A logical argument against the free will assumption in natural sciences

Tobias C. Sutter BSc MSc

Corresponding author(s). E-mail(s): tobias.sutter@quantagon.at;

Abstract

The debate about the fundamental axioms underlying science does not gain broad attention from the physics community. Only few dedicated groups spend their time investigating these basics. We contribute to this discussion by giving a simple logical argument leading to a contradiction within most scientific theories, based on five reasonable implicit hypotheses. All necessary terms are suitably defined to avoid misunderstandings and to allow a subsequent mathematical treatment. Furthermore, any misleading meanings of terms such as superdeterminism are clarified. After discussing how to potentially avoid the inconsistency, we present our main conjecture which asserts the rejection of free will. This seems to be the most rational route to more comprehensive and complete theories about nature. We further support this conjecture in view of the potential answers it gives to relevant foundational problems in physics. In passing we stress that our conjecture does not eliminate measurement independence. Lastly, it is elaborated how other thinkers approached this question of avoiding the contradiction.

Keywords: Foundations of Science, Implicit Hypotheses, Free Will, Superdeterminism, Measurement Independence

1 Introduction

In the broadest sense, the main goal of all natural sciences is to find causal relations among observational facts of the world. We try to capture natural laws to predict what happens in well-defined experimental situations. To this end, we create theories and interpret them in terms of measurable quantities – only thus does our mathematical description obtain epistemological content. Finally, testing a theory by comparing its predictions with experimental data gives us the ultimate tool for critically deciding between corroboration or refutation of that theory. In other words, we require our theories to be empirically adequate.

An often overlooked trouble in this procedure concerns the postulation of reasonable hypotheses as the foundations of our theories. These serve as our unquestioned presuppositions about the universe, i.e. statements which are not proven to be true by hard empirical facts. This inevitably brings a certain degree of belief into science.

Some of these propositions have to be stated explicitly to meaningfully build up a theory – we call those *explicit hypotheses*. They are specific to and characteristic for each theory, and therefore are debated at great length during development. If there is an unsettling aspect about one or more of those, it is usually resolved after a while. A historic example is the apparent difference between the independently developed matrix mechanics and wave mechanics of quantum theory; after some time, not only their predictive power but also their foundations were rigorously shown to be equivalent [1]. On the other hand, there are also *implicit hypotheses*. These are often taken for granted for a wide variety of different theories and are usually not spelled out at all. Possible troubles stemming from them are more difficult to spot and resolve.

This is exactly where the current short text is meant to step in. The main purpose is to give a feeling that all natural sciences rest upon their unproven axioms, which in turn should be chosen carefully. We do this by critically scrutinising the assumption that humans possess free will and the implications thereof. This is also called *measurement dependence* in the physics literature and has only been seriously investigated seldomly on physical grounds [2–7].

First, we establish the contradictory nature of (some of) the implicit hypotheses of contemporary natural science in section 2. For this we state one possible definition of free will with the additional benefit of making a mathematical treatment possible with some further reasonable restrictions [8]. Some conclusions we may draw from the inconsistency and how to avoid the contradiction is discussed in subsection 3.1. On the way we criticise the fallacious meaning of the term *superdeterminism* mostly attributed to it in the literature. This discussion leads us straight to our main conjecture, presented and shortly defended in subsection 3.2. In the final section 4 we review and comment on existing arguments concerning free will.

2 A logical contradiction in natural sciences

After this brief introduction we will state our main argument right away. It demonstrates how four implicit hypotheses and one basic observation lead to a contradiction within any theory accepting them as true. We stress the fact that the reasoning can indeed be conducted *inside* such a theory without invoking a metalevel. The argument can be summarized as follows:

(P1)	Classical logic is valid.
(P2)	Nature is entirely governed by natural laws.
(P3)	Our theories about nature can and should be logically consistent.
(P4)	Humans are a part of nature.
(P5)	Humans are not entirely governed by natural laws.
(C)	A theory with the above premises is self-contradictory.

Before describing how these five premises lead to a contradiction, let us take a closer look at each one of them to avoid any misunderstandings. Afterwards, it will also be clear that these assumptions really fit most contemporary scientific theories.

(P1) Classical logic is valid. This enables us to make inferences based on a set of premises. It is indispensable for our argument and also for every theory about nature which wants to use inference rules. Without (P1), we are not able to conclude a contradiction from two contradicting assumptions. Basic logic is also needed for developing the tools we use to rigorously describe nature as mathematics can be built upon set theory.

(P2) Nature is entirely governed by natural laws. This vague assumption is fundamental for all natural sciences. We strongly suspect (although we may never be certain) that there are regularities in nature which we can somehow analyze and understand. Still, we want to emphasize here that this premise does neither mention determinism nor indeterminism[1]; it does also not even exclude the possibility that the natural laws could change from one spacetime point to another; and theories capturing regularities of different natural phenomena (e.g. different fundamental forces of nature) may also forever be incompatible. We only require that in principle we can understand every aspect of nature in some detail and ascribe a (potentially random) process to it. Thus, (P2) is compatible with both indeterministic theories such as quantum mechanics, and deterministic ones such as general relativity.

(P3) Our theories about nature can and should be logically consistent. This is a statement about our capability to describe nature. It asserts that if we can understand parts of nature at all, then the corresponding theory does not lead to contradictions. On its own, it neither claims anything about how much of our universe we can analyze in total, nor about whether there is a unified theory of everything. Without this assumption, a contradiction within a theory would not be problematic as all possible theories might be full of contradicting statements. Before proceeding, we want to mention that it is erroneous to conclude that this premise would follow from an argument involving (P1) and (P2). In particular, this could only yield something like "nature is logically consistent", which contains no information about our theories.

(P4) Humans are a part of nature. This is the aforementioned basic observation we need for our main argument. From an objective standpoint this can hardly be denied. The refutation of it would need another hypothesis explaining why humans are special and are to be placed outside of nature – a case which is hard to defend scientifically. Often, such arguments are accompanied by the postulation of an immaterial soul. These viewpoints mostly originate from religion, thereby putting humanity above nature but below god. We leave speculations of this kind to theologians and metaphysicians, and do not consider the rejection of (P4) in the next section.

(P5) Humans are not entirely governed by natural laws. This will be called the *free will hypothesis*. To justify this naming, let us start with a definition: we say that any physical system possesses *free will* if its change of state (its actions) can be

[1] To avoid misconceptions, we mention the distinction between *random* and *lawless* behaviour: we say that fundamentally random systems are subject to natural laws if they are biased in specific situations, e.g. towards some measurement outcome when certain measurement settings are employed. Quantum mechanics with its probabilistic predictions fits this definition. If a physical systems behaviour is always unbiased regardless of the information one might have about it, we say it does not follow any causality relation. So our theories might describe random but not lawless behaviour.

inherently unpredictable for an outside observer. The deliberate phrasing *can be* in the definition allows for calculable behaviour at least some of the time and within some theories. The word *inherently* stresses that not even with all information in the world we could reasonably ascribe probabilities to such a process, i.e. they show lawless behaviour[2]. This definition certainly includes all classical notions of free will, while others (which are usually not labeled as free will, e.g. completely lawless behaviour) are included as well.

(C) A theory with the above premises is self-contradictory.[3] First, premise (P1) is needed to justify the subsequent inferences. Then, by combining (P2) and (P4), we conclude that humans are governed by natural laws without exception, directly contradicting (P5). Finally, the self-contradiction emerges when we apply (P3); it specifically excludes the possibility that our descriptions of nature lead to contradictions. Hence, the set of assumptions state their own inconsistency!

3 How to avoid the self-contradiction

3.1 Alternative hypotheses

Now that we have seen that a theory with implicit hypotheses (P1)-(P5) is inconsistent, let us examine some possibilities of recovering consistency. It is clear that accepting the original premises while adding further, uncontradictory ones is no solution to this problem. Consequently, we must propose alternative premises to replace their original versions in the above argument. There are two main paths we can take. One is to reject classical logic altogether, discussed below as alternative (P1'). The other one is to accept classical logic and modify at least one other premise, explored by (P2'), (P3'), and (P5'). As already mentioned earlier, we abstain from questioning (P4), i.e. the observation that humans are a part of nature.

(P1') Classical logic is invalid. This is definitely the most radical route we can take as it would invalidate our whole argument. In this scenario, basic logical rules might be wrong and a priori we would not know how to arrive at a conclusion from a set of premises. Accordingly, we would also not be entitled to infer the falsity of at least one premise from the contradictoriness of a deductive conclusion. But this was the starting point of considering (P1') in the first place. To break the cycle, we either need to accept a modified version of classical logic or propose a new logic altogether. Regardless, this would need to be incorporated into (P1') to regain an analog of classical logical deduction.

(P2') Not all of nature is governed by natural laws. This is equivalent to the assertion that there are some physical systems devoid of regularities[4]. As a

[2] Note that the authors of [9] defined free will as the property of being independent of all information *available* to a physical system. For this to be meaningful, they fix a causal structure of the spacetime. While this might be the right approach to mathematically cope with it, our definition here is less restrictive. In particular, we do not assume *any* causal structure. In other words: even if one could send a signal back in time to inform their past self about the decisions they are going to make, the recipient of the signal could still decide to do otherwise. In particular, they could decide to never send the signal in the first place. This may already indicate that paradoxes can arise from (P5) on a global spacetime scale.

[3] All of the following reasoning can be carried out *within* any theory that accepts (P1)-(P5) as true – it is essential for the power of the final conclusion that there is no metalevel involved.

[4] (P2) asserts that all physical systems x are subject to some natural law or regularity R, i.e. $\forall x R(x)$. By classical logic, the negation yields $\neg\forall x R(x) \Leftrightarrow \exists x \neg R(x)$.

consequence, any system can be classified depending on its relation to natural laws, which give rise to two disjoint sets of natural phenomena. We may further assume that none of the two sets is empty (we disregard the cases when either all of nature follows certain rules, contradicting (P2'), or when there are no such regularities at all, making our scientific endeavor futile). A thorough analysis of this alternative must answer many difficult questions. How can we decide whether some part of nature does not follow any laws at all, or whether we simply have not understood it properly? And has there always been some portion of the universe without regularities, i.e. already at the big band? Or did some of these regularities get lost at a later (spacetime-)point (or hypersurface)? Maybe only when life or conciousness emerged? What was the event that caused it? Whatever it was, as it happened in an era of complete lawfulness, such vanishing of regularities from physical systems would be a natural law itself. But how can a predictable and calculable event cause the loss of predictiveness? We certainly do not have answers at the present stage of science. However, any proponent of (P2') must acknowledge these questions as valid and eventually answerable.

(P3') Our theories about nature might lead to inconsistencies. This allows our theories to be infested with serious contradictions. By accepting (P3') we cannot refute any theory about nature anymore on purely logical grounds. Moreover, also empirical falsifiability would meet its demise as we might get "correct" but contradicting theoretical predictions about any one experiment. This makes a meaningful comparison of predictions with measured data impossible. As a consequence we might get empirically *in*adequate theories, the scientific value of which can be genuinely doubted.

(P5') Humans are entirely governed by natural laws. In contrast to (P5) we dub this the *no-free-will hypothesis*. It asserts that humanity follows objective rules just as anything else in the universe, in accord with (P2). The assertion of (P5') raises the question: how is the perception of free will possible when there really is no such thing, i.e. how can nature trick us in this regard? This is straightforwardly accessible to scientific methods (e.g. by a collaboration of different research fields such as psychology or neuroscience) – a fact which is in stark contrast to the questions raised by (P2'). As a philosophical consequence we would need to accept that there is nothing special about consciousness; it is rather an emergent phenomenon of physical systems we call alive. Before moving on, we want to clarify the nomenclature. Theories accepting the no-free-will hypothesis are often called superdeterministic (a term first used by John Bell [10] p. 244). We argue that this is misleading as (P5) alone does not imply determinism of nature or human decisions! According to (P2), natural laws can be either deterministic or indeterministic. And (P5') only states that humans underlie some of them, without specifying which ones.

3.2 Main conjecture

Elaborating on possible ways to avoid the self-contradiction of our theories is only the first step: it is one thing to list the options, yet another one to actively choose one. This choice, however, is necessary in order to progress on our quest for more comprehensive theories. Let us shortly consider the reasonableness of the alternatives in order.

First, we do not want to make the rejection of classical logic in the form of (P1') appear as an impossibility. However, it would certainly change our perception of the universe drastically. From our perspective this is not very likely, which is why we leave futher foundational work to mathematicians and logicians. Second, the existence of some physical systems devoid of regularities, asserted by (P2'), is certainly also a valid option. But any theoretical framework accepting this must pose more exact questions than the ones presented above, and answer them in a convincing way. We could not find a thorough and objective discussion of these massive complications in the light of our most corroborated theories anywhere in the literature. Either the questions were not noticed, ignored, or trivialised. As we also do not have any fruitful and formalisable approach, we do not consider (P2') as an alternative to (P2) at the moment. Nevertheless, anyone objecting this decision is invited to convey us. Third, (P3') is also not tenable in our eyes as it leads to empirical inadequateness. It therefore strips our scientific theories from their practical value.

This leaves us with (P5'), which leads us to suggest the following.

Conjecture. *The most reasonable modification to the implicit premises of our theories is to reject the free will hypothesis by replacing* (P5) *by* (P5').

We support this claim by three arguments concerning the simplicity of nature, completeness/closure of theories, and the concept of physical reality, respectively.

First, "reasonableness" in the conjecture is tantamount to pragmatism. Namely, (P5') is the only option leading to directly investigatable questions without requiring additional axioms. Hence, if we value *logical simplicity* by minimizing the fundamental set of hypotheses, it seems unreasonable to go for a different approach. Regarding quantum theory (which is generally assumed to accept (P2')), one step towards our conjecture was prominently made by Everett [11]. Philosophically, his proposal is to model observers as (potentially very complicated) physical automata, which undoubtably follow the same natural laws as everything else. By doing so, he was able to abandon the quantum mechanical measurement postulates altogether, leading to a logically simpler framework. It is interesting to note that his interpretation can be seen as originating from the mathematical formalism of quantum theory itself, rather than being forced onto it from the outside.

The second argument in favor of our conjecture is linked to the fact that humans interact with other parts of nature. We conclude that if there exists a comprehensible theory involving this interaction, it should also incorporate the experimenters decisions. Specifically, decisions of experimenters must not be uncalculable quantities to be inserted into the theory from outside – else, we could call the theory *incomplete* or *not closed* (cf. [12] p.173). In case we cannot (yet) calculate the (probabilistic) decisions of humans, the theory should at least allow for this possibility.

Thirdly, rejecting free will also eviscerates the philosophical problem of observer-independent facts [13] which is linked to the Wigner's friend setup [14]. One can interpret (P5') as including the observer into *physical reality* instead of requiring the independence of the two concepts. Any observer would need to be aware that they are also fully subject to nature's will. Consequently, in case our theories dictate it (which

quantum theory does), they are forced to accept that their actual physical state might be in a superposition and change drastically upon interaction with other systems. How they perceive such situations is not clear as relevant experiments are unfeasible, at least in the near future.

On a final note we want to mention that the no-free-will hypothesis does not exclude the possibility of *measurement independence*. The experimenters choices of possible measurement settings can still be uncorrelated to the physical systems under investigation. This means that the arbitrarily chosen split between observer and observed system may be justified in many situations. Assuming this lack of correlation, the Bell-type inequalities [15][16] can be derived nonetheless. The corresponding calculation with the assumption of (P5') is shown in [8]. Hence, quantum mechanics in its current formulation might still predict nonlocality[5].

4 Concluding Remarks

Of course, no one is forced to accept our main conjecture or even our argument in general. Unfortunately, this lies at the heart of any philosophical debate on the fundamental assumptions of our universe: one can only be more or less convinced by it. This is a basic fact which most writers of philosophy constantly neglect, thereby getting involved in debates where no one can objectively prevail. For that reason we want to repeat the two purposes of this manuscript. First, it is *not* meant to convince anyone unconditionally. We only tried to confront the reader with a set of sensible hypotheses from which they can choose some, but not all at once. Our main conjecture then appears to be the most logical consequence. Yet, one can also get rid of the contradiction otherwise, for example by defining free will differently. This decision is left to the reader. Second, finding the most reasonable set of premises for our theories is a hard task with many subtleties and it is almost certainly far from being solved. The more people think about it, the more our scientific thinking will diversify, potentially leading to more accurate and objective theories. Therefore, the other reason of the current text is to liven the discussion about this topic. Especially because for almost 100 years every serious attempt of questioning the roots has been refuted by the vast majority of scientists without much consideration. Only a handful of researchers and philosophers of science developed the ideas further.

That being said, how could the obvious contradiction we presented above be overlooked by almost everyone arguing in favor of free will? How could e.g. Bell [10] (p. 154), or Shimony, Horne, and Clauser [17] all argue against our no-free-will-hypothesis? As these are all very rational people, they certainly dropped one of the assumptions (P1)-(P4) implicitly to avoid a contradiction. But which one? It seems that they all chose the path of (P2'), namely asserting that not all of nature is governed by comprehensible laws. According to them, the prime example for this is simply any

[5] However, if one consequently follows the Copenhagen interpretation in that the term "reality" can only be meaningfully attributed to measured quantities, there is no problem with locality. For example, applied to a Bell experiment, this means that for either observer, the other ones measurement setting and their result only attain reality upon measurement of them. This measurement can only be realised via local interaction (e.g. communicating the measurement outcome to each other via a classical channel). Thus, the correlations of the measurement outcomes also only occur locally. Nonlocality arises by attributing an independent reality to each observer.

conscious human. While this is an easy statement to make at first, we tried to show that this is not so. The refutation of (P2) in favour of (P2') leads to many difficult and quite inaccessible questions. And it is exactly these questions which were possibly overlooked or at least trivialised by the above mentioned authors.

In any case, we do not wish to attribute thoughtlessness to them as they clearly tried to answer the question: what brings us ultimately closer to a sound understanding of our universe? This is a very noble question indeed! But as it demands a subjective answer, we cannot in good conscience pretend to have a definitive one. We only demand that everyone ponders on it deliberately. Maybe then can we find a decisive answer to the question posed by Kuhn's theory on scientific progress: when is it time for a paradigm shift?

References

[1] von Neumann, J., Taub, A.W., Taub, A.H.: The Collected Works of John Von Neumann: 6-Volume Set. Reader's Digest Young Families, New York (1963)

[2] Brans, C.H.: Bell's theorem does not eliminate fully causal hidden variables. Int. J. Theor. Phys. **27**, 219–226 (1988) https://doi.org/10.1007/BF00670750

[3] 't Hooft, G.: The Cellular Automaton Interpretation of Quantum Mechanics, 1st edn. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-41285-6

[4] Pütz, G., Gisin, N.: Measurement dependent locality. New J. Phys. **18** (2016) https://doi.org/10.1088/1367-2630/18/5/055006

[5] 't Hooft, G.: Free Will in the Theory of Everything (2017)

[6] Hossenfelder, S., Palmer, T.: Rethinking superdeterminism. Front. Phys. **8:139** (2020) https://doi.org/10.3389/fphy.2020.00139

[7] Šupić, I., Bancal, J.-D., Brunner, N.: Quantum nonlocality in presence of strong measurement dependence (2022)

[8] Sutter, T.C.: Measurement Dependence and a generalised CHSH-inequality. To be published

[9] Conway, J., Kochen, S.: The Free Will Theorem. Found Phys 36, 1441-1473 (2006) https://doi.org/10.1007/s10701-006-9068-6

[10] Bell, J.S., Aspect, A.: Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy, 2nd edn. Cambridge University Press, Cambridge (2004). https://doi.org/10.1017/CBO9780511815676

[11] Everett, H.: "Relative State" Formulation of Quantum Mechanics. Rev. Mod. Phys. **29**, 454–462 (1957) https://doi.org/10.1103/RevModPhys.29.454

[12] Peres, A.: Quantum Theory: Concepts and Methods. Springer, Dordrecht (2002). https://doi.org/10.1007/0-306-47120-5

[13] Brukner, C.: A no-go theorem for observer-independent facts. Entropy **20**(5) (2018) https://doi.org/10.3390/e20050350

[14] Wigner, E.P.: In: Mehra, J. (ed.) Remarks on the Mind-Body Question, pp. 247–260. Springer, Berlin, Heidelberg (1995). https://doi.org/10.1007/978-3-642-78374-6_20

[15] Bell, J.S.: On the Einstein Podolsky Rosen paradox. Physics Physique Fizika **1**, 195–200 (1964) https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195

[16] Clauser, J.F., Horne, M.A., Shimony, A., Holt, R.A.: Proposed experiment to test local hidden-variable theories. Phys. Rev. Lett. **23**, 880–884 (1969) https://doi.org/10.1103/PhysRevLett.23.880

[17] Bell, J.S., Shimony, A., Horne, M.A., Clauser, J.F.: An exchange on local beables. Dialectica **39**(2), 85–110 (1985). Accessed 2023-09-05