the copy of his letter to Gödel, Herbrand contrasts his consistency proof with that of Ackermann:

> In my arithmetic the axiom of complete induction is restricted, but one may use a variety of other functions than those that are defined by simple recursion: in this direction, it seems to me, that my theorem goes a little farther than yours.[83]

This is hardly a description of a class of functions that is deemed to be of fundamental significance! However, a detailed account of the evolution of Herbrand's schema, as well as the precise characterization of the provably total functions of Herbrand's system of arithmetic (1931d), has to wait for another occasion.

2. Kleene (1987a, 491) emphasized that the approach to effective calculability through λ-definability had "quite independent roots (motivations)" and would have led Church to his main results "even if Gödel's paper (1931) had not already appeared." Perhaps Kleene is right, but I doubt it. The flurry of activity surrounding Church's *A Set of Postulates for the Foundation of Logic* (published in 1932 and 1933) is hardly imaginable without knowledge of Gödel's work, in particular, not without the central notion of representability and, as Kleene pointed out, the arithmetization of metamathematics. Since the fall of 1931 the Princeton group of Church, Kleene, and Rosser knew of Gödel's theorems

through a lecture of von Neumann: Kleene (1987a, 491) reports that through this lecture "Church and the rest of us first learned of Gödel's results" (cf. also Rosser 1984). The centrality of representability for Church's considerations comes out quite clearly in his lecture on Richard's Paradox given in December 1933 and published in 1934. According to Kleene (1981, 59), Church already had formulated his thesis for λ-definability in the fall of 1933, so it is not difficult to read the following statement as an extremely cautious statement of the thesis (Church 1934, 358):

> . . . it appears to be possible that there should be a system of symbolic logic containing a formula to stand for every definable function of positive integers, and I fully believe that such systems exist.[84]

One has only to realize from the context that (1) "definable" means "constructively definable," so that the value of the function can be calculated, and (2) that "to stand for" means "to represent." In a letter to Bernays dated January 23, 1935, Church claims explicitly that the λ-calculus may be a system that allows the representability of all constructively defined functions:

> The most important results of Kleene's thesis concern the problem of finding a formula to represent a given intuitively defined function of positive integers (it is required that the formula shall contain no other symbols than λ, variables, and parentheses). The results of Kleene are so general and the possibilities of extending them apparently so unlimited that one is led to conjecture that a formula can be found to represent any particular constructively defined function of positive integers whatever. It is difficult to prove this conjecture, however, or even to state it accurately, because of the difficulty in saying precisely what is meant by "constructively defined." A vague description can be given by saying that a function is constructively defined if a method can be given by which its values could be actually calculated for any particular positive integer whatever. Every recursive definition, of no matter how high an order, is constructive, and as far as I know, every constructive definition is recursive.

One and a half years later, Church sent Bernays a copy of a letter he had written on June 8, 1937, to the Polish logician Jozef Pepis. Pepis earlier had informed Church about his project of constructing a numerical function that is effectively calculable, but not general recursive. In his response, Church confessed himself to be "extremely skeptical—although this attitude is of course subject to the reservation that I may be induced to change my opinion after seeing your work." Church stated his impression that Pepis might "not fully appreciate the consequences which would follow from the construction of an effectively calculable non-recursive function" and went on to formulate such consequences—giving a most sophisticated defense of Church's Thesis by an argument that makes implicit use of a concept close to Gödel's notion of absoluteness:

> For instance, I think we may assume that we are agreed that *if a numerical function* f *is effectively calculable then for every positive integer* a *there must be a positive integer* b *such that a valid proof can be given of the proposition* f(a) = b (at least *if we are not agreed on this then our ideas of effective calculability are so*

different as to leave no common ground for discussion) [my emphasis]. But it is proved in my paper in the *American Journal of Mathematics*[85] that if the system of *Principia Mathematica* is omega-consistent, and if the numerical function f is not general recursive, then, whatever permissible choice is made of a formal definition of f within the system of *Principia*, there must exist a positive integer a such that for no positive integer b is the proposition $f(a) = b$ provable within the system of *Principia*. Moreover, this remains true if instead of the system of *Principia* we substitute any one of the extensions of *Principia* which have been proposed (e.g., allowing transfinite types), or any one of the forms of the Zermelo set theory, or indeed any system of symbolic logic whatsoever which to my knowledge has ever been proposed.

Because of the metamathematical facts and the assumed minimal agreement on effective calculability, Church concludes that to discover an effectively calculable nonrecursive function "would imply discovery of an utterly new principle of logic, not only never before formulated, but never before actually used in a mathematical proof—since all extant mathematics is formalizable within the system of *Principia*, or at least within one of its known extensions." Yet the final line of defense is what I called Church's Central Thesis:

> Moreover, this new principle of logic must be of so strange, and presumably complicated, a kind that its metamathematical expression as a rule of inference was [sic!] not general recursive (for this reason, if such a proposal of a new principle of logic were ever actually made, I should be inclined to scrutinize the alleged effective applicability of the principle with considerable care).

Substantiating my claim of the "dependent" development in Princeton requires further detailed (historical) work I am not going to pursue here; but I add that this is not an issue of settling priorities, but of elucidating the role of central informal notions.

Notes

Acknowledgements—It is a pleasure to acknowledge that the considerations in parts 2 and 3 have profited from discussions with Guglielmo Tamburrini and Robin Gandy. As to the published literature, the original papers of Gödel, Church, Turing, Post, and Kleene, and the essays of Gandy (1980, 1988), were particularly important to me; I also learned a great deal from the historical papers of Davis (1982) and Kleene (1981, 1988). To Clark Glymour I owe thanks for bringing Herbrand's remarkable letter to Claude Chevalley to my attention, and to Catherine Chevalley for permission to quote from it. Thanks for improvements are due to C. Bartusis, A. Behboud, J. Byrnes, A. Ignjatovic, G. H. Müller, P. Odifreddi, T. Seidenfeld, and, especially, to J. W. Dawson. Finally, I am grateful to the editors of Gödel's *Collected Works* (Solomon Feferman, John W. Dawson, Warren Goldfarb, and Charles D. Parsons) for access to the correspondence between Herbrand and Gödel, Bernays and Gödel, and to Gödel's as yet unpublished lectures of 1933 and 1951. The Institute for Advanced Studies at Princeton gave me permission to quote from this unpublished material; similarly, the Bibliothek der ETH, Zürich, allowed me to use Bernays' correspondence with Church, Herbrand, and Turing.

1. From Wittgenstein (1980), section 1096. I first read this remark in Shanker (1987), where it is described as a "mystifying reference to Turing machines."

2. The speculation is taken up briefly in the last section of this chapter and, in detail, in my paper with Tamburrini, "Does Turing's Thesis Matter?"

3. But see Tamburrini (1988), and the critical survey of the literature given there. The present paper is part of a book project Tamburrini and I have been pursuing for a number of years. A detailed review of the classical arguments is in Kleene (1954), sections 62, 63, and 70; section 6.4 of Shoenfield (1967) also contains a careful discussion of Church's Thesis; and, finally, the first chapter of Odifreddi (1989) provides a broad perspective for the whole discussion.

4. Mendelson (1990) intends "to renounce the standard views concerning the nature of Church's thesis" and concludes that the thesis is *true* on account of "Turing's analysis of the essential elements involved in computation" (p. 233). Very standardly, however, he emphasizes (1) that the "independently proposed" explications of Church, Post, and Turing are "quite different," and (2) that Turing used his machines directly as mathematical models "to capture the essence of computability." (The real target of) Turing's analysis and the source of the restrictive, "normalizing assumptions" for Turing machine computations are not mentioned at all; see section 3.2.

5. Turing (1939); see the reprint in Davis (1965, 160). I want to warn the reader against misinterpretations of Turing's Thesis by "mechanists"—as in Webb (1980, 9), where it is claimed that it is a very strong thesis indeed, "for it says that *any* effective procedure whatever, using whatever 'higher cognitive processes' you like, is after all mechanizable"—but also against the misunderstanding of the thesis and an emphasis on absolutely misleading issues by "anti-mechanists"—as in Searle (1990); in particular pp. 24–28. On p. 26, for example, Searle claims that the standard definition of "digital computer", which he traces back to Turing, seems to imply: "For any object there is some description of that object such that under that description the object is a digital computer."

6. Here and in section 4.2 I draw on my paper (1990a) and refer to it for additional and relevant details. For a comprehensive discussion of Leibniz's views, see Spruit and Tamburrini (1991); Krämer (1988) traces the historical development of calculi in a very informative way. Note that I focus here on the—for my purposes—most relevant background and do not discuss, for example, Babbage's (theoretical) work; for that see Gandy (1988).

7. There is much mathematical work, partly related to proof theory, that started with Weyl's "Das Kontinuum" and early lectures of Hilbert's presented in the second volume of *Grundlagen der Mathematik*. During the last decade important and most relevant work was done in "reverse mathematics"; see my review (1990b).

8. Van Heijenoort (1967, 101).

9. P. 237. But he was careful to emphasize (in other writings) that all of thinking "can never be carried out by a machine or be replaced by a purely mechanical activity," Frege (1969, 39). He went on to claim: "Wohl läßt sich der Syllogismus in die Form einer Rechnung bringen, die freilich auch nicht ohne Denken vollzogen werden kann, aber doch durch die wenigen festen und anschaulichen Formen, in denen sie sich bewegt, eine grosse Sicherheit gewährt."

10. He added parenthetically: "This has the consequence that there can never be any doubt [as] to what cases the rules of inference apply, and thus the highest possible degree of exactness is obtained."

11. Another significant influence was the sharpening of the "hypothetico-deductive method" within mathematics; a sharpening that brought about a separation of syntax

semantics for mathematical theories. This separation was clear to Dedekind, for example. Wiener's talk at the meeting of the Deutsche Mathematiker Vereinigung in Halle, succinctly summarized in his Wiener (1891), made this methodological point very forcefully and impressed Hilbert strongly. For this development compare Guillaume (1985, 766–777).

12. Clearly, an adequate representation of content was taken for granted. Cf. Kreisel (1968) for the following discussion.

13. I assume that Herbrand had in mind Dirichlet's theorem mentioned previously. See also Herbrand (1930, 187), where he hopes that his approach will allow the elimination of "transcendental methods" from proofs of arithmetic theorems.

14. Support for this claim is given in Sieg (1990a, 271–272). The "mediating" role of the program was not only described by the immediate members of the Göttingen school, but also, for example, by Herbrand; see Herbrand (1971, 211–212).

15. These connections are elaborated in Sieg (1984); as to the lively interest in Kronecker's ideas in Germany in the twenties, see Pasch (1918) and Kneser (1925). When describing the central features of "intuitionist" mathematics, for example in 1931a and c, Herbrand emphasized exactly Kronecker's points; see Herbrand (1971, 273, and footnote 5, 288–289).

16. I will come back to these remarks in section 4.2.

17. That is expressed in the Nachtrag to Bernays (1930), reprinted in Bernays (1976, 61). In a letter to Gödel, written on September 7, 1942, Bernays emphasized that the methodological points of the above character do not correspond to a "strict *formalist* standpoint"; he continued: "... aber einen solchen habe ich niemals eingenommen, insbesondere habe ich mich in meinem (Sommer 1930 geschriebenen) Aufsatz 'Die Philosophie der Mathematik und die Hilbertsche Beweistheorie' deutlich davon distanziert, und noch mehr dann in dem (Ihnen wohl bekannten) Vortrag 'Sur le platonisme dans les mathématiques'."

18. Herbrand (1930a, 188); the same point is made in (1930b, 214).

19. Herbrand thought that this assumption was not restrictive: "And in general," he wrote (1930b, 213), "we can contrive so as to make all usual mathematical arguments in theories that have only a determinate finite number of hypotheses. Thus we can see the importance of this problem, whose solution would allow us to decide with certainty with regard to the truth of a proposition in a determinate theory."

20. In Gandy (1988, 64–65), one finds the remark that this idea of requiring bounds turns up over and over "like a bad penny"; but in the context of the issues Herbrand and others were working on it is a most natural constructivity requirement. However, and here I agree with Gandy, in a general theory of computability there is no good reason to "mix together constructive and nonconstructive notions of existence." This point will come up again in the discussion of Gödel's notion of general recursive function that was based, as Gödel put it, on a suggestion of Herbrand.

21. Löwenheim's work on the decision problem was done in the Boole–Schröder tradition of algebraic logic. He had established results that could be used to obtain (partial) answers to the decision problem also for Frege's *Begriffsschrift*; namely, he solved the problem for monadic predicate logic and reduced that for full predicate logic to the fragment with just binary predicates. Independently, Behmann (1922) proved these results directly for a system of symbolic logic building on the work of Frege, and Whitehead and Russell.

22. Herbrand (1930a, 176), compare also Herbrand (1929b, 42).

23. That was already explicit in Löwenheim (1915); see van Heijenoort (1967, 246). Cf. also Herbrand (1930b, 207), where Herbrand speaks of an "experimental certainty"

that *Principia Mathematica* allows the representation of all mathematical statements and arguments. That point was made forcefully also in Herbrand (1930a, 48).

24. Hilbert (1927), translated in van Heijenoort (1967, 475).

25. This observation extends to the interpretation of the incompleteness and undecidability theorems of Gödel, Church, and Turing. The most striking and contentious "consequence" of these particular metamathematical results is briefly formulated as: Minds are (not) machines. As to the unnegated statement, see Myhill (1952) and Webb (1980, 1990); the negated statement has been defended in, for example, Löb (1959) and Lucas (1961).

26. Herbrand and Church reacted to the incompleteness theorems in the same way: "They can't apply to my formalism!" As to Gödel's interest in the first issue, see section 2.4. That the second issue was central for Gödel should be clear from the following text. I thus agree with Shapiro (1983), who emphasized that the problem of generalizing Gödel's incompleteness theorems was a "central item" in the development of a theory of computability. Gödel was also concerned about the third problem as evidenced by his discussion of Diophantine problems (1931, 1934). That naturally connects to Hilbert's Tenth Problem and other mathematical problems requiring decision procedures, like Thue's word problem for semigroups; see Gandy (1988, 60–61).

27. See footnote 7 of Post (1936), reprinted in Davis (1965, 291). Post's proposal is discussed in section 3.1, where I also describe the major difference with Turing's.

28. In section 9 of his 1888 work; see, in particular, theorem 126 and its applications in sections 11–13.

29. Tait (1981) argues for this claim. Gödel's system A (1933) seems to be just *PRA* and is claimed to contain all of finitist mathematics (actually used by "Hilbert and his disciples").

30. Gödel (1934), reprinted in Davis (1965, 61).

31. Gödel added for the publication of the lecture notes in Davis (1965): "This statement is now outdated; see the Postscriptum, pp. 369–371." He refers to the postscriptum appended to the lectures for Davis' volume.

32. As to the background for Herbrand's proposal, see section 2.2. Kalmar (1955) pointed out that the class of functions satisfying such functional equations is strictly greater than the class of (general) recursive functions.

33. In a letter to van Heijenoort of April 23, 1963, excerpted in the introductory note to Herbrand (1931d); see Herbrand (1971, 283). (Gödel refers to his 1934 lectures.) The background for, and content of, the Herbrand–Gödel correspondence is described in Dawson (1991).

34. The very notion of partial recursive function had been introduced in Kleene (1938).

35. The connection between the "different" proposals is also discussed by Gödel in correspondence with van Heijenoort, partially contained in van Heijenoort (1971) and in footnote 34 of Gödel (1934); that note was expanded in 1964. An earlier discussion of this issue is found in a letter to J. R. Büchi that was written by Gödel on November 26, 1957.

36. Herbrand considered the Ackermann function to be a finitist function, as he asserts—without giving a hint of an argument why (b) is satisfied—that it can be introduced according to the above schema; cf. n. 37.

37. This holds in spite of the explicit (but in this case definitely false) claim that the Ackermann function is among those that can be introduced intuitionistically. How is this confusing state of affairs to be understood? The discrepancy between Herbrand's conjecture in April and that in July is elucidated, it seems to me, by letters of Bernays

to Gödel from this period. Bernays and Herbrand had been in contact, in Berlin and also in Göttingen; Herbrand even sent Bernays a copy of his letter to Gödel—with an interesting accompanying letter dated, as the letter to Gödel was, April 7, 1931. In his letter of April 20, 1931, Bernays asked Gödel why the recursive definition of arithmetic truth could not be formalized in Z and why Ackermann's consistency proof could not be carried out in Z. Without waiting for Gödel's response, Bernays conjectured in his next letter to Gödel of May 3, 1931, that the answer to his questions lay in the unformalizability of certain types of recursive definitions in Z, the definition of truth and that of the Ackermann function being among them. (As to the Ackermann function, Bernays is still not right: it cannot be introduced in the fragment of arithmetic with induction for Σ_1^0 formulas, but in the fragment with Π_2^0 induction it can be introduced.)

38. In a letter to van Heijenoort of August 14, 1964; see van Heijenoort (1985a, 115–116).

39. In the postscriptum, Davis (1965, 73), Gödel asserts that the question raised in footnote 3 of the lectures can be "answered affirmatively" for recursiveness as given in section 9, "which is equivalent with general recursiveness as defined today." As to the contemporary definition, he seems to point to μ-recursiveness. But I do not understand how that definition could have convinced Gödel that "all possible recursions" are captured; nor do I understand how the Normal Form Theorem—as Davis (1981, 11) indicates—could do so without assuming some version of Church's Central Thesis. Indeed, such arguments seem to me crucially to require an appeal to that thesis and are, essentially, reformulations of Church's argument analyzed in the following text. That holds also for the appeal to the recursion theorem in (1954, 352), when Kleene argues that "Our methods ... are now developed to the point where they seem adequate for handling any effective definition of a function which might be proposed."

40. Church, in a letter to Kleene, dated November 29, 1935, and quoted in Davis (1982, 9).

41. As to the evolution of the concept of λ-definability and an earlier formulation of the thesis, see Appendix.

42. Church (1936a), reprinted in Davis (1965, 89–90). f is the characteristic function of the proposition; that 2 is chosen to indicate "truth" is, as Church remarked, accidental and nonessential.

43. Church (1936a, footnote 3), reprinted in Davis (1965, 90).

44. Ibid., 100.

45. An argument pertaining quite closely to the first method is given in Shoenfield (1967, 120). Church grappled with the connection of intuitive (effective) definability and representability in a system of symbolic logic already in Church (1934).

46. Church (1936a), reprinted in Davis (1965, 101). As to what is intended, namely for **L** to satisfy epistemologically motivated restrictions of the sort mentioned previously, see Church (1956, section 07, in particular pp. 52–53).

47. Compare footnote 20 in Davis (1965, 101), where Church remarks: "In any case where the relation of immediate consequence is recursive it is possible to find a set of rules of procedure, equivalent to the original ones, such that each rule is a (one-valued) recursive operation, and the complete set of rules is recursively enumerable."

48. The remark is obtained from footnote 19 of Church (1936a), reprinted in Davis (1965, 101), by replacing "an algorithm" by "a system of symbolic logic." Cf. Church's letter to Józef Pepis quoted in part 2 of Appendix.

49. It is most natural and general to take the underlying generating procedures directly as finitary inductive definitions. That is Post's approach via his production systems; using Church's Central Thesis to fix the restricted character of the generating

steps guarantees the recursive enumerability of the generated set. Cf. Kleene's discussion of Church's argument in (1954, 322–323). To see how pervasive this kind of argument is, compare note 39 and part 2 of Appendix.

50. As to the former, compare Gödel (1986, 170 and 176); as to the latter, see Davis (1965, 58).

51. There is no indication of an argument for the absoluteness of computability; I can think only of proofs of the Normal Form Theorem type. Also, Gödel did not compare the class of computable functions with other classes of functions, except for remarking that, "In particular, all recursively defined functions, for example, are already computable in classical arithmetic, that is, the system S_1." But here, I assume, he used "recursive" either in the sense of "primitive recursive" or "recursive of arbitrarily high order," but not "general recursive." The concept of recursion of arbitrarily high order is used in Gödel (1936b), a review of Church (1935a), in the context of λ-definability.

52. The content of Gödel's note was presented in a talk on June 19, 1935. See Davis (1982, 15, footnote 17) and Dawson (1986, 39).

53. "Remark printed on p. 83" of Davis (1965) refers to the remark concerning absoluteness that Gödel added in proof to the original German publication.

54. Church's remark about the "necessity of proving preliminary theorems" can be easily clarified: in my description of his argument for the recursiveness of the function F that is calculable in a logic I glossed over the very last step; to take it, Church refers to an earlier theorem (IV) in his paper, asserting that the class of recursive functions is closed under the μ-operator—in the "normal case."

55. As to the crucial points of difference, see Kleene's discussion in (1954, 361), where it is also stated that this treatment "is closer in some respects to Post 1936."

56. In Feferman (1991, 1–2) the case is made "for the primary significance for practice of the various notions of relative (rather than absolute) computability, ..." Indeed, Feferman argues later (p. 25) that "notions of *relative* computability have a much greater significance for practice than those of *absolute* computability." The reason given is that the organization and control of computational devices have to be structured into "*conceptual levels* and at each level into interconnected *components*." Although I can hardly disagree with that remark, it is heeded in the theory of absolute computability and, furthermore, if some process carried out by a device is to be called a "computation," it will certainly have to satisfy the general conditions formulated for absolute computability.

57. Post (1936), reprinted in Davis (1965, 289). Post remarks that the infinite sequence of boxes can be replaced by a potentially infinite one, expanding the finite sequence as necessary.

58. The emphasis is mine. To clarify some of the difficulties here, one has to consider other papers of Post's. A good starting point might be the discussion in Davis (1982, 21–22) and Post's remarks on finite methods in Davis (1965, 426–428).

59. Post (1936), reprinted in Davis (1965, 291).

60. This is from Kleene (1988, 34). In Gandy (1988, 98), one finds the pertinent and correct remark on Post's 1936 paper: "Post does not analyze nor justify his formulation, nor does he indicate any chain of ideas leading to it." In his review of that paper Church is also quite critical. But compare the second part of note 58.

61. My emphasis. This justification is discussed in section 3.3.

62. I am following the useful convention of Gandy's whereby a human carrying out a computation is a *computor*, whereas *computer* refers to some machine or other. In the *Oxford English Dictionary* the meaning of "mechanical" as applied to a person is given by: "resembling (inanimate) machines or their operations; acting or performed without the exercise of thought or volition; ..."

Turing (1936), reprinted in Davis (1965, 135).

4. A formulation of a computor operating in a two-dimensional computing space be given in Sieg (forthcoming); such a computor satisfies appropriately generalized initeness conditions, and it can be shown that his computations can be carried out by "linear" computor. So this step is indeed without theoretical consequence. This nalysis is of a significantly more general character than the investigations of generalized Turing machines in Kleene (1954).

65. In Davis (1965, 135) a very similar reason is given for restricting the number of states of mind. My account here is not a pure reconstruction, but joins Turing's considerations for restricting the number of symbols and states of mind with the (later) ones on immediate recognizability.

66. Turing relates state of mind to memory in section 1, p. 117, for his machines: "By altering its *m*-configuration [i.e., its state of mind] the machine can effectively remember some of the symbols which it has 'seen' (scanned) previously." This point is also emphasized by Kleene (1988, 22): "A person computing is not constrained to working from just what he sees on the square he is momentarily observing. He can remember information he previously read from other squares. This memory consists in a state of mind, his mind being in a different state at a given moment of time depending on what he remembers from before."

67. Turing actually argues that the changed squares must satisfy similar conditions as the observed squares and, for that reason, can be taken as being among them; in addition, the changes can be carried out one square at a time.

68. This last condition is specific for a "linear" computor; in general, one would have to require that there is a fixed finite number of configurations that can serve as "paths" from one observed symbolic configuration to the next. (Cf. n. 64.)

69. Turing (1936), reprinted in Davis (1965, 137).

70. Ibid., 138.

71. Church, in the letter to Kleene of November 29, 1935, quoted in Davis (1982, 9).

72. Postscriptum to Gödel (1934), reprinted in Davis (1965, 72). Cf. also Turing's remark in which formal systems are characterized as "mechanical" ones: Turing (1939), reprinted in Davis (1965, 194).

73. Turing (1936), reprinted in Davis (1965, 139).

74. It should be possible to present counterexamples that would show—as those in Gandy (1980)—that weakening of the conditions (1.1)–(2.2) leads to "omniscient computors."

75. Turing (1936), reprinted in Davis (1965, 135).

76. Post (1944), reprinted in Davis (1965, 310). See also footnote 1 of Post (1965).

77. Compare Webb (1990) for a detailed discussion that in my view is mistaken—mainly because Webb accepts the premise of Gödel's argument; that premise is clearly congenial to Webb's understanding of Turing's Thesis. (For the latter, see my note 5.)

78. It is informative to compare this statement of Gödel's with the (incorrect) translation on p. 241 *and*, most significantly, with the corresponding remark in Gödel (1972a, 273). In the latter, Gödel in effect added "from a reflection upon" to "insights that derive" in this remark.

79. There is a rich literature dealing with the "iterative conception of set," including papers by Parsons and Wang; this literature cannot be discussed here. For references to it, see the second edition of *Philosophy of Mathematics*, edited by P. Benacerraf and H. Putnam and published by Cambridge University Press in 1983, and also Parsons (1990).

80. The categoricity of the second-order theory of complete ordered fields does not argue against this point; as another example of a theory exhibiting similar features, consider the theory of dense linear orderings without end points.

81. Post (1965), reprinted in Davis (1965, 423).

82. The remarks in brackets are due to Goldfarb, the editor of Herbrand (1971).

83. This is a literal rendition of Herbrand's remark. Bernays, in his letter to Gödel of April 20, 1931, pointed out that Herbrand had misunderstood him in an earlier discussion: he, Bernays, had not talked about a result of his, but rather about Ackermann's consistency proof.

84. Clearly, Church assumed the converse of this claim.

85. Church alluded to the last page (1936a), that is, to Davis (1965, 107).

References

Ackermann, Wilhelm. 1928. Zum Hilbertschen Aufbau der reellen Zahlen. *Mathematische Annalen* 99, 118–133.

Aczel, Peter. 1977. An introduction to inductive definitions. In J. Barwise, ed., *Handbook of Mathematical Logic*. Amsterdam, 739–782.

Behmann, Heinrich. 1922. Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem. *Mathematische Annalen* 86, 163–229.

Bernays, Paul. 1922. Über Hilberts Gedanken zur Grundlegung der Arithmetik. *Jahresbericht der Deutschen Mathematiker Vereinigung* 31, 10–19.

———. 1923. Erwiderung auf die Note von Herrn Aloys Müller: "Über Zahlen als Zeichen." *Mathematische Annalen* 90, 159–163.

———. 1930. Die Philosophie der Mathematik und die Hilbertsche Beweistheorie. Blätter für Deutsche Philosophie 4, 326–367; reprinted in Bernays (1976), 17–61.

———. 1936. Review of Church (1936a and b). *Journal of Symbolic Logic* 1(2), 73–74.

———. 1976. *Abhandlungen zur Philosophie der Mathematik*. Darmstadt.

Bourbaki, Nicolas. 1950. The architecture of mathematics. *Mathematical Monthly* 57, 221–232.

Church, Alonzo. 1932. A set of postulates for the foundation of logic. *Annals of Mathematics* 33(2), 346–366.

———. 1933. A set of postulates for the foundation of logic, second paper. *Annals of Mathematics* 34(2), 839–864.

———. 1934. The Richard Paradox. *American Mathematical Monthly* 41, 356–361.

———. 1935a. A proof of freedom from contradiction. *Proceedings of the National Academy of Sciences, U.S.A.* 21, 275–281.

———. 1935b. An unsolvable problem of elementary number theory. Preliminary report (abstract). *Bulletin of the American Mathematical Society* 41, 332–333.

———. 1936a. An unsolvable problem of elementary number theory. *American Journal of Mathematics* 58, 345–363; reprinted in Davis (1965), 89–107.

———. 1936b. A note on the Entscheidungsproblem. *Journal of Symbolic Logic* 1(1), 40–41.

———. 1937a. Review of Turing (1936). *Journal of Symbolic Logic* 2(1), 42–43.

———. 1937b. Review of Post (1936). *Journal of Symbolic Logic* 2(1), 43.

———. 1956. *Introduction to Mathematical Logic*. Princeton, University Press.

Davis, Martin. 1958. *Computability and Unsolvability*. New York, McGraw-Hill.

———. 1965. *The Undecidable*. New York, Raven Press.

———. 1982. Why Gödel didn't have Church's Thesis. *Information and Control* 54, 3–24.

Dawson, John W. 1986. A Gödel chronology. In Gödel (1986), 37–43.
———. 1991. Prelude to recursion theory: the Gödel–Herbrand correspondence. Manuscript.
Dedekind, Richard. 1888. *Was sind und was sollen die Zahlen.* Braunschweig.
———. 1890. Letter to Keferstein. In van Heijenoort (1967), 98–103.
Feferman, Solomon. 1988. Turing in the land of $O(z)$. In Herken (1988), 113–147.
———. 1991. Turing's "Oracle": From absolute to relative computability—and back. Forthcoming in *Proceedings of the Symposium on Structures in Mathematical Theories* (held in San Sebastian, Spain, September 25–29, 1990).
Frege, Gottlob. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* Halle, Verlag Louis Nebert.
———. 1893. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet.* Jena, Verlag H. Pohle.
———. 1969. *Nachgelassene Schriften.* H. Hermes, F. Kambartel, and F. Kaulbach, eds. Hamburg, Felix Meiner Verlag.
———. 1984. *Collected Papers on Mathematics, Logic, and Philosophy.* B. McGuinness, ed. Oxford, Oxford University Press.
Gandy, Robin. 1980. Church's Thesis and principles for mechanisms. In Barwise, Keisler, and Kunen, eds. *The Kleene Symposium.* Amsterdam, 123–148.
———. 1988. The confluence of ideas in 1936. In Herken (1988), 55–111.
Gödel, Kurt. 1931. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. *Monatshefte für Mathematik und Physik* 38, 173–198.
———. 1933. The present situation in the foundations of mathematics. Lecture delivered at the meeting of the Mathematical Association of America in Cambridge, Massachusetts, December 29–30, 1933, to appear in Gödel's *Collected Works,* Vol. III.
———. 1934. On undecidable propositions of formal mathematical systems; lecture notes, Princeton; reprinted in Davis (1965), 39–71, and *Collected Works,* Vol. I, 346–371.
———. 1936a. Über die Länge von Beweisen. *Ergebnisse eines mathematischen Kolloquiums* 7, 23–24.
———. 1936b. Review of Church (1935a): A proof of freedom from contradiction. in Gödel (1986), 398–401.
———. 1946. Remarks before the Princeton bicentennial conference on problems in mathematics; reprinted in Davis (1965), 84–88, and *Collected Works,* Vol. II, 150–153.
———. 1951. Some basic theorems on the foundations of mathematics and their implications. Unpublished Gibbs Lecture, to appear in Gödel's *Collected Works,* Vol. III.
———. 1958. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica* 12, 280–287; reprinted in *Collected Works,* Vol. II, 240–250.
———. 1972a. On an extension of finitary mathematics which has not yet been used; reprinted in *Collected Works,* Vol. II, 271–280.
———. 1972b. Some remarks on the undecidability results. In *Collected Works,* Vol. II, 305–306.
———. 1986. *Collected Works,* Vol. I. Oxford, Oxford University Press.
———. 1990. *Collected Works,* Vol. II. Oxford, Oxford University Press.
Guillaume, Marcel. 1985. Axiomatik und Logik. In Jean Dieudonné, ed., *Geschichte der Mathematik, 1700–1900.* Braunschweig, 748–881.
Herbrand, Jacques. 1928. On proof theory. In Herbrand (1971), 29–34.

———. 1929a. The consistency of the axioms of arithmetic. In Herbrand (1971), 35–37.
———. 1929b. On the fundamental problem of mathematics. In Herbrand (1971), 41–43.
———. 1930a. Investigations in proof theory. In Herbrand (1971), 44–202.
———. 1930b. The principles of Hilbert's logic. In Herbrand (1971), 202–214.
———. 1931a. On the fundamental problem of mathematical logic. In Herbrand (1971), 215–259.
———. 1931b. Unsigned note on Herbrand's thesis, written by Herbrand himself. In Herbrand (1971), 272–276.
———. 1931c. Note for Jacques Hadamard. In Herbrand (1971), 277–281.
———. 1931d. On the consistency of arithmetic. In Herbrand (1971), 282–298.
———. 1971. *Logical Writings.* Warren Goldfarb, ed. Cambridge, Mass., Harvard University Press.
Herken, Rolf, ed. 1988. *The Universal Turing Machine—a Half-century Survey.* Oxford, Oxford University Press.
Hilbert, David. 1900a. Über den Zahlbegriff. *Jahresbericht der Deutschen Mathematiker Vereinigung* 8, 180–194.
———. 1900b. Mathematische Probleme. *Archiv der Mathematik und Physik* 1(1901), 44–63, 213–237.
———. 1904. Über die Grundlagen der Logik und der Arithmetik; reprinted in van Heijenoort (1967), 129–138.
———. 1925. Über das Unendliche; reprinted in van Heijenoort (1967), 367–392.
———. 1927. Die Grundlagen der Mathematik; reprinted in van Heijenoort (1967), 464–479.
Hilbert, David, and Wilhelm Ackermann. 1928. *Grundzüge der theoretischen Logik.* Berlin.
Hilbert, David, and Paul Bernays. 1935. *Grundlagen der Mathematik,* Vol. I. Berlin, Springer.
———. 1939. *Grundlagen der Mathematik,* Vol. II. Berlin, Springer.
Kalmar, Laszlo. 1955. Über ein Problem, betreffend die Definition des Begriffes der allgemein-rekursiven Funktion. *Zeitschr. f. math. Logik und Grundlagen d. Math.* 1, 93–95.
Kleene, Stephen C. 1936. General recursive functions of natural numbers. *Mathematische Annalen* 112(5), 727–742.
———. 1938. On notation for ordinal numbers. *Journal of Symbolic Logic* 3, 150–155.
———. 1954. *Introduction to Metamathematics.* Groningen, Wolters-Noordhoff.
———. 1976. The work of Kurt Gödel. *Journal of Symbolic Logic* 41(4), 761–778.
———. 1981. Origins of recursive function theory. *Annals of the History of Computing* 3, 52–66.
———. 1987a. Reflections on Church's Thesis. *Notre Dame Journal of Formal Logic* 28, 490–498.
———. 1987b. Gödel's impression on students of logic in the 1930s. In P. Weingartner and L. Schmetterer, eds., *Gödel Remembered.* Naples, Bibliopolis, 49–64.
———. 1988. Turing's analysis of computability, and major applications of it. In Herken (1988), 17–54.
Kneser, Adolf. 1925. Leopold Kronecker. *Jahresbericht der Deutschen Mathematiker Vereinigung* 33, 210–228.
Krämer, Sybille. 1988. *Symbolische Maschinen*—Die Idee der Formalisierung in geschichtlichem Abriß. Darmstadt, Wissenschaftliche Buchgesellschaft.
Kreisel, Georg. 1952. On the interpretation of non-finitist proofs II. Interpretation of number theory. Applications. *Journal of Symbolic Logic* 17(1), 43–58.

———. 1968. A survey of proof theory. *Journal of Symbolic Logic* 33(3), 321–388.

Kronecker, Leopold. 1887. Über den Zahlbegriff. *Crelle, Journal für die reine und angewandte Mathematik* 101, 337–355.

Löb, Martin H. 1959. Constructive truth. In A. Heyting, ed., *Constructivity in Mathematics*. Amsterdam, North-Holland, 159–168.

Lucas, J. R. 1961. Minds, machines and Gödel. *Philosophy* 36, 112–127.

Löwenheim, Leopold. 1915. Über Möglichkeiten im Relativkalkül; reprinted in van Heijenoort (1967), 228–251.

Mendelson, Elliott. 1990. Second thoughts about Church's Thesis and mathematical proofs. *The Journal of Philosophy* 87(5), 225–233.

Myhill, John. 1952. Some philosophical implications of mathematical logic. *The Review of Metaphysics* 6(2), 165–198.

Odifreddi, Piergiorgio. 1989. *Classical Recursion Theory: The Theory of Functions and Sets of Natural Numbers*. Amsterdam.

———. 1990. About logics and logicians—A palimpsest of essays by Georg Kreisel, Vol. II: Mathematics. Manuscript.

Parsons, Charles D. 1990. The structuralist view of mathematical objects. *Synthese* 84, 303–346.

———. In press. Quine and Gödel on analyticity. In Paolo Leonardi and Marco Santambrogio, eds., *On Quine*. Cambridge University Press.

Pasch, Moritz. 1918. Die Forderung der Entscheidbarkeit. *Jahresbericht der Deutschen Mathematiker Vereinigung* 27, 228–232.

Post, Emil. 1936. Finite combinatory processes. Formulation I. *Journal of Symbolic Logic* 1, 103–105; reprinted in Davis (1965), 289–291.

———. 1944. Recursively enumerable sets of positive integers and their decision problems. *Bulletin of the American Mathematical Society* 50, 284–316; reprinted in Davis (1965), 305–337.

———. 1965. Absolutely unsolvable problems and relatively undecidable propositions, account of an anticipation. In Davis (1965), 340–433. Originally submitted for publication in 1941.

Rosser, J. Barkley. 1984. Highlights of the history of the Lambda-Calculus. *Annals of the History of Computing* 6(4), 337–349.

Searle, John R. 1990. Is the brain a digital computer? *Proceedings and Addresses of the American Philosophical Association* 64(3), 21–37.

Shanker, Stuart G. 1987. Wittgenstein versus Turing on the nature of Church's Thesis. *Notre Dame Journal of Formal Logic* 28(4), 615–649.

Shapiro, Stewart. 1983. Remarks on the development of computability. *History and Philosophy of Logic* 4, 203–220.

———. In press. Metamathematics and computability. In Ivor Grattan-Guinness, ed., *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*. London, Routledge.

Shoenfield, Joseph R. 1967. *Mathematical Logic*. Reading, Mass., Addison-Wesley.

Sieg, Wilfried. 1984. Foundations for analysis and proof theory. *Synthese* 60, 159–200.

———. 1989. Turing's computors and Gandy's machines. Manuscript.

———. 1990a. Relative consistency and accessible domains. *Synthese* 84, 259–297.

———. 1990b. A Review of "S. G. Simpson, Friedman's research on subsystems of second order arithmetic." *Journal of Symbolic Logic* 55, 870–874.

———. 1991. Herbrand analyses. *Archive for Mathematical Logic* 30, 409–441.

Sieg, Wilfried, and Guglielmo Tamburrini. 1991. Does Turing's Thesis matter? (Or: on mechanical and other effective procedures). Manuscript.

Skolem, Thoralf. 1923. Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich; Skrifter utgit av Videnskapsselskapet i Kristiania, I. Matematisk-naturvidenskabelig klasse, no. 6, 1–38; reprinted in van Heijenoort (1967).

Spruit, Leen, and Guglielmo Tamburrini. 1991. Reasoning and computation in Leibniz. *History and Philosophy of Logic* 12, 1–14.

Tait, William W. 1981. Finitism. *Journal of Philosophy* 78, 524–546.

Tamburrini, Guglielmo. 1988. Reflections on mechanism. Ph.D. diss., Columbia University, New York.

Turing, Alan. 1936. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* 42, 230–265; reprinted in Davis (1965), 116–151.

———. 1939. Systems of logic based on ordinals. *Proceedings of the London Mathematical Society* 45, 161–228; reprinted in Davis (1965), 155–222.

———. 1948. Intelligent machinery. In *Machine Intelligence* 5, 3–23. Written in September 1947 and submitted to the National Physical Laboratory in 1948.

———. 1986. Lecture to the London Mathematical Society on 20 February 1947. In B. E. Carpenter and R. W. Doran, eds., *A. M. Turing's ACE Report of 1946 and Other Papers*. Cambridge University Press, 106–124.

van Heijenoort, Jean, ed., 1967. *From Frege to Gödel*. Cambridge, Mass.

———. 1971. Introductory note to Herbrand (1931d). In van Heijenoort (1967), 618–620, and with minor editorial changes in Herbrand (1971), 282–284.

———. 1982. L'oeuvre logique de Jacques Herbrand et son contexte historique. In J. Stern, ed., *Proceedings of the Herbrand Symposium* (Logic Colloquium '81). Amsterdam, 57–85.

———. 1985a. *Selected Essays*. Naples.

———. 1985b. Jacques Herbrand's work in logic and its historical context. In van Heigenoort (1985), 99–122. Translated and expanded version of van Heijenoort (1982).

von Neumann, Johan. 1927. Zur Hilbertschen Beweistheorie. *Mathematische Zeitschrift* 26, 1–46.

Wang, Hao. 1974. *From Mathematics to Philosophy*. London, Routledge & Kegan Paul.

Webb, Judson C. 1980. *Mechanism, Mentalism, and Metamathematics*. Dordrecht, Reidel.

———. 1990. Introduction to Remark 3 of Gödel (1972b) in *Collected Works*, Vol. II, 292–304.

Wiener, H. 1891. Über Grundlagen und Aufbau der Geometrie. *Jahresbericht der DMV* 1, 45–48.

Wittgenstein, Ludwig. 1980. *Remarks on the Philosophy of Religion*, Vol. 1. Oxford, Blackwell.