Undecidable problems in quantum field theory

Yuji Tachikawa

Received: 1 February 2023 / Accepted: 11 April 2023 / Published online: 1 September 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
We point out that some questions in quantum field theory are undecidable in a precise mathematical sense. More concretely, it will be demonstrated that there is no algorithm answering whether a given 2d supersymmetric Lagrangian theory breaks supersymmetry or not. It will also be shown that there is a specific 2d supersymmetric Lagrangian theory which breaks supersymmetry if and only if the standard Zermelo-Fraenkel set theory with the axiom of choice is consistent, which can never be proved or disproved as the consequence of Gödel’s second incompleteness theorem. The article includes a brief and informal introduction to the phenomenon of undecidability and its previous appearances in theoretical physics. A 15-minute video presentation of the content is also available at https://www.youtube.com/watch?v=H548i3dnsWE.

Keywords Undecidability in theoretical physics · Incompleteness theorems

1 Introduction

Many aspects of life are often unpredictable, and physics is no exception. By now, physicists are familiar with two types of unpredictability: one is the inherent probabilistic nature of quantum mechanics, and another is the chaotic behavior of various classical systems, where a minute difference in the initial condition is exponentially amplified, leading to the loss of predictability. In this article, we would like to introduce the audience to a third type of unpredictability, whose existence does not seem to be too widely appreciated.

To set up the stage, consider physicists working on a class of systems $X$. They are often interested in a certain property $P$ of those systems, and they often spend their entire academic career trying to answer the following question:

Given a system $x \in X$ in this class, does it have the property $P$?
Examples of such \((X, P)\) are:

|   | \(X\)                  | \(P(x)\) for \(x \in X\)                                      |
|---|------------------------|---------------------------------------------------------------|
| 1 | quantum spin chains    | \(x\) is gapless in the infinite volume limit                |
| 2 | supersymmetric theories | \(x\) spontaneously breaks supersymmetry                      |
| 3 | condensed matter systems| \(x\) is superconducting at room temperature and pressure    |
| 4 | gauge theories         | \(x\) confines                                               |

The unpredictability which we will discuss comes in two types, one generic and another specific. The generic one is of the following form:

There can be no single algorithm which decides for every \(x \in X\) whether \(x\) has the property \(P\),

whereas the specific one has the form:

There is a concrete \(x \in X\) for which it is not possible to prove or disprove whether \(x\) has the property \(P\).

As we will see below, undecidability in both senses would usually arise together, so we will simply call such a problem \((X, P)\) undecidable.

Such situations are well-known in the theory of computation and in mathematical logic. There is also a long history of works showing that there are similarly undecidable questions in theoretical physics.\(^1\) The earliest properly formulated statement of this kind in theoretical physics that the author of the present article is aware of is\(^2\) the work \([1]\), where the question

\[
X = \{\text{Classical mechanical systems}\},
\]

\[
P(x) = x \text{ leads to a specified outcome}
\]

\(^1\) That said, the author of this article was totally unaware of such series of works, and naively thought, before Beni Yoshida introduced him to \([7]\) in May 2021, that whether 1d spin chains are gapless and whether 2d supersymmetric Lagrangian theories break supersymmetry would be algorithmically decidable in the future when the theoretical physics would be sufficiently developed. The author wanted to share his surprise with his colleagues and gave an informal lunch talk on this topic in autumn 2021. The reaction of mathematicians there was, however, that the author of the present article was too naive and was behind mathematicians by about 100 years. Therefore it is perfectly plausible that many physicists reading this article might similarly laugh at the naïvété of the author. But the author still wanted to convey his surprise to a broader community, which was the main motivation for the author to prepare this article.

\(^2\) The author based this historical information on the introduction of an article \([20]\) suggested by one of the referees. (The main physics claim of \([20]\), presented in its Sec. 2, was too sketchy for the author of this present article to judge its validity. But the introduction does contain valuable encyclopedic information.) The same referee also suggested the author of the present article to cite \([21, 22]\). The former article \([21]\) is an essay on the relation of Gödel’s incompleteness theorem and physics, the validity of which is beyond the ability of the author of the present article to decide. The latter article \([22]\) is a philosophical analysis of the series of works starting in \([4]\). As the author of the present article is not well-versed in the philosophy of science, it is again beyond the capability of the author of the present article to discuss the validity of the content, but the author of the present article still wants to mention that two out of the three main points listed under bullet points in their summary section contain factual errors in the interpretation of \([4]\) and quantum spin systems in general. Namely, the first claims that the local form of the Hamiltonian of \([4]\) depends on the size of the lattice, which is not the case. Similarly, the second of the three claims that any finite quantum spin systems can be numerically computable on an actual computer, thus making the undecidability claim of \([4]\) meaningless. But an actual computation of a large quantum spin system is exponentially hard due to the exponential growth of the number of states. As such, any such philosophical analysis needs to tackle directly the question of the meaning of philosophical computability of a computation practically impossible, which was not adequately addressed in that article.
was shown to be undecidable. Shortly afterwards, works [2, 3] also discussed undecidable
questions in classical mechanics. More recently, it was shown in [4, 5] from 2015, that the question
\[ X = \{2d \text{ quantum spin systems}\}, \]
\[ P(x) = x \text{ is gapless in the infinite volume limit} \]  
was similarly demonstrated to be undecidable. This result was later strengthened to the
undecidability of gaplessness of 1d quantum spin systems in [6, 7] and extended to the
undecidability of the phase space of 2d quantum spin systems and of the outcome of the real-space
renormalization group flow in [8, 9], respectively. Similarly, in [10], it was shown that
whether a quantum optimal control protocol can be implemented or not is undecidable, and
in the work [11, 12], it was demonstrated that whether 1d quantum spin systems thermalize
is undecidable. In the present article, we will also see that the question
\[ X = \{2d, N = (2, 2) \text{ supersymmetric Lagrangian theories}\}, \]
\[ P(x) = x \text{ breaks supersymmetry} \]  
is equally undecidable. In each case, the undecidability is shown by reducing it to the most
basic and fundamental case of undecidability, the halting problem of computer programs:
\[ \Xi = \{\text{computer programs}\}, \]
\[ \Pi(\xi) = \text{the execution of } \xi \text{ halts in finite time}. \]

As will be recalled later, there can be no algorithm to decide \( \Pi(\xi) \) for each \( \xi \in \Xi \). Furthermore, one can write a program \( \xi_0 \) which halts in finite time if and only if the Zermelo-Fraenkel
set theory with the axiom of choice, the standard foundation of mathematics of our times,
is inconsistent. Due to Gödel’s second incompleteness theorem, that \( \xi_0 \) halts in finite time
can neither be proved nor disproved. Therefore, the remaining task to establish the unde-
cidability of a particular question \((X, P)\) is to construct a concrete map \( f : \Xi \to X \)
such that \( P(f(\xi)) = \Pi(\xi) \), i.e. to find a way to construct systems \( f(\xi) \in X \) encoding computer
programs \( \xi \) such that \( f(\xi) \) has the property \( P \) if and only if the program \( \xi \) halts in finite time.

The rest of the article is organized as follows. In Sec. 2, we collect basic facts concerning
undecidability in computer science and mathematical logic. In more detail, we very briefly
review the halting problem of Turing machines in Sec. 2.1, summarize Gödel’s incomple-
teness theorems in Sec. 2.2, discuss what their combination means in Sec. 2.3. We then recall
in Sec. 2.4 their relation to the negative solution to Hilbert’s 10th problem, i.e. the fact that
the solvability of Diophantine equations is undecidable.

We then come to Sec. 3, which contains three physics manifestations of undecidability. In
Sec. 3.1, we recall the arguments of [1–3] who found the undecidability in classical mechan-
ics. In Sec. 3.2, we briefly review the results of [4, 5], which showed the undecidability of the
gaplessness of 2d spin systems. Then in Sec. 3.3, we show that whether 2d supersymmetric
Lagrangian theory breaks supersymmetry is undecidable, which is a simple consequence of
the negative solution to Hilbert’s 10th problem. We will conclude with discussions in Sec. 4.

2 Preliminaries

In this section, we give a brief and informal overview of fundamental results we need from
theoretical computer science and mathematical logic. Interested readers should consult stan-
dard textbooks for more details. For those who read Japanese, an extremely readable account
for non experts is [13].
2.1 Halting problem

Computers are ubiquitous in modern life, but the theory of computing predates modern digital computers and goes back to the early 20th century. At that time, multiple proposals were made to formalize precisely what it means to compute algorithmically. Various proposals were made, including Turing machines, Church’s $\lambda$ calculus, and recursive functions. These were soon understood to be all equivalent since an algorithm expressed in one formulation can be translated to an algorithm expressed in another. We can then call something computable when it is computable in any one of these equivalent formulations; this is the Church-Turing thesis.

In the discussion below, we typically consider Turing machines, which are in some sense the closest among the formulations to the actual computers we use. A Turing machine $\xi$ has a finite-length program and a finite number of internal states, and works with an infinite amount of memory. It operates step by step, according to its program, by updating its internal state and modifying a finite amount of memory at each step.

We note that a Turing machine accepting a finite sequence of natural numbers $n_1, n_2, \ldots$ can also be thought of as a Turing machine accepting a single natural number. This is because e.g. we can encode $n_1, n_2, \ldots$ into a single natural number $N := 2^{n_1}3^{n_2} \cdots (p_k)^{n_k} \cdots$ where $p_k$ is a $k$-th prime number, and that we can recover $n_1, n_2, \ldots$ from $N$ algorithmically. This allows us to be flexible about the number of arguments a Turing machine takes.

Consider a Turing machine $\xi$ accepting a natural number $n$ as the input. It then tries to compute the result, emit it as an output $\xi(n)$, and halt. It can also go into an infinite loop and does not halt, as often happens when one first learns to program. In this case the output is undefined. Below, we simply say $\xi(n)$ halts or does not halt, to distinguish these two behaviors.

We need two fundamental facts of Turing machines. One is the existence of a universal Turing machine $\upsilon$, which is a single program which emulates arbitrary Turing machines in the following sense. Note that a program describing a Turing machine $\xi$ is a finite string of alphabets. It can be saved in a text file, which is simply a string of bits (zeros and ones). It can then be regarded as a natural number $\lceil \xi \rceil$ written in binary digits. A universal Turing machine $\upsilon$ accepts two natural numbers, $i$ and $n$. If $i$ is of the form $\lceil \xi \rceil$ for a Turing machine $\xi$, i.e. if $i$ describes a valid program, $\upsilon$ emulates the behavior of $\xi$ operating on $n$, and emits $\xi(n)$ as the output if $\xi$ halts, and $\upsilon$ runs forever if $\xi$ runs forever. Schematically, we have $\upsilon(\lceil \xi \rceil, n) = \xi(n)$. The existence of $\upsilon$ is nontrivial, but we do use this fact every day: the CPU can run arbitrary programs stored in the memory, since it is a universal Turing machine.

Another fundamental fact of Turing machines is that it is impossible to write a Turing machine $\eta$ such that

$$
\eta(\lceil \xi \rceil, n) = \begin{cases} 
0 & \text{if } \xi(n) \text{ halts}, \\
1 & \text{if } \xi(n) \text{ does not halt}. 
\end{cases}
$$

(2.1)

In other words, the halting problem of Turing machines is algorithmically undecidable. It can be proved as follows. Suppose there is such an $\eta$. We can then write a Turing machine $\delta$, which accepts a natural number $i$ as the input and does the following:

- Compute $\eta(i, i)$.
- If the result is $0$, go to an infinite loop.
- If the result is $1$, halt.
Now consider the behavior of $\delta(\lceil \delta \rceil)$. Expanding the definition, we see that it does the following:

- Compute $\eta(\lceil \delta \rceil, \lceil \delta \rceil)$. The result is 0 if $\delta(\lceil \delta \rceil)$ halts, and 1 if $\delta(\lceil \delta \rceil)$ does not halt.
- If the result is 0, go to an infinite loop.
- If the result is 1, halt.

Therefore, $\delta(\lceil \delta \rceil)$ halts if $\delta(\lceil \delta \rceil)$ does not halt, and $\delta(\lceil \delta \rceil)$ does not halt if $\delta(\lceil \delta \rceil)$ halts, which is a contradiction.

### 2.2 Incompleteness theorems

We now proceed to an informal discussion of the incompleteness theorems of Gödel. First we need to introduce a few technical terminologies. A formal system, or equivalently a theory, is a set of axioms and inference rules. Let us denote such a theory by $T$. A model $M$ of $T$ is a mathematical object satisfying these axioms and rules. As an example, we can consider a theory $G$ of groups, consisting of the usual axioms of groups. Tautologically, a model of this theory $G$ is a group. As another example, we can take the Zermelo-Fraenkel set theory with the axiom of choice, which is usually abbreviated as ZFC. This is often taken as the foundation of modern mathematics. In the case of ZFC, we do not usually think of using a particular model of it, but we are effectively working in one.

Consider a mathematical statement $\phi$ in a theory $T$. A proof $\pi$ of $\phi$ in the theory $T$ is a series of valid inferences starting from the axioms of $T$. If there is a proof $\pi$ of $\phi$ in the theory $T$, the statement $\phi$ is true in any model $M$. This is known as the soundness of logic. A more nontrivial fact is that if a statement $\phi$ is true in every model $M$, $\phi$ has a proof in the theory $T$. This is known as the completeness of logic. A theory $T$ is called consistent if there is no proof of a contradiction in the theory. This is known to be equivalent for the theory $T$ to have a model $M$. We assume below that all theories we discuss are consistent, unless otherwise mentioned.

For example, we can consider the theory $G$ of groups, and let $\phi_{Ab}$ be the statement that all elements commute with each other. The statement $\phi_{Ab}$ is true or false in a model $G$ of $G$, depending on whether $G$ is Abelian or not. Then, the statement $\phi_{Ab}$ can neither be proved nor disproved in the theory $G$ of groups, and whether $\phi_{Ab}$ is true or not depends on the particular model $G$ of the theory $G$. The combinations $G + \phi_{Ab}$ and $G + \neg \phi_{Ab}$ are the theory of Abelian groups and the theory of non-Abelian groups, respectively.

In general, it should not be surprising that a certain statement $\phi$ in a theory $T$ can neither be proved nor disproved, and whether $\phi$ is true or not depends on the chosen model $M$ of $T$. Such statement $\phi$ is said to be independent of the theory $T$.

In the early 20th century, when people first started to formalize mathematics, it was realized that the set theory can be used as a foundation of any other known mathematics, by encoding other mathematical objects as sets; it is similar to the use of the assembly language to implement higher-level programming languages. It was then hoped that the set theory would either prove or disprove any statement $\phi$, and therefore assign whether each statement $\phi$ is true or false uniquely.

Gödel’s first incompleteness theorem says that this is impossible. Namely, it says that for any consistent theory $T$ capable of describing elementary arithmetic of natural numbers, there is a statement $\zeta_T$ of this theory which is independent of $T$. As we said, this means that $\zeta_T$ can neither be proved nor disproved in this theory, and therefore the validity of the statement $\zeta_T$ depends on the model of $T$. The proof of this theorem uses the idea that the statements $\phi$ and the proofs $\pi$ in the theory $T$ can be saved in text files, whose content can
be considered as (gigantic) natural numbers $[\phi]$ and $[\pi]$. These are called Gödel codes of $\phi$ and $\pi$, respectively. Then, whether a statement $\phi$ can be proved or not can itself be rephrased as an arithmetical statement about its Gödel code $[\phi]$ which can be treated within the theory $T$.

To state Gödel’s second incompleteness theorem, we consider the statement that there is no proof in the theory $T$ which leads to a contradiction. This statement can also be translated into an arithmetical statement $\text{Con}(T)$, which says there is no natural number representing the content of a text file containing a proof of a contradiction. Then Gödel’s second incompleteness theorem says that $\text{Con}(T)$ can neither be proved nor be disproved in $T$. Our discussions have been extremely rough, but even more roughly speaking, this means the following: if a theory $T$ is consistent and is capable of describing elementary arithmetic of natural numbers, it can never prove or disprove its own consistency.

These incompleteness theorems apply to ZFC, the Zermelo-Fraenkel set theory with the axiom of choice, which is usually taken as the standard foundation of modern mathematics. This means that ZFC can never prove or disprove $\text{Con}(\text{ZFC})$. ³

Adding $\text{Con}(\text{ZFC})$ as an axiom does not improve the situation, since the consistency of the combination $\text{ZFC} + \text{Con}(\text{ZFC})$ cannot be proved from it either. A more peculiar point is that the combination $\text{ZFC} + \neg\text{Con}(\text{ZFC})$ is also consistent if ZFC is consistent. This somewhat contradictory situation is possible because, in a model $M$ of $\text{ZFC} + \neg\text{Con}(\text{ZFC})$, the statement ‘there is a natural number which encodes the proof of a contradiction’ is true in $M$, but what $M$ thinks as the set of natural number can be larger than what $\mathbb{N}$ is for us.

### 2.3 Incompleteness theorems and Turing machines

Let us now combine what we learned about Turing machines and incompleteness theorems. Suppose one wants to know if a statement $\phi$ is provable in a theory $T$. Let us write $\phi$ in an editor, save it into a text file, and regard its content as a binary representation of a natural number $[\phi]$. We can then write a program $\gamma_T([\phi], n)$ which does the following:

- Regard the natural number $n$ as the binary representation $[\pi]$ of the content of a text file containing a purported proof of $\phi$.
- Check whether $\pi$ is a valid proof of $\phi$ in the theory $T$.

We note that such programs $\gamma_T$ are known as proof verifiers and their development is an active area of research.⁴

Using $\gamma_T$, it is possible to write a program $\delta_T([\phi])$ which does the following:

- Enumerate natural numbers $n = 1, 2, 3, \ldots$ one by one.
- Checks whether $n$ is a proof of $\phi$ using $\gamma_T([\phi], n)$.

This program finds a proof of $\phi$ and halts in finite time if $\phi$ is provable in $T$, but it does not halt if $\phi$ is not provable in $T$.

---

³ Analyzing the structure of ZFC mathematically using ZFC might sound circular, but it is simply a special case of studies of various formal systems $T$ using ZFC assuming the consistency of ZFC, where $T$ happens to be ZFC. As such, we can learn a lot about $T = \text{ZFC}$, for example the fact that it cannot prove its own consistency. The author thinks that it is vaguely analogous to the situation of the validity of the Born rule and the theory of measurement in quantum mechanics. There, one analyzes the combined system of the experimental apparatus and the target system by applying the Born rule to the entire system. In this manner one can learn a lot about the measurement process, although one cannot prove the validity of the Born rule as applied to the entire system.

⁴ For example, Gödel’s second incompleteness theorem itself was checked several years ago in [23, 24] using the proof verifier Isabelle/HOL. For informed opinions of pure mathematicians who advocate the use of proof verifiers from their own experiences of wrong proofs, see e.g. [25, 26].
We now consider the program \( \xi_T := \delta_T([0 = 1]) \). This program halts if and only if there is a derivation of this contradiction from \( T \). In other words, this program halts if \( \text{Con}(T) \) is false, and does not halt if \( \text{Con}(T) \) is true. Gödel’s second incompleteness theorem says that which of the two alternatives is chosen is independent of \( T \). In particular, the program \( \xi_{\text{ZFC}} \) halts if and only if \( \text{ZFC} \) is inconsistent, and whether it does so can neither be proved nor disproved from \( \text{ZFC} \). We note that a variant of \( \xi_{\text{ZFC}} \), an explicit Turing machine whose halting behavior is independent of \( \text{ZFC} \), was explicitly written down in [14], which only has 7910 internal states.

### 2.4 Diophantine equations and Hilbert’s 10th problem

So far, we recalled that there is no algorithm which decides for each Turing machine \( \xi \) whether \( \xi \) halts in finite steps, and that there is a specific program \( \xi_{\text{ZFC}} \) such that whether it halts or not is undecidable in \( \text{ZFC} \). Here we explain, again very briefly, how these facts were used to answer Hilbert’s 10th problem negatively. For more details, the readers are referred e.g. to Davis’ review [15] or Matiyasevich’s textbook [16].

Hilbert’s 10th problem is the question of finding an effective method to decide if a given Diophantine equation has a solution or not. Here, a Diophantine equation is a polynomial of many variables with integer coefficients over natural numbers,

\[
P(x_1, \ldots, x_k) = 0, \quad x_i \in \mathbb{N}.
\]  

A long series of works by Davis, Putnam, Robinson and finally by Matiyasevich established that for any Turing machine \( \xi \) one can construct a Diophantine equation \( P_\xi(x_1, \ldots, x_k) = 0 \) such that it has a solution if and only if \( \xi \) halts. This immediately means that there is no algorithm to tell whether a Diophantine equation has a solution or not, and that there is a specific Diophantine equation \( P_{\xi_{\text{ZFC}}} \) such that whether it has a solution is independent of \( \text{ZFC} \).

Explicitly writing down this equation is a difficult problem. One method is to use the Diophantine form \( P_\nu \) of the universal Turing machine \( \nu \), which allows us to translate whether a program \( \xi \) halts on the input \( x \) into whether the Diophantine equation

\[
P_\nu([\xi], x; x_1, \ldots, x_k) = 0
\]  

has a solution \( x_1, \ldots, x_k \in \mathbb{N} \) or not. An example of such a polynomial \( P_\nu \) was constructed in [17], and has the form

\[
0 = P_\nu(v, x; a, b, c, d, e, f, g, h, u, j, k, l, m, n, o, p, q, r, s, t, w, z, u, y, \alpha, \gamma, \eta, \theta, \lambda, \tau, \phi)
\]

\[
= (((zu) + u)^2 + y - v)^2 + (elg^2 + \alpha - (b - x)(q - b)^2)^2 + (q - b^{560})^2
\]

\[
+ (\lambda + q^4 - 1 - \lambda b^5)^2 + (\theta + 2z - b^2)^2 + (l - u - t\theta)^2 + (e - y - m\theta)^2 + (n - q + 16)^2
\]

\[
+ (r - (g + eq^3 + \eta q)^2 + (2e - z\lambda)(1 + xb^5 + g)^4 + \lambda b^5 + \lambda b^5 q^4)q^5(n^2 - n)
\]

\[- (q^3 - bl + l + \theta \lambda q^3 + (b^5 - 2)q^5)(n^2 - 1)^2
\]

\[
+ (p - 2ws^2 rn^2)^2 + (p^2 k^2 - k^2 + 1 - r)^2 + (4(c - kn)^2)^2 + \eta - k^2)^2
\]

\[
+ (k - (r + 1 + hp - h))^2 + (a - (wn^2 + 1)kn)^2 + (c - 2r - 1)^2
\]

\[
+ (d - (bw + ca - 2c + 4aw - 5\gamma))2 + (d^2 - (a^2 - 1)c^2 - 1)^2
\]

\[
+ (f^2 - (a^2 - 1))^2 c^4 - 1)^2
\]

\[
+ ((d + of)^2 - ((a + f^2(d^2 - a))^2 - 1)(2r + 1 + jc)^2 - 1)^2.
\]  

\[\textsf{Springer}\]
Then, for a specific choice of $v$ and $x$, the existence of a solution to this Diophantine equation is equivalent to the inconsistency of ZFC, and therefore can neither be proved nor disproved from ZFC. Finding explicit natural numbers $v$, $x$ for $\xi_{ZFC}$ is still a tedious question, however. Another method is to use the Diophantine equation representing whether a statement $\phi$ is provable in a certain theory $T$ including elementary arithmetic, which was written down in [18].

Before proceeding, we point out two minor facts. First, the question of the existence of a simultaneous solution to a set of Diophantine equations $P_1 = 0$, $P_2 = 0$, ..., $P_\ell = 0$ can be reduced to the question of the existence of a solution to a single Diophantine equation

$$0 = P := (P_1)^2 + (P_2)^2 + \cdots + (P_\ell)^2.$$  

(2.5)

In fact, the formula (2.4) is obtained by using this trick from a set of equations given in [17]. Second, in the theory of Diophantine equations, it is common to let unknowns $x_1, \ldots, x_k$ to vary over natural numbers as we did above, but the existence of a solution to Diophantine equations over integers is equally undecidable. This can be shown using Lagrange’s four square theorem, saying that any natural number $x$ can be written as $x = y_1^2 + y_2^2 + y_3^2 + y_4^2$, where $y_i$ are integers. Indeed, then, the question of the existence of a solution to $P(x_1, \ldots, x_k) = 0$ with $x_i \in \mathbb{N}$ can be rewritten as the question of the existence of a solution to

$$P(y_{11}^2 + y_{12}^2 + y_{13}^2 + y_{14}^2, y_{21}^2 + y_{22}^2 + y_{23}^2 + y_{24}^2, \ldots, y_{k1}^2 + y_{k2}^2 + y_{k3}^2 + y_{k4}^2) = 0$$  

(2.6)

with $y_{ia} \in \mathbb{Z}$, showing that the question of the existence of a solution over integers is undecidable. We will use the undecidability in this form later in this article in Sec. 3.3.

3 Physics manifestations

3.1 Undecidability in classical mechanical systems

The earliest realizations that the undecidability in mathematics also manifests itself in theoretical physics were given, to the author’s knowledge, in the context of classical mechanics in the late 1980s to the early 1990s in [1–3]. Here we would like to review their arguments.

The first reference [1] only gave a brief argument on this point, that one can build a Turing machine within classical mechanics. This then immediately translates to the fact that it is undecidable if a given outcome is reached in a given classical mechanical system.

The second reference [2] gave the argument somewhat more precisely. In it, it was pointed out that one can devise a non-linear continuous mapping $f$ of a square onto itself depending on a given Turing machine, such that $f(f(f(\cdots((x_0)\cdots))))$ reaches a certain subregion of the square if and only if the Turing machine halts. Then the author of [2] outlined, although not very precisely, that one can design a potential such that a classical point particle moving under its influence will follow the mapping $f$, thus arguing that the outcome of a given classical mechanical system is undecidable.

The third reference [3] is of a somewhat different nature. It uses a mathematical result of [19] saying that it is undecidable to conclude if a real function given by an explicit mathematical expression composed from elementary functions is actually a zero function or not. Then, the author of [3] considers classical systems whose Hamiltonian is given by explicit mathematical expressions. Clearly, such systems inherit the undecidability associated to the fact that it is undecidable to know if a certain term in its Hamiltonian is zero or not.
3.2 Undecidability of ground state properties of 2d quantum spin systems

We next provide a brief review of the wonderful work [4, 5] in which the authors showed that whether the gaplessness in the infinite volume limit of 2d spin systems with the finite-range interaction is undecidable. This is done by finding a way to construct 2d spin systems from a given Turing machine $\xi$ so that the property of its ground state depends on whether $\xi$ halts or not.

The main technical result of [4, 5] is the following: there exists a nearest-neighbor translationally-invariant Hamiltonian $H_\xi$ on a 2d spin system with the Hilbert space $H = \bigotimes_i V_i$, where the local Hilbert space is $V_i := \mathbb{C}^{d_\xi}$, such that the ground state energy $E_0$ on a lattice of size $L^2$ behaves as

$$E_0 \leq -1/2 \quad \text{for all } L \text{ if } \xi \text{ does not halt},$$

$$E_0 \geq +1/2 \quad \text{for all } L \geq L_\xi \text{ if } \xi \text{ halts},$$

where $L_\xi$ is a constant related to the running time of the program $\xi$.

The papers [4, 5] actually describe two different versions of this result. One is their Lemma 8, where the dimension $d_\xi$ of the local Hilbert space is at most proportional to the number of states of the Turing machine $\xi$, but with the bonus that the ground state in the non-halting case is a gapped, product state. Another is their Corollary 54, where $d_\xi$ is now independent of $\xi$, but without the guarantee of the nature of the ground state in the halting case. We use the first version in the rest of the section.

The construction of this Hamiltonian $H_\xi$ requires real ingenuity, which the interested readers can marvel at by going over their detailed description in [5], and which the author of this present article dare not to provide. The remaining task is to translate this undecidability to arbitrary properties of the ground state. For this purpose, we take two auxiliary 2d nearest-neighbor spin systems whose behavior we do understand, $H_W$ acting on $\bigotimes_i W_i$ with $W_i = \mathbb{C}^a$, and $H_Z$ acting on $\bigotimes_i Z_i$ with $Z_i = \mathbb{C}^b$, both normalized so that the ground state has zero energy. Let us write the nearest-neighbor interactions of $H_\xi$, $H_W$ and $H_Z$ as $h_\xi^{(i,j)}, h_W^{(i,j)}$ and $h_Z^{(i,j)}$. We now fuse three systems as follows. We take the local Hilbert space per site of the combined system to be $U_i = V_i \otimes W_i \oplus Z_i$, and we take the new nearest-neighbor Hamiltonian to be given by

$$H_{\text{combined}} = \sum_{i,j} (h_\xi^{(i,j)} + h_W^{(i,j)} + h_Z^{(i,j)}) + \lambda \sum_i P_{V_i \otimes W_i} P_{Z_i}.$$  \hfill (3.2)

Here, the term $h_\xi^{(i,j)} + h_W^{(i,j)}$ now acts on $V_i \otimes W_i \otimes V_j \otimes W_j \subset U_i \otimes U_j$ extended by zero outside by abuse of notations, the term $h_Z^{(i,j)}$ now acts on $Z_i \otimes Z_j \subset U_i \otimes U_j$ similarly extended by zero, $\lambda$ is a large positive constant, and $P_{V_i \otimes W_i}$ and $P_{Z_i}$ are projectors to the respective components.

Let us study the ground state of the combined system. The term proportional to $\lambda$ gives a huge penalty to wavefunctions supported both on the $V \otimes W$ part and on the $Z$ part. Therefore, the ground state comes either from wavefunctions on the $V \otimes W$ part only, or those on the $Z$ part only. The ground state on the $Z$ part is that of $H_Z$ and the energy is $E_0^{(Z)} = 0$. The ground state on the $V \otimes W$ depends on whether $\xi$ halts or not. If it halts, the energy $E_0^{(V \otimes W)} \geq +1/2$. If it does not halt, $E_0^{(V \otimes W)} \leq -1/2$, and the ground state has all the features of that of $H_W$, since the ground state of $H_\xi$ in this case is guaranteed to be a gapped product state. Therefore, the ground state of the combined system is that of $H_W$ when $\xi$ does not halt, and is that of $H_Z$ when $\xi$ does halt.
This means that essentially any property of the ground state of 2d spin systems is undecidable. For example, by taking $H_W$ to be gapped and $H_Z$ to be gapless, one finds that the gaplessness of the 2d spin systems is undecidable. Similarly, by taking $H_W$ and $H_Z$ to be different symmetry protected topological (SPT) phases protected by a symmetry group $G$, one finds that determining the SPT phase given a $G$-symmetric gapped system is also an undecidable problem.

3.3 Undecidability of supersymmetry breaking

Let us finally come to a case of undecidability in quantum field theory. This subsection is the only new result in this article, and yet it is an immediate consequence of the basic materials reviewed above.5

Our claim is the following. First, there is no algorithm which decides, for each 2d $\mathcal{N}=(2, 2)$ supersymmetric Lagrangian theory given to it, whether it breaks supersymmetry. Second, there is a specific 2d $\mathcal{N}=(2, 2)$ supersymmetric Lagrangian theory which breaks supersymmetry if and only if ZFC is consistent, which can never be proved nor disproved from ZFC.

We show this by resorting to the negative solution to Hilbert’s 10th problem6 we reviewed in Sec. 2.4. Take a Diophantine equation

$$P_\xi(x_1, \ldots, x_k) = 0 \quad (3.3)$$

over $\mathbb{Z}$ corresponding to a Turing machine $\xi$. We now consider a Wess-Zumino model with $2k + 1$ chiral superfields $Y, Z_1, \ldots, Z_k$ and $X_1, \ldots, X_k$, with the superpotential

$$W_\xi = Y P_\xi(X_1, \ldots, X_k)^2 + \sum_a Z_a (\sin 2\pi i X_a)^2. \quad (3.4)$$

Let us now look for supersymmetric vacua of this model. The condition $\partial W_\xi / \partial Z_a = 0$ imposes the condition $X_a \in \mathbb{Z}$. Then the condition $\partial W_\xi / \partial Y = 0$ imposes the condition $P_\xi = 0$. When these two conditions are met, the remaining F-term conditions $\partial W_\xi / \partial X_a = 0$ are automatically satisfied, without restricting the vacuum expectation values of $Z_a$ and $Y$.

We found that this model breaks supersymmetry if and only if the Diophantine equation $P_\xi$ does not have any solution over integers, i.e. if and only if the Turing machine $\xi$ does not halt.

Now, the undecidability of the halting of Turing machines immediately means that there is no uniform algorithm to decide if a given 2d $\mathcal{N}=(2, 2)$-supersymmetric Wess-Zumino model is supersymmetric or not. Furthermore, by using the Turing machine $\xi_{\text{ZFC}}$ which looks for a contradiction of ZFC, we can construct a Wess-Zumino model with the superpotential $W_{\xi_{\text{ZFC}}}$ which breaks supersymmetry if and only if ZFC is consistent, which can neither be proved nor be disproved by means of ZFC itself.

---

5 This was the reason why the author of the present article felt it impossible to make this document public unless as a bedtime reading as a preprint on April Fools’ day. The author thanks an editor of this journal for the invitation to submit and the referees for consideration to publish.

6 Possible relevance of the negative solution to Hilbert’s 10th problem to the issues surrounding string landscape was previously studied in [27, 28]. It is to be noted that the undecidability of the quantum control in [10] also reduces the issue to the undecidability of the Diophantine problem.
4 Discussions

In this paper, we reviewed basic mathematical concepts surrounding undecidability in an informal manner. We then proceeded to recall the mechanisms how such mathematical decidability was translated to the realm of theoretical physics, first in [1–3] in the case of classical mechanics, and second in [4, 5] and the subsequent works in the case of quantum mechanical systems. Finally we presented a minor but new result, that the existence of supersymmetric vacua in $\mathcal{N}=(2, 2)$-supersymmetric Wess-Zumino models is similarly undecidable.

As already mentioned in the introduction and explained in more detail in the main text, the strategy is always the same. Namely, to show that a particular property of a certain class of systems in theoretical physics is undecidable, we only have to devise a way to encode arbitrary Turing machines into the class of systems under consideration. It can be done directly as in [4, 5], or indirectly by first translating to the Diophantine problem as in this work or in [10], or to other mathematical formulations as in [3] which used [19].

This makes it clear that the easier the construction of undecidable problems become, the broader the class of physical systems one wants to consider. For example, in the paper [1] we reviewed, the entire classical physics was considered, in which clearly a computer can be implemented. Therefore it was almost immediate that their question at hand was undecidable, and there was no element of surprise in it.

In contrast, in the case of [4, 5], the class of systems they considered, namely quantum spin systems with finite-dimensional Hilbert space at each site with translation-invariant and finite-range interactions, was felt to be narrow enough, at least to a large percentage of the researchers in that field, in the understanding of the author of the present article. This explains the surprise with which the results of [4, 5] were received, and also the flurry of subsequent works generalizing and extending them.

How should we judge the only new result of this paper, presented in Sec. 3.3, in view of these considerations? It should be pointed out that the superpotential of $\mathcal{N}=(2, 2)$ supersymmetric quantum field theory in two dimensions can be an arbitrary holomorphic function. This was the crucial property which allowed us to encode Diophantine equations easily into this class of systems. Therefore it would not be very fair to say that the result is of great importance, except possibly to the author of this article himself, since it allowed him to dispel his childlike trust in the collective intellect of theoretical physicists in solving questions in theoretical physics.

For a possible direction of future research, we can consider, for example, if we can encode Turing machines into the properties of renormalizable quantum field theories in four dimensions, supersymmetric or otherwise. This is the standard class of quantum field theories high-energy physicists habitually work in, and include the Standard Model of particle physics and their various extensions. The main difficulty, compared to the two-dimensional case considered in this paper, is that the (super)potential of this class of theories only allows polynomials of degrees three (in the case of superpotentials of supersymmetric theories) or four (in the case of potentials of non-supersymmetric theories). Therefore the class is much more restricted, and the encoding of Turing machines would accordingly be more difficult.

The author of the present article does not know if this would be possible or not. Would we learn anything new, if it would turn out to be possible? That is debatable, as the many works e.g. [8–11], which followed the seminal paper [4], only show that the same technique of embedding of Turing machines works in various types of properties of various classes of systems, and the encoding into four-dimensional quantum field theories merely provides
one more. That said, it would surely be intellectually interesting, at least to the author of the present article.

**Acknowledgements** The author thanks Beni Yoshida for his tweet (https://twitter.com/rougeteaviolet/status/1395922985614299137) which introduced me to the wonderful paper [7]. He also thanks helpful discussions with Justin Kaidi and Kantaro Ohmori. The content of this paper was presented first in a lunch talk at IPMU in the autumn, 2021; the author thanks the audience members for the lively and stimulating discussions there. The author is supported in part by WPI Initiative, MEXT, Japan at Kavli IPMU, the University of Tokyo.

This article originally appeared on the arXiv preprint server on April Fool’s day of 2022, in the hope of providing some intellectual entertainment in that dark year. The author thanks the editor of this journal for suggesting him to submit the preprint, which he did not originally have any intention to publish. The editor’s suggestion to submit led to a fruitful refereeing process, allowing the author to improve the manuscript greatly. For this the author would also like to thank the two referees.

This article was written down during a week in January 2022 when he and his wife needed to take care of the children by themselves because the kindergarten was closed due to the omicron variant. One day, while preparing the table for the supper, he told his wife that he finally finished an article for April Fools’ day, to which his wife exclaimed: “We can work only half as usual, because one of us needs to take care of the kids while the other is working, and then you were using that precious time to write an article for April Fools’ day??!” It took the author a couple of minutes to explain that the article was not just to make stupid jokes but to make somewhat unusual philosophical takes on theoretical physics. The author is not sure if that explanation successfully convinced his wife.

**References**

1. Geroch, R., Hartle, J.B.: Computability and physical theories. Found. Phys. 16, 533–550 (1986). https://doi.org/10.1007/BF01886519
2. Moore, C.: Unpredictability and undecidability in dynamical system. Phys. Rev. Lett. 64, 2354–2357 (1990). https://doi.org/10.1103/PhysRevLett.64.2354
3. da Costa, N.C.A., Doria, F.A.: Undecidability and incompleteness in classical mechanics. Int. Journal of Theor. Phys. 30, 1041–1073 (1991). https://doi.org/10.1007/BF00671484
4. Cubitt, T.S., Perez-Garcia, D., Wolf, M.M.: Undecidability of the spectral gap. Nature 528, 207–211 (2015). https://doi.org/10.1038/nature16059
5. Cubitt, T.S., Perez-Garcia, D., Wolf, M.M.: Undecidability of the spectral gap (full version), arXiv:1502.04573
6. Bausch, J., Cubitt, T.S., Lucia, A., Perez-Garcia, D.: Undecidability of the spectral gap in one dimension. Physical Review X 10, (2020). arXiv:1810.01858 https://doi.org/10.1103/physrevx.10.031038
7. Cubitt, T.S.: A note on the second spectral gap incompleteness theorem, arXiv:2105.09854
8. Bausch, J., Cubitt, T.S., Watson, J.: Uncomputability of phase diagrams. Nature Comm. 12, 452 (2021). https://doi.org/10.1038/s41467-020-20504-6. arXiv:1910.01631
9. Watson, J., Onorati, E., Cubitt, T.S.: Uncomputably complex renormalisation group flows. Nature Comm. 13, 7618 (2022). https://doi.org/10.1038/s41467-022-35179-4
10. Bondar, D.I., Pechen, A.N.: Uncomputability and complexity of quantum control. Sci. Rep. 10, 1195 (2020). arXiv:1907.10082
11. Shiraishi, N., Matsumoto, K.: Undecidability in quantum thermalization. Nature Communications 12 (2021). arXiv:2012.13889 https://doi.org/10.1038/s41467-021-25053-0
12. Shiraishi, N., Matsumoto, K.: Undecidability of the fate of relaxation in one-dimensional quantum systems. arXiv:2012.13890
13. 朝井一二三，コンピュータは数学学者になるのか？— 数学基礎論から証明とプログラムの理論へ—，青書社，2015.
14. Yedidia, A., Aaronson, S.: A relatively small Turing machine whose behavior is independent of set theory. Complex Systems 25, 297 (2016). https://doi.org/10.25088/ComplexSystems.25.4.297
15. Davis, M.: Hilbert’s tenth problem is unsolvable. Amer. Math. Monthly 80, 233–269 (1973). https://doi.org/10.2307/2318447
16. Matiyasevich, Y.V.: Hilbert’s tenth problem, Foundations of Computing Series, MIT Press, Cambridge, MA, 1993. Translated from the 1993 Russian original by the author, With a foreword by Martin Davis. https://mitpress.mit.edu/books/hilberts-10th-problem
17. Jones, J.P.: Universal Diophantine equation. J. Symbolic Logic 47, 549–571 (1982). https://doi.org/10.2307/2273588
18. Carl, M., Moroz, B.Z.: On a Diophantine representation of the predicate of provability. Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 407(130), 77–104. https://doi.org/10.1007/s10958-014-1830-2

19. Richardson, D.: Some undecidable problems involving elementary functions of a real variable. Journal of Symbolic Logic 33, 514–520 (1968)

20. Noce, C., Romano, A.: Undecidability and quantum mechanics. Encyclopedia 2, 1517–1527 (2022). https://doi.org/10.3390/encyclopedia2030103

21. Ben-Yaacov, U.: Gödel’s incompleteness theorem and universal physical theories. J. Phys.: Conf. Ser. 1391, 012067 (2019). https://doi.org/10.1088/1742-6596/1391/1/012067

22. Ippoliti, E., Caprara, S.: Undecidability of the spectral gap: An epistemological look. J. Gen. Philos. Sci. 52, 157–170 (2021). https://doi.org/10.1007/s10838-020-09549-9

23. Paulson, L.C.: A mechanised proof of Gödel’s incompleteness theorems using nominal Isabelle. J. Automat. Reason 55, 1–37 (2015). https://doi.org/10.1007/s10817-015-9322-8

24. Paulson, L.C.: A machine-assisted proof of Gödel’s incompleteness theorems for the theory of hereditarily finite sets. Rev. Symb. Log 7, 484–498 (2014). https://doi.org/10.1017/S1755020314000112

25. Voevodsky, V.: Univalent foundations. https://www.math.ias.edu/vladimir/sites/math.ias.edu.vladimir/files/2014_IAS.pdf

26. Buzzard, K.: The future of mathematics? https://www.andrew.cmu.edu/user/avigad/meetings/fomm2020/slides/fomm buzzard.pdf

27. Cvetic, M., Garcia-Etxebarria, I., Halverson, J.: On the computation of non-perturbative effective potentials in the string theory landscape: IIB/F-theory perspective. Fortsch. Phys. 59, 243–283 (2011). https://doi.org/10.1002/prop.201000093

28. Halverson, J., Plesser, M., Ruehle, F., Tian, J.: Kähler Moduli Stabilization and the Propagation of Decidability. Phys. Rev. D 101,(2020). https://doi.org/10.1103/PhysRevD.101.046010

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.