Absolutely undecidable problems and the power of mathematical reasoning*

Johannes Groessl

Harvard University

12/15/2009

Abstract

This paper discusses the implications of Gödel's Incompleteness Theorems on epistemology, especially on the question of computability of mathematical reasoning. Arguments are presented by John Lucas, David Chalmers, William Reinhardt and Timothy Carlson.

1 Gödels Dichotomy

Can we determine if a certain mathematical system - a bunch of axioms and inference rules - comprises *all of mathematics*? Mathematics is the study of structure, order and relation[1] and therefore primarily

*This essay has been written as a final paper for the seminar "Topics in the philosophy of mathematics", taught by Peter Koellner at Harvard University in fall term 2009.

[1] Art. "mathematics." Encyclopædia Britannica. Encyclopaedia Britannica 2008 Ultimate Reference Suite. Chicago: Encyclopædia Britannica, 2008.

an action performed by an intelligent lifeform. It includes all logical problems such a person can possibly formulate, since all mathematical propositions can be rewritten as logical propositions. But it is disputed if a mathematical problem has to be not only formulatable but also provable in order to be considered a part of mathematics.

Objective Mathematics is defined as "the body of those mathematical propositions which hold in an absolute sense, without any further hypothesis", whereas Subjective Mathematics is "the body of all humanly demonstrable (or knowable) mathematical truths, i.e., all the propositions which the human mind is capable of demonstrating."[2] Kurt Gödel argued in 1951 that it is contradictory to perceive a certain axiomatic system to be objective mathematics, because the consistency of this system would always remain as an expressable but provably unprovable sentence. If we knew that a certain system comprises all of mathematics, we would also know its consistency - which entails that our reasoning can impossibly be comprised by this particular system.[3]

Theorem 1 (Gödel)**.** *No well-defined system of correct axioms can contain all of mathematics.*

The fundamental question about the epistemological status of mathematics can be formulated as "Do subjective and objective mathematics coincide?" If we negate this question, we postulate mathematical propositions which are *absolutely undecidable*, and if we believe in the meaningfulness of every expressable problem, this decision leads us to the view that mathematics cannot be a human creation. If we,

[2]Feferman 2006, 2.
[3]C.f. Gödel 1951, 309.

however, believe that they coincide, we face an even more challenging consequence, namely that human beings are able to determine all or at least some undecidable problems and therefore, as Gödel expresses it, that "the human mind (...) infinitely surpasses the powers of any finite machine". The restriction to these two answers has come to be known as *Gödels dichotomy*.[4]

2 Independence and undecidability

According to our notion of whether a theorem can be acknowledged as intuitively acceptable, the studies of reverse mathematics determine the least amount of axioms that are needed in order to write down a straightforward mathemtical proof for this theorem. This is what leads us to accept set theory as the only plausible (sub-)model of reasoning, because every weaker system either lacks certain proofs[5] or will never be able to determine some meaningful (but not yet proved) statements.[6] Peter Koellner argues that "taking the power set operation seriously" automatically "throws you up" into a higher system, all the way to ZFC.

Gödel discovered in 1936 that every consistent system that includes the notion of natural numbers and induction (Peano Arithmetic) contains undecidable sentences and is therefore incomplete. This form of

[4]Feferman 2006, 1-3.

[5]E.g.: The Hydra Theorem is beyond Predicate Recursive Arithmetic (Finitism), Kruskal's theorem is beyond ATR_0 (Predicativism), the Graph Minor Theorem is beyond Π^1_0-CA.

[6]E.g. the claim that all projective sets are Lebesgue measurable is even beyond Zermelo-Fraenckel set theory with the axiom of choice.

undecidability is a *relative* undecidability, since it is possible to prove such a sentence in a stronger system. For example, the consistency of Peano Arithmetic can be proved in the system PRA + PR-TI(ϵ_0)[7].

If we have two systems with a different provability power we can compare them by determining the place of the universe of the systems in the *interprability hierarchy* of sets. For every system there is a subclass of the universe of sets that has a model which is mutually interpretable with the system and therefore has the same provability power. A set V_β, where β is any ordinal, is the set union of all sets beneath in the hierarchy:

$$V_\beta = \bigcup_{\alpha<\beta} V_\alpha$$

Now V_ω (where ω is the ordinal reflecting the amount of natural numbers) is a model of PA (where PA is mutually interpretable with ZFC without the axioms of infinity and replacment). Two steps up - after adding two axioms - we arrive at Zermelo-Fraenckel set theory, of which V_κ is a model, κ being a strongly inaccessible cardinal[8]. In order to further climb up the hierarchy, one can add more and more large cardinal axioms.

It is, however, not always the case that adding a new axiom makes the newly created system a stronger one. As Gödel has shown, the consistency sentence as well as its negation are independent from any sufficiently powerful system.[9] But while one "jumps up the hierarchy" by adding the consistency sentence, one does not by adding its

[7]Predicate Recursive Arithmetics + transfinite induction up to ϵ_0 for primitive recursive predicates. Vgl. Gentzen 1936.

[8]$\kappa = \sup\{\omega, 2^\omega, 2^{2^\omega}, ...\}$

[9]This is Gödels second Incompleteness Theorem. 'Sufficiently powerful' means that

negation: If a system is inconsistent, it proves anything, even its own consistency; adding the inconsistency sentence as a new axiom does not make any difference. We say that a consistency axiom creates a *single jump*. If a sentence as a new axiom as well as its negation as a new axiom both established a more powerful system, this would be called a *double jump*, although there are no natural examples for this case.[10] It is more interesting to look at these sentences that are independent but do not bring about any jump at all.

2.1 The Continuum Hypothesis

One example for this case is the Continuum Hypothesis (CH), introduced by Cantor and as old as set theory itself. This hypothesis basically says that $\aleph_1 = \beth_1$, where $\aleph_1$ being the cardinality of the set of all countable ordinal numbers and $\beth_1 = 2^{\aleph_0}$. The Generalized Continuum Hypothesis (GCH) similarly states that for all ordinals α, $\aleph_\alpha = \beth_\alpha$.

Gödel and Cohen showed that CH and its negation are both independent from ZFC. But adding CH or its negation as a new axiom does not create a more powerful system; the systems are mutually interpretable: [11]

Theorem 2. $ZFC \equiv ZFC + CH \equiv ZFC + \neg CH$

the system has to be able to express any formula in its language as a natural number; therefore every system that is an extension of Peano Arithmetic is considered sufficiently powerful.

[10]C.f. Koellner 2010, 38f (Ch. 2.3.2).

[11]C.f. Koellner, Peter: *Strong Logics of First and Second Order*, to appear in Bulletin of Symbolic Logic, 16.

Due to this type of independence, CH cannot be settled by any large cardinal axioms discovered so far. Yet there are certain proposed axioms “beyond” large cardinals that settle CH. For example, ZF + V=L (the assertion that every set is constructible) proves GCH; however, certain large cardinals refute this assertion, like the existence of measurable cardinals (Scott 1960). Furthermore, Woodin’s Strong Ω-Conjecture implies the existence of a ‘good’ axiom that implies $\neg$CH.

Although CH has for a long time been regarded as a good candidate to be an absolutely undecidable problem, we know today that there might be axioms that settle CH or $\neg$CH respectively. The problem here is to give criteria for a reasonable justification of new axioms.

2.2 Principle of reasoning

What justification do we have to accept axioms? Let us start with the intuitive concept of natural numbers covered by Robinson Arithmetic Q. To accept the self-evident theory Q but not accepting any extensions of it can be shown to be incoherent by the principle of reasoning:

Theorem 3 (Principle of Reasoning)**.** *If one accepts a certain theory to be a model of reasoning, one also has to accept the consistency of this thoery.*

Together with a serious notion of powersets (that is to assert that the powerset operation is meaningful and not inherently vague) this principle leads us all the way up to accepting set theory. What about going further? Gödel introduced the criterion of *fruitful consequences* for the (extrinsic) justification of new axioms.[12] Since large cardinal axioms can settle quite a few statements which would remain absolutely undecidable when regarding ZFC as the "final theory", we are justified to accept LCAs as new axioms. A similar point can be made for the axiom $V = L_S^{\Omega}$.[13]

2.3 The Penrose-Lucas-Argument

If I can know anything at all, human reasoning must be consistent. From this premise John Lucas developed in 1961 an argument against functionalism. This argument was further developed by Roger Penrose and David Chalmers.

Claim 4. *(If I can know anything at all,) I know that if there is an ultimate model for human reasoning, then this model has to be consistent.*

For any Turing machine that can simulate human reasoning by a certain kind of *knowledge* or *unassailable belief* (Chalmers) or *intuitive provability* (Reinhardt) function, this function cannot determine the consistency of itself. Invoking Gödel, the question "Is there any sentence which I can prove to be *not* an unassailable belief" cannnot be answered by this Turing machine. Thus, for any Turing machine T I can know that if T captures my reasoning powers, I would be able to know T's consistency and can therefore assert that T does not capture *all* of my reasoning powers.[14]

[12] C.f. Gödel 1947, 261.
[13] C.f. Koellner / Woodin 2010, Ch. 8.2.2.
[14] C.f. Lucas 1961 and Chalmers 1995.

3 The epistemology of provability

In his article "Reflexive Reflections" and in "Representation & Reality", Hilary Putnam advances the view that "all epistemic methods employed in human inquiry are, formalized, susceptible to Gödel's theorems."[15] Therefore, Gödel sentences cannot only be constructed in certain axiomatic systems but also in a system capturing human reasoning and epistemology - as far as such a system exists.

3.1 Epistemic Arithmetic

Epistemic Arithmetic (EA) is a modal theory of arithmetic developed by William Reinhardt and S. Shapiro.[16] EA is an extension of PA including a quantified version of the modal logic S4. Using Gödels Incompleteness Theorem, Reinhardt proved that the following sentence is inconsistent with EA: "I am a Turing machine and I know which one."[17] But he also formulated a weaker schema, called the Strong Mechanistic Thesis (SMT) which he conjectured to be consistent with EA:[18]

Claim 5 (SMT)**.** *For any formula $\phi(x)$, I know that the set of x for which I know $\phi(x)$ is recursively enumerable.*

If human reasoning can be modeled by an axiomatic system, then it can also be simulated by a Turing machine:

[15] Buechner 2006, 30.
[16] C.f. Carlson 1999, 2.
[17] Reinhardt 1986, 427-474.
[18] C.f. Carlson 1999, 4.

Corollary 6 (Post-Turing thesis). *'Humanly provable' is equivalent to provability by some Turing machine.*

Timothy Carlson gives the definition that an entity is a *knowing machine* if it is recursivley enumerable.[19] In his paper "Knowledge, Machines, and the Consistency of Reinhardt's Strong Mechanistic Thesis" he proves that EA as well as EA+SMT are such knowing machines.[20] But this proof cannot be regarded as a proof of functionalism, because it already presupposes functionalism by asserting that human knowledge is fully captured by the system EA.

3.2 The belief predicate

William Reinhardt introduced a belief predicate to use Gödels argument in epistemology, presuming that the Post-Turing thesis is correct. He classifies this predicate as a "one-place sentential connective (modal operator)"[21] It is not a standard predicate because its parameter can only be a boolean sentence. The operator shall give an "exact account of the formal axioms for 'provable' in the 'intuitive' or 'absolute' sense."

Assuming that human reasoning powers can be captured by a formal system, then this system would include such a belief operator. According to Reinhardt, the following axioms hold for this system:

Axiom 7 (1). $Bp \rightarrow p$

Axiom 8 (2). $B(p\rightarrow q) \rightarrow (Bp \rightarrow Bq)$

[19]Ibid., 11.
[20]Ibid., 23-28.
[21]Reinhardt, 318.

Axiom 9 (3)**.** $Bp \to BBp$

Axiom 10 (4)**.** $B(\forall n P(n)) \to \forall n B P(n)$

The first axiom is the *principle of epistemological infallibilism* and basically follows from the correctness of our system and mathematical realism.

Let us consider a system T that at least includes Robinson Arithmetic Q and is a model of human reasoning including a belief operator B as defined above. Now it is possible to construct a fixed point to obtain an undecidable B-sentence. The proof herefore is quite similar to Gödel's method. Let us take a look at the construction of self-referential sentence with the method of diagonalization:

Lemma 11 (Gödel/Smullyan)**.** *For every formula F with one free variable having the Gödel number g_F, there exists a sentence E such that $E(n) \Leftrightarrow \forall v_1 : (v_1 = n) \to F(v_1)$. The function $e(g_F, n)$ can be defined to produce the Gödel number of any $E(n)$. The diagonalization function will be defined as $d(x) \Leftrightarrow e(x, x)$ and outputs the Gödel number of a predicate that is equivalent to a function, which is given over its own Gödel number as an argument.*

Let $\tilde{P}$ stand for the set of all non-provable sentences, and $\tilde{P}^*$ for its diagonalization. The provability predicate is shown by Gödel to be arithmetic in systems including PA. Let $H(v_1)$ be the function expressing $\tilde{P}^*$, having the Gödel number h. The diagonalization of H corresponds to $E_h(h)$ and has the Gödel number $d(h)$. Hence, $E_h(h)$ is only true if and only if it is not provable, since if it is true, its Gödel number $d(h)$ is an element of $\tilde{P}$.[22] Therefore $E_h(h)$ is an undecidable sentence.

Reinhardt constructs a theorem that is similar to Gödel's, only working with an extended system that includes a belief operator.

3.3 Reinhardt's Theorem

Theorem 12 (Reinhardt). *"If the Post-Turing thesis obtains, then there must be absolutely undecidable sentences of arithmetic."*

If the Post-Turing thesis obtains, then every humanly decidable sentence is recursively enumerable. One can create a "belief system" T that is an extension of Robinson or Peano Arithmetic, containing a belief operator and the associated axioms shown above.

Definition 13 (Reinhardt). *Let F be a formula with one free variable in T, such that for each p either "$F[p] \to p$" or "$F[p] \to Bp$" is a belief sentence in T. F is therefore by definition a predicate that is a sufficient condition for truth or knowledge (belief) of a sentence, similar to the provability predicate.*

Lemma 14 (Reinhardt). *A belief system T with a Gödel sentence s proves both Bs and $B\neg F[s]$.*

Proof. Consider the tautology $(X \to Y) \to ((\neg X \leftrightarrow Y) \to Y)$. Every tautology is a belief sentence. Substitute F[s] for X and s for Y:

$$B(F[s] \to s) \to ((\neg F[s] \leftrightarrow s) \to s)$$

[22]C.f. Smullyan 2002, 36f and 49f.

According to the second belief axiom, the belief operator B can be distributed into material implications, which yields

$$B(F[s] \rightarrow s) \rightarrow (B(\neg F[s] \leftrightarrow s) \rightarrow Bs)$$

$B(F[s] \rightarrow s)$ holds by definition of F, and $B(s \leftrightarrow \neg F[s])$ holds because s is a Gödel sentence. Hence

$$T \vdash Bs$$

We can also apply the B-distibution to the definition of the fixed point construction, yielding $Bs \rightarrow B\neg F[s])$. From Bs follows

$$T \vdash B\neg F[s].$$

■

Theorem 15 (Reinhardt)**.** *If T is ω-complete, there is a sentence s such that $T \vdash s \wedge \neg Bs$, saying that s is absolutely undecidable.*

Proof. Let U(e,x) be any formula with two free variables e,x. Suppose in addition that for each Predicate P, $T \vdash \exists e : \forall x : BP(x) \leftrightarrow U(e, x)$. This basically says that every humanly decidable sentence is recursively enumerable. Then there is a formula q(e) with one free variable e such that $T\exists e : q(e) \wedge \neg Bq(e)$. If T is ω-complete, then $Ts \wedge \neg Bs$.[23]

■

3.4 David Chalmers

David Chalmers wrote a paper called "Minds, machines, and Gödel" commenting on John Lucas' and Roger Penroses' argument against

[23]C.f. Reinhardt 1985, 326.

functionalism. According to him, a system S, which "unassailably believes in its own consistency"[24], leads to a contradiction. At first, one assumes that $B(X)$ represents "a system's reasoning about its own belief". The symbol $A(X)$ should stand for the ability of a system to assert unassailably that X. Chalmer's argument can then be stated as following:

(1) If X is an unassailable sentence in S, then the belief in X is also unassailable. (2) Modus ponens is valid for belief sentencenes and this validity is an unassailable belief. (3) The system knows unassailably that if it beliefs a certain sentence, then it also has to belief that it beliefs this sentence. (4) The system knows unassailably that it is consistent, that is, that it knows that it does not belief in any contradictory sentence f. (5) A Gödel sentence can be construed saying "I don't belief in G". (6-12) From these premises a proof can be shown leading the system to assert an unassailable belief in its own inconsistency.

$$A('X') \Rightarrow A('B(X)') \tag{1}$$

$$A('B(X) \wedge B(X \rightarrow Y) \rightarrow B(Y)') \tag{2}$$

$$A('B(X) \rightarrow B(B(X))') \tag{3}$$

$$A('\neg B(f)') \tag{4}$$

$$A('G \leftrightarrow \neg B(G)') \tag{5}$$

[24]Chalmers 1995

$$A('B(G) \to B(\neg B(G))') \quad \text{[from (5)]} \tag{6}$$

$$A('B(G) \to B(B(G))') \quad \text{[from (3)]} \tag{7}$$

$$A('B(G) \to (B(\neg B(G) \wedge B(G)))') \tag{8}$$

$$A('B(G) \to B(f)') \tag{9}$$

$$A('f \to G') \quad \text{[ex falso quodlibet]} \tag{10}$$

$$A('B(f) \to B(G)') \quad \text{[from (10),(2)]} \tag{11}$$

$$A('G \leftrightarrow \neg B(f)') \quad \text{[from (9),(11)]} \tag{12}$$

$$A('B(G)') \quad \text{[from (12),(4),(1)]} \tag{13}$$

$$A('B(f)') \quad \text{[from (13),(9)]} \tag{14}$$

Since neither the premises (1-3) can be doubted nor the inclusion of Peano Arithmetic into our belief system which makes the construction of a Gödel sentence possible (5), the only premise that not only *can* but because of the contradiction this proof leads to *must* be rufuted is premise number 4. Chalmers conclusion therefore is: "Perhaps we are sound, but we cannot know unassalably that we are sound."[25]

Penrose's reply to Chalmers was basically the retention of premise (4), at the same time refuting the very idea of an axiomatic belief system. This concept would already presuppose functionalism and is inconsistent, as Chalmers argument has shown.[26]

3.5 A hierarchy of knowledge

The development from rationalism to criticism or even empiricism was a process of disputing the assumption that humans can obtain abso-

[25] Ibid.

[26] C.f. Penrose 1996, 3.6.

lute knowledge. Already David Hume argued that most knowledge is based on induction, and induction is a good, but not an absolute justification. Furthermore, Kuhn, Feyerabend and Quine criticized the method of falsification for scientific theories or paradigms. The *Duhem-Quine thesis* states that every hypothesis requires axiliary assumption an can therefore not be tested or falsified in isolation. Today, in science as well as in metaphysics, we build certain models of reality and compare these models according to criteria like consistency, coherency with sensual experience (including predictability power) and ontological parsimony. To hold not only a scientific pluralism (pluralism of theories) but also a pluralistic methodology makes objective comparison of theories impossible. Even Feyerabend was careful when he thought about the absolute incommensurability of theories,[27] seeing that this might lead to scientific 'anarchism'.

One important result of debating the pluralistic approaches of knowledge for philosophy of mathematics and epistemology is the following: If one can know anything at all, one knows its own consistency. To be more precise, if knowledge is comparable according to its grade of certainty, then the grade of the knowledge of one's consistency is above all other belief sentences. So every belief sentence carries a covert conditional "if I am consistent". For example, if I argue for a theory T and I say "I believe T", I actually mean "I believe$_1$ that my reasoning cannot prove false sentences AND I believe$_2$ that T is a reasonable theory, whereas believe$_1$ is of a higher grade of certainty than believe$_2$."

If human reasoning can be captured in a belief system, as Rein-

[27]Feyerabend, Against Method, 114 and 225.

hardt and Chalmers argue, the consistency of the belief system is absolutely undecidable. But since every possible belief depends on the consistency of this system, every such belief is of the same (or lower) epistemological grade of justification. Therefore the assertion of an axiomatic belief system leads to an epistemological egalitarianism in which every belief sentence has an equal grade of justification: absolutely undecidable. This is why Chalmer's argument concludes to the unassailable belief of a false sentence; no distinction between more or less justified can be made any more.

4 Conclusion

Gödels dichotomy cannot be resolved so far. The only tenable way to resolve it would be giving up the possiblity of *any* knowledge, as Nietzsche did by denying the intelligibility of the world. But this form of nihilism is incoherent to our everyday actions and the basic process of thinking, which presumes a distinction between more and less reasonable beliefs in order to work.

A reasonable consequence of the dichotomy is, in my opinion, a form of mathematical realism (or platonism), as Gödel pleaded for. The case is similar to Nietzsche's objection to the possibility of knowledge: Unless mathematics is not completely meaningless, it cannot be a human invention. If human mathematical reasoning can be captured by a formal system like Epistemic Arithmetic, the consistency of this system would be absolutely undecidable.

Maybe the functionalist thesis is right and human reasoning is equal to some formal epistemic system. Then we have to simply trust that our system is consistent and sound. But is trust not usually based on a belief and again belief a form of knowledge?

I come to the conclusion that if mathematical reasoning has the power to assert any truths at all, then Gödel is right by saying that the human mind infinitely surpasses the power of any finite machine. This does not mean that there are no absolutely indecidable problems. But it entails that there are some ‘finitely’ indecidable problems that are considered knowledge by our intellect - like the consistency of thinking.

I would like to conclude this essay with a 750-year-old quotation by Thomas Aquinas, maybe already anticipating the contemporary debate:

> Our intellect, furthermore, extends to the infinite in understanding. [28]

[28] Thomas Aquinas, Summa contra Gentiles I, cap. 43, n. 10.

Bibliography

[Buechner 2008] Buechner, Jeff: *Gödel, Putnam, and Functionalism. A New Reading of Representation and Reality*, Cambridge - London 2008.

[Carlson 1999] Carlson, Timothy: *Knowledge, Machines, and the Consistency of Reinhardt's Strong Mechanistic Thesis*, in: Annals of Pure and Applied Logic, Vol. 105, Issues 1-3, 1 November 2000, 51-82.

[Chalmers 1995] Chalmers, David: *Minds, Machines, and Mathematics*, in: Psyche 2 (1995), 11-25.

[Feferman 2006] Feferman, Solomon: *Are There Absolutely Unsolvable Problems? Gödels Dichotomy*, in: Philosophia Mathematica (III).

[Feyerabend 1975] Feyerabend, Paul: *Against Method: Outline of an Anarchistic Theory of Knowledge*, London 1975.

[Gödel 1947] G¨odel, Kurt: What is Cantors Continuum Problem?, in Gödel, Kurt: *Collected Works*, Vol. II. Publications 1938-1974, edited by S. Feferman et al., New York: Oxford University Press 1990.

[Gödel 1951] Gödel, Kurt: *Some basic theorems on the foundations of mathematics and their implications*, in Gödel, Kurt: *Collected Works*, Vol. III. Unpublished Essays and Lectures, edited by S. Feferman et al., New York: Oxford University Press 1995, 304323.

[Koellner 2010] Koellner, Peter / Woodin, W. Hugh: *Foundations of Set Theory*, to be published.

[Lucas 1961] Lucas, John: Minds, Machines and Gödel, in: Philosophy 36 (1961), 112- 127.

[Penrose 1996] Penrose, Roger: *Beyond the Doubting of a Shadow. A reply to Commentaries on Shadows of the Mind*, in: Psyche 2 (1996).

[Reinhardt 1985] Reinhardt, William: *Absolute Versions of the Incompleteness Theorem*, in: Nous 19 (Sep. 1985), 317-346.

[Reinhardt 1986] Reinhardt, William: *Epistemic Theories and the Interpretation of Gödel's Incompleteness Theorems*, in: Journal of Philosophical Logic 15 (1986), 427-474.

[Smullyan 2002] Smullyan, Raymond: *Gödels Incompleteness Theorems*, New York - Oxford ^{2}2002.