Synthese (2023) 202:136
https://doi.org/10.1007/s11229-023-04366-1

ORIGINAL RESEARCH

Why computational complexity may set impenetrable barriers for epistemic reductionism

Michael H. Herzog[1] · Adrien Doerig[2] · Christian Sachse[3]

Received: 21 April 2023 / Accepted: 23 September 2023 / Published online: 24 October 2023

Abstract

According to physicalism, everything is physical or metaphysically connected to the physical. If physicalism were true, it seems that we should – in principle – be able to reduce the descriptions and explanations of special sciences to physical ones, for example, explaining biological regularities, via chemistry, by the laws of particle physics. The multiple realization of the property types of the special sciences is often seen to be an obstacle to such epistemic reductions. Here, we introduce another, new argument against epistemic reduction. Based on mathematical complexity, we show that, under certain conditions, there can be "complexity barriers" that make epistemic reduction – in principle – unachievable even if physicalism were true.

Keywords Physicalism · Epistemic reductionism · Complexity · Cryptography · Functional reduction · Multiple realization

✉ Christian Sachse
christian.sachse@unil.ch

Michael H. Herzog
michael.herzog@epfl.ch

Adrien Doerig
adoerig@uni-osnabrueck.de

[1] Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[2] Institute of Cognitive Science, University of Osnabrück, Osnabrück, Germany

[3] Institute of Philosophy, University of Lausanne, Lausanne, Switzerland

1 Introduction

In the last century, the special sciences, such as sociology, psychology, and biology, have developed their own theories and models, which are highly different from the physical ones. Nevertheless, physicalism proposes that the regularities and laws of the special sciences are "just" high-level descriptions of the laws and entities of fundamental physics. With physicalism we refer to the metaphysical position according to which everything is either physical or metaphysically connected to the physical, for instance by a realization or grounding relation (see Stoljar, 2023 for an overview and the Discussion below).

In terms of the standard model of particle physics, all matter in the universe is made of configurations of fermions and all interactions between them are fully governed by the basic four forces mediated by bosons (strong and weak nuclear forces, gravity, electromagnetism). So, intuitively, one should – at least *in principle* – be able to express the processes of the special sciences in terms of particle physics (Oppenheim & Putnam, 1958; cf. Kim, 1998; Papineau, 2001). This is the "classic" idea of the unifying epistemic reductionist program, strongly motivated by the well-known hypothesis of physical closure that does not allow space for non-physical entities and forces (see van Riel & van Gulick, 2019 for different views on this classic idea).

For example, subjective brightness perception of a human is captured by Fechner's law, which states that an increase in perceived brightness (B) of a light patch is equal to the logarithm of the increase of physical luminance (L) of the patch: $B = c \log(L/L0)$, with c being a constant depending on observer, and L0 being the base luminance. Fechner's law may then next be explained in terms of the neural wiring in the early visual system and, if successful, we thus could reduce subjective perception to systems neuroscience. The neural processes may then be explained by physiological processes of the neurons in terms of membrane processes, etc., which in turn are explained by chemical processes of ion concentrations, and so on, eventually finding a fulfilment in terms of particle physics.

To this day, most successful cases of reduction occur mainly in physics, such as the reduction of temperature to statistical physics. Overall, the list of successful theory reductions is puzzlingly short, given how straightforward the reductionist framework seems to be *prima facie*. How can we explain that reduction was achieved centuries ago in some cases, but that most other cases still resist reduction? One very influential explanation is the well-known multiple realization argument (classically, Fodor, 1974; Putnam, 1975; Polger & Shapiro, 2016), which undermines theory reduction.

Importantly, multiple realization is generally taken as *compatible* with physicalism and local reductive, physical explanations (see Chalmers, 1996; Kim, 1998; Kim, 2005), but this compatibility has limits and is controversial (e.g., Hemmo & Shenker, 2022). Here, we introduce *another* argument why epistemic reductions are so rare and may be impossible even if physicalism were true. If, as we call them, "complexity barriers" exist in nature, then epistemic reduction may not be taken for granted. Our argument applies to weak and strong(er) versions of physicalism (see Hemmo & Shenker, 2022; Kim, 1998; Papineau, 2001; Stoljnar, 2023).

2 The complexity argument

Let us introduce our argument with a real-world example. Huntington's disease is a lethal and currently uncurable neurological disorder, i.e., the directly relevant mechanisms are on the brain level. On the clinical level, there are clear-cut symptoms, such as uncontrolled movements. On the genetic level, Huntington's disease is indicated by abnormal long repeats of three base pairs (CAG) on chromosome 4 (Walker, 2007). Clinical symptom severity and mortality are highly correlated with the number of CAG repeats, which vary strongly within the patient population. The more repeats there are, the earlier patients die (with some variability). Hence, there are law-like links between the genetic level and the clinical level. However, at the causally relevant neurobiological level of neurons, glia cells, neurotransmitters, etc., it is impossible to predict mortality or diagnose the disease. Hence, reduction from clinical assessment to the genetic level *via* neuroscience is (at least currently) impossible.

Rationale. Let us formalize the example in a simplified way. There are three levels of scientific explanation: L1 (genetics), L2 (neuroscience), and L3 (clinics), all having proper types and regularities (laws). For instance, there are law-like relations on the clinical level, e.g., symptom X is often associated with symptom Y, and there are law-like molecular processes on the genetic level. In addition, there are law-like relationships between the genetic level L1 and the clinical level L3. For example, as mentioned, the severity of the disease is well predicted by the genetic repeats on chromosome 4. Physicalism implies at least that the regularities (symptom associations) at level L3 supervene on the neural level L2, which supervenes on the genetic level L1 (which ultimately supervenes on the fundamental physics level). So why is reduction so difficult to carry out?

Cryptography and complexity theory may provide an explanation. Well-known mathematical results explain under what conditions it may be *im*possible to find a scientific explanation for symptoms at level L3 based on the neuroscience level L2, even though it causally brings about the disease. These results also explain why we can find law-like relationships between L1 (genetics) and L3 (symptoms), even though we can never find an explanation on level L2, for principled reasons. The simple guiding explanation is: there may be discoverable mappings from L1 to L3, and vice versa, but neither between L1 and L2, nor between L2 and L3. In these cases, reduction of L3 to L1 *via L2* is impossible (Fig. 1). A pathologist can study as many post-mortem brains as they want, without ever being able to figure out what, at the neural level L2, predicts the severity of the disease. This can be true, even though they have knowledge of how the genetic level L1 (# of repeats) is lawfully related to the clinical diagnosis L3. In the following, we present these arguments in detail.

2.1 Formal arguments

1. *Functions cannot always be inferred from data*.

In mathematics, it is well known that there are functions f: X->Y for which one cannot discover f even after seeing arbitrarily many input-output pairs x&y (see Appendix).

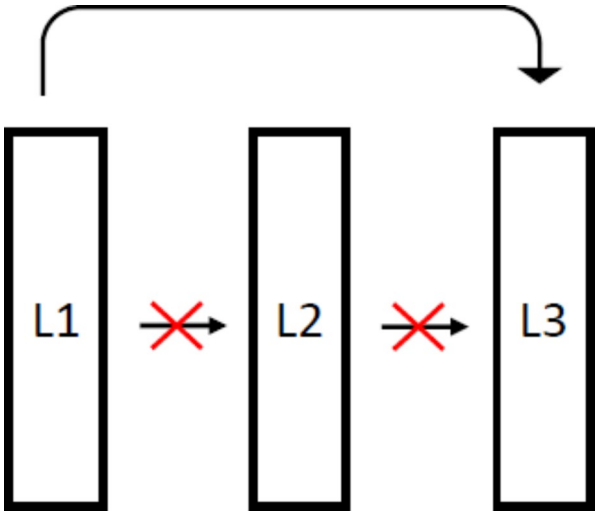

Fig. 1 We may find law like relationships between levels L1 and L3 but not between L1 and L2 and/or L2 and L3 because of complexity barriers (red crosses). Complexity theory and cryptography ensure that, in cases with such a complexity barrier at L2, there can be, still, real links from L1 to L2 and from L2 to L3 that cannot be discovered even when the L1->L3 law-like relationship is known

In other words, for certain functions, observing examples does not help to interpolate how the function behaves in unobserved cases. In the example above, X may be the neural level L2 and Y the clinical outcome level L3: one cannot learn the corresponding function just from the data.

Importantly, these mathematical results are only true for certain functions, not all. For example, linear functions $f(x)=y=ax+b$ can be determined by observing only two input-output pairs. For example, the input-output pairs (0,1) and (3,2) fully determine *f*, leading to $f(x)=1/3x+1$. Now, for all other input values *x*, we can easily know the corresponding output *y*. But for other kinds of functions, it can be shown that they cannot be learned in the same *efficient* way (see Appendix). For example, Boolean functions cannot be learned even if countlessly many input-output pairs are given (and there are many other kinds of unlearnable functions, such as cryptographic functions (Barak, 2017) or the large class of NP-complete problems (Aaronson, 2013). Importantly, this holds true even if we only want to *approximately* learn the function (Kearns & Valiant, 1994).

In summary, if certain bridges between levels of explanation are more like Boolean functions than linear functions, scientists will never be able to discover them and, therefore, reduction cannot be carried out.

2. *Complexity barriers can exist and can be side-stepped.*

Prima facie, one might think that in cases with unlearnable functions L1-> L2 and/or L2->L3, it should also be impossible to learn the function L1->L3. In the Huntington disease example, this would mean that, if the neural level L2 is a complexity barrier, it should be impossible to side-step it by linking symptoms L3 directly to genetics L1. In other words, one may think that complexity barriers should completely cut off higher levels from lower ones.

However, cryptography tells us that this is not true. Cryptographic systems, such as the Diffie–Hellman system or the RSA, have provided explicit methods to reli-

ably transmit information between a sender and a receiver, which we identify with the scientific levels L1 & L3, via an encrypted intermediate state, which we identify with the intermediate level L2. In these systems, the sender sends a message, which is encoded by combining a public and a private key, then the encoded message is transmitted (to L2), and decoded by a receiver using the same keys (at L3). Except for the private key, everything is publicly observable, yet, the encrypted message cannot be decoded without the private key. In other words, what cryptography has achieved is to find a way to "hide" information in complexity, even though the complex state in which the information is hidden is fully observable. These systems exist in real life and are used everyday millions of times, e.g., for safe banking and internet communication.

In our analogy, this means that scientists may have full access to levels L1, L2 and L3, yet be unable to decode the information from the encrypted level L2 if they do not know the private key (i.e., "how nature has encoded the information").

In summary, in cryptographic cases, no matter how many *sent_message->encrypted_state->received_message* observations are carried out, it is impossible to find the general function to decode the *encrypted_state* for unobserved samples. If such cases exist in scientific contexts, reduction is impossible. For example, if the mapping from the genetics L1 to the symptoms at level L3 via neural states at level L2 is cryptographic, then, L2 is a complexity barrier that cannot be learned, even though L1, L2 and L3 are all fully observable. Nevertheless, this complexity barrier can be "side-stepped" by going directly from the genetics level L1 to the symptoms level L3 with clear law-like relationships.

One might optimistically argue that this will not happen in scientific contexts, and that it is only a question of time until a clear-cut picture emerges on the neural level L2 in the Huntington example. Indeed, we can hope that "nature is nice to us" (Lewis, 1994), allowing for the feasibility of epistemic reduction. In the current context, this would mean that mappings between all scientific levels can be learned from data. However, as mentioned, unlearnable functions and cryptography suggest that nature may in fact not always be nice and rather keeps its secrets. Physicalism is silent about the likelihood of complexity barriers and thus cannot entail the – in principle – feasibility of epistemic reduction.

Hence, whether or not complexity barriers exist in the special sciences is an empirical question, but there are reasons to think that they do. Krakauer (2015) has argued that cryptography may be ubiquitous in biological systems. Lazebnik (2002) argued that small radios, which are much simpler than a living cell, cannot be understood in terms of their constituent parts using the techniques commonly used by biologists. Jonas and Kording (2017) showed that simple microprocessors with far fewer components than a small brain region cannot be understood in terms of their constituent parts using tools commonly used in neuroscience.

Next, we will present a simple example to illustrate how easily complexity barriers can occur.

A thought experiment illustration. Assume a new animal species is discovered. Researchers find out that, if an animal of this species is presented with a red patch, it always moves the right limb. Being presented with a green patch, it always moves the left limb. Researchers analyse the brain and find 60 neurons, which are each either

active (1) or inactive (0) at each moment of time. In this example, the data at L1 is the external stimulus (red, green), L2 is the 60 neurons, and L3 is the motor behaviour. The function from L1 to L3 is f(red)=right, f(green)=left. Experimenters show the animal either the red or green patch and monitor the activity of each neuron (yielding an activity vector of sixty 1s and 0s) in each case, giving them a fully observable neural state at L2. Performing this experiment with many trials, researchers realize that each patch presentation elicits a *new* activity vector almost every time, i.e., a new vector of sixty 1s and 0s. Even after thousands of trials, they are unable to find a pattern in the neural activity vectors (L2), i.e., there is no way to tell which brain activity is related to "red" and which is related to "green" (and the corresponding limb behaviour).

This situation is analogous to the cases explained earlier. Even though L1 (red/green) and L3 (right/left) are lawfully linked, observers cannot infer from data how this link is mediated by the neurons in L2. They can discover "nothing", so to speak, that the neural activity vectors have in *common* when the red/green patch is followed by the right/left movement, even though the neurons *deterministically* link color sensations to limb movements. Put differently, it is impossible to find a mapping or rule allowing the researchers to link neural activity to outputs (and inputs) even though there are only two possible inputs and outputs and the inputs are law-like linked to the outputs. The complexity and cryptography results described above ensure that this situation may occur.

For such complexity barriers to occur, it is crucial that coding is truly *combinatorial*, i.e., focusing on single neurons does not exhibit any information in the above example. To illustrate, let us simplify the example even more and assume the animal's brain contains only two neurons coding for red and green in the following way: for red, either both neurons are activated (1,1) or neuron 1 is activated but not neuron 2 (1,0); green is coded by (0,1) and (0,0). Assume all four possibilities occur with the same probability. In this case, focusing on neuron 1 shows a series of equally frequent 1s and 0s. There is absolutely no correlation with the colors. Same with neuron 2. Hence, decoding is impossible. Decoding is only possible when focusing on *both* neurons. This means that we need to decode in the *combinatorial space*. With two neurons, the combinatorial space is very small (4 states), so decoding is possible. The crucial fact is that decoding becomes impossible if the number of states just moderately increases. For example, if the brain contains just 60 neurons, it is simple and quick to measure the activation of each neuron at a given moment of time. However, the number of neural activity vectors is 2^{60}, a number larger than the number of seconds in the universe (counted on from the Big Bang). Hence, a scientist measuring one 60-dimensional activity vector per second would need longer than the age of the universe to map out the full state space. In other words, *very small structures* can create *very large state spaces*. What matters for the behaviour is not the number of the states of the *single* neurons (60) but the number of their combinations (2^{60}).

This is the essence of complexity: one variable of the problem, e.g., the number of neurons in the above example, grows linearly whereas the number of relevant coding states grows *exponentially* when coding is truly combinatorial.

Importantly, these complex combinatorial problems are not *complicated* to solve, they are *impossible* to solve given the *finiteness* of the world. Indeed, whatever the

resources of science are, we can always find a state space whose number of combinatorial states clearly exceeds the resources. For example, if brain measurements can be carried out 1000 times per second (instead of 1 time per second as in the previous example, i.e., a factor of 1000 times faster), then measuring the full state space of 69 neurons (instead of 60 in the previous example, i.e., just 9 more neurons) takes more time than the age of the universe. Hence, no matter the speed and resources of scientific equipment, an unsolvable problem can be created. 69 neurons is minuscule compared to the human brain, which contains more than 100 billion neurons, and even compared to a small animal such as the 1 mm long nematode *Caenorhabditis elegans* with its 302 neurons.

Potential objections. First, it might be objected that scientists can do more than observing states. For example, they might study the *connectivity* of the neurons and, thus, find the causally relevant structure to discover the system's rule. This objection may be inspired by having a mechanistic model of explanation in mind (e.g. Craver, 2006). Indeed, the connections between neurons account for the different forms of neural activity and, thereby, of different behaviours. Still, the very same reasoning applies: the pattern of connectivity generally is even more complex than the neural states themselves. For example, there are 2^{n^2} directed graphs with n neurons, i.e., there are more ways of connecting $n=8$ neurons than seconds since the Big Bang (assuming each neuron has a binary on/off connection to each neuron including itself).

Second, it might be argued that there are successful examples of reduction, such as the reduction of thermodynamics to statistical mechanics. This is true. However, our claim is not that reduction is *always* impossible. Our claim is that there *might* be cases where reduction is impossible. Importantly, the reason why reduction was possible in the case of thermodynamics is that it was (relatively) easy to find *types* of micro-states corresponding to temperature T (namely, all micro-states with an average energy per atom of $E=(3/2)kT$). In other words, this is far from a complex situation. However, imagine if the temperature depended on the specific location and speed of each atom in a much more complicated way – a way that can demonstrably not be identified in finite time, like the Boolean example given in the appendix. In that case, even though the temperature is linked to the microstates by this complicated law, science could never find it. It would be hidden in complexity. This shows that what matters for reduction are not simply the degrees of freedom of a system (e.g., the number of neurons in the above examples). Rather, what matters is whether there are easily identifiable regularities (i.e., types), allowing to find a description with fewer "effective" degrees of freedom. When they do not exist, there can be no reduction.

3 Discussion

According to our best physical theories, all matter is composed of fermions and all processes in nature are emanations of the basic four forces, mediated by bosons. Strong physicalism is built on the controversial hypothesis of the causal, nomological and explanatory completeness of physics (cf. Papineau, 2001). This is often further

specified as an ontological reductionist position in the sense of *token-identity* of any *causally efficient* higher-level property with its physical supervenience base (cf. Kim, 1998). Other versions of physicalism (see Stoljar, 2023 for an overview) and important notions like "causation", "law", "explanation" (see Ben-Menahem, 2018) are under debate. Importantly, our argument is *independent* of the specific type of physicalism, the nature of causation, etc., as long as there are different levels of explanations, such as L1, L2, L3, etc. in the special sciences. Our argument casts doubt on the implied epistemic suggestion that we should – in principle – be able to reduce theories, models, laws, and regularities of the special sciences to the fundamental physical theories, models, laws, and regularities (or at least to build reductive physical explanations for these entities and processes) over the years of scientific progress. Such epistemic theory reduction and reductive explanations are only possible if there are no complexity barriers. If there are barriers, epistemic reductionism is impossible. Metaphorically speaking, the only thing scientists could do in such situations is to create a telephone-book-like *list* linking arbitrary *tokens* of one scientific level to another one. However, this is not reduction, which classically requires correlating "higher-level" to "lower-level" *types*, or at least linking types to species-specific types (cf. Kim, 1998; Kim, 2005) or to functionally defined sub-types (cf. Esfeld & Sachse, 2011). Again, metaphorically speaking, reduction requires structures, not just tokens. In addition, even if one would wish for such a "telephone book", it would take longer than the age of the universe to create it.

Philosophical considerations. First, our argument can be complexified in terms of synchronic reduction and diachronic causation. For example, the genetic level L1 at time t0 *diachronically causes* the neuroscientific brain state at a later time t1, e.g., by determining which proteins are produced, which in turn alter neural function. In contrast, reduction is generally understood as a *synchronous* relationship asking, for example, whether the clinical level L3 at time t2 can be *synchronically reduced* to the neuroscientific brain state at the same time t2. Importantly, our argument applies both to diachronic causation and synchronic reduction because our argument relies only on the notion of the functions between the scientific description levels and not on the mechanisms underlying the functions.

Second, as said before, the aims of our analysis are of *epistemological* nature only. Even if one adopts a relatively strong version of *physicalism*, like that of *token-identity* of any higher-level property with something physical, complexity barriers block epistemic reduction and reductive explanations since both require the identification of *types.*

Third, our complexity argument is different from the well-known multiple realization argument. The latter essentially proposes that the special sciences point out salient effects that configurations of physical property tokens have *in common* under standard environmental conditions, although they are actually composed in *different* manners and therefore are of *different* physical types (Fodor, 1974; Kim, 1998, 2005). In other words, different physical types (say "P_1", "P_2", "P_3") can each be of a different *realizer type* (synchronic relation) of a certain high-level biological property type (say "B"). For instance, certain types of physically/molecularly different DNA sequences can lead to the same protein and thus to the same salient phenotypic effect (characteristic functional effect of "B") because of the redundancy of the genetic

code. Still, each realizer type has regularities and applies to a *set of tokens* (it is not merely a single realizer token), which may allow us to cope with multiple realization and to build eventually (realizer type specific) reductive explanations (see Chalmers, 1996; Kim, 1998; Kim, 2005) that may also be understood as mechanistic explanations (Craver, 2006).

Simplified, think of a gene type "B" (coding for a protein) that is multiply realized in a population by molecularly different DNA sequences ("P_1", "P_2", "P_3"). Given the knowledge of molecular genetics, a reductive explanation can be built about how each of these three types of DNA sequences leads to the production of the very same protein (Esfeld & Sachse, 2011). Importantly, each reductive explanation allows us to explain the functioning of *many* (not only a single token) of the studied organisms since genes may only sometimes vary molecularly from one organism to the other. Furthermore, given the knowledge of the redundancy of the genetic code, it is possible to *predict* what further, not yet observed, molecular realizations (say "P_3", and "P_4") of the very same gene are possible.

In contrast, a crucial part of our complexity argument is that there may be *no discoverable realizer types at all* (and thus no types of reductive explanations even within an individual). In terms of our thought experiment above, there simply may be no types of the neural activity vectors that can be identified as realizer types for linking the red-patch-input to the right-limb-behaviour. In contrast, in multiple realization, it is assumed that not only within an individual, such as our animal above, there are realizer types but also that these realizer types apply to *many* tokens, such as the entire species for example. Hence, our argument is much stronger than arguments from multiple realization. Therefore, in contrast to the multiple realization argument, which poses problems in cases where we want to find what all low-level implementations of a high-level type have in common, our argument applies within *one single system.*

Fourth, there is nothing fundamentally wrong with the general framework of functional reduction (Kim, 1998, 2005), including its eventually possible extension to a fully-fledged theory reduction (Esfeld & Sachse, 2011). Let us here shortly summarize the three characterizing steps of functional reduction (cf. Kim, 2005, Chap. 4) and then comment on it. Firstly, any functional property is defined in terms of a causal role, and that causal role requires a realization by some configuration of fundamental physical tokens. Secondly, scientists discover such physical realizers because they bring about the effects that characterize the functional property in question. Thirdly, one shows how the physical description of any such configuration of tokens explains why the configuration in question is a case of the functional property in question (reductive explanation). The complexity argument is not against any of these three steps, but it lays stress on an *implicit* weak point in such a reductionist position, namely that it takes for granted that discovering realizer tokens on a given occasion is sufficient to identify realizer *types* for then building reductive explanations that apply *in general* (e.g., to more than one token, to more than one organism at one single moment of its life).

Fifth, there is an interesting link worth exploring further between the complexity argument and the ongoing debate on *causal specificity*. Beyond the classical question of distinguishing *causal* from *non-causal* relationships, there is an increasing focus

on various distinctions *among* causal relations, for instance in terms of causal "stability", "proportionality" and "specificity" (Woodward, 2010). The central idea of causal specificity is that "some causes allow a much more fine-grained control over their effect variable than others" (Weber, 2017, 575). Generally, reductive explanations, by adding causal knowledge, involve an identification of what is causally more (or most) specific in a causal relationship. For example, the so-called "coding" parts of DNA have been identified to be causally more specific for the protein synthesis than other causally involved parts (in the causal relationship between genotype and phenotype (Weber, 2017). In the presence of complexity barriers, reductive explanation may not allow the identification of causally more specific parts. No proper subset of the entire activity vector of 60 neurons is causally more specific than any other for bringing about the behaviour of the animal; there is nothing, so to speak, that allows for a more fine-grained control.

Sixth, our results may be seen as limits of physicalism on one hand. On the other hand, our argument may explain why the special sciences and its laws and regularities *appear* to be fundamentally different than the laws of particle physics. In fact, much of the discussion of multiple realization comes exactly from this apparent difference, fuelling the debate. For this reason, weaker versions of physicalism give the special science a kind of independent status. Others deny this special status, question multiple realization, and give reasons why the special sciences appear as so different (cf. Hemmo & Shenker, 2022). Here, our argument is relevant because complexity barriers may hinder reduction and, thus, detach the special sciences from physics epistemologically.

Scientific considerations. Our complexity argument states that it may be impossible to reduce higher scientific levels to lower ones because there are complexity barriers. Whether such barriers exist is an empirical question. We will never be able to prove that a topic includes a complexity barrier. All we can do is prove that a system has *no* complexity barriers, by finding how to reduce it. It may well be that certain topics appear to be complex today, but will be simplified with future methods. However, several factors suggest that complexity barriers really occur in nature, as mentioned above.

Noise and deterministic chaos can create complexity but they are very different from barriers. In the animal example discussed earlier, strong noise would create an impossible decoding situation for the scientist as well as a barrier but the noise would also corrupt the link between perception (red/green) and movement (right/left). In contrast, complexity barriers, based on cryptography, do not hinder robust functioning, they simply hide information in the complex state.

Importantly, complexity barriers do not preclude scientific progress. Even when there is a complexity barrier, scientists can simply "ignore" it and work on the levels for which clear-cut patterns can be observed, as in the example of Huntington's disease. Likewise, no biologist tries to reduce mitosis to fundamental physics. Instead, biologists work at the biological level, where many stable lawful relationships can be observed, and particle physicists work at the particle physics level, with its own lawful relationships. The fact that there is likely a complexity barrier in between does not hinder this intra-level research. Problems may arise if the level with the complexity barrier contains crucial aspects for a scientific question. For example, if the cure

for Huntington's disease could only be found on the neural level, then our inability to link it with symptoms would be a serious problem. Hence, our results may help scientists to decide what to do when their science gets "stuck", for example when more and more factors are discovered underlying a phenomenon, each explaining very little of the variance.

Our results are not aimed to define complexity, a notoriously difficult task (Ladyman et al., 2013; Simon, 1991). For example, Ladyman et al. (2013) discusses several aspects, such as linearity, numerosity, hierarchy, and shows that no single one of them is necessary and sufficient to capture the intuition about complexity. Barriers might be an additional aspect of complex systems. In addition, our arguments support the notion that numerosity of the system's elements is not necessary for a complex system since a few elements can create large state spaces. What matters is whether or not the state space is truly combinatorial.

In summary, using mathematical results from complexity and cryptography, we have provided an argument suggesting that, even under physicalist assumptions, epistemic reduction cannot be taken for granted. This kind of reasoning is in line with Aaronson (2013), who proposed that complexity has many important things to tell philosophy. While future work is needed to refine our characterization and understanding of complexity barriers and when they occur, the present results offer a framework to understand the relative scarcity of successful reductions.

Appendix

In the above example, the experimenter must infer from a number m of neural recordings the corresponding behavior/stimulus. Let each recording be given by an n-vector $x = (x_1,\ldots,x_n)$ with the x_i being the activity of one neuron, i.e., $x_i \in \{0,1\}$. Each activity corresponds to one behavior, hence, x is mapped onto 1 (left foot behavior) or 0 (right foot behavior), $\phi(x) = \phi(x_1,\ldots,x_n) \rightarrow \{0,1\}$. Here, we show that, for certain mappings ϕ, it is impossible to infer this mapping from a polynomial number of observations. We argue by means of reduction using results from intractability of learning (Kearns & Vazirani, 1994).

If, for example, the mapping is a 3-term logical DNF formula, it is impossible to infer this mapping from a polynomial number of observations. A 3-term logical DNF formula ϕ is, for example, of the form $(x_1 \wedge \neg x_3 \wedge x_8) \vee (x_2 \wedge x_3 \wedge \neg x_7) \vee (\neg x_4 \wedge x_5 \wedge x_6)$ with $\wedge$ the logical AND, $\vee$ the logical OR, $\neg$ the negation. Thus, as an example, $\phi(0,1,1,0,0,1,0,1) = 1$. The experimenter must be able to efficiently construct such a mapping ϕ which is consistent with the observations. Pitt and Valiant (1988) show that this precise step cannot be carried out efficiently from a polynomial number of observations if a widely believed conjecture, namely $P \neq NP$, holds. It can be shown that it is even impossible to infer hypotheses ϕ' which are "close" to ϕ in the sense that the probability that ϕ' differs from ϕ on a random sample is less than ϵ: $\phi - \phi' \leq \epsilon$. In addition, Kearns and Valiant (1994) have shown that under the assumption that certain number-theoretical problems are hard to solve, similar results hold independently from the representation of the target formula.

We do not claim that all mappings $x \rightarrow \{0,1\}$ cannot be learned. Quite to the contrary, many mappings are easy to learn, e.g. $x \rightarrow \{0,1\}$, $\phi(x)=x_1$. We claim only: there are mappings which cannot be learned, e.g. the 3-term logical DNF-formulae. Hence, we have shown that reduction *may* be impossible. It is an empirical question whether or not this is the case in a particular situation.

Acknowledgements We would like to thank Hilal Lashuel for advice on the Huntington Disease argument, Friedrich Eisenbrand for mathematical insights and Michael Esfeld and Patrice Soom for useful discussions. This work was funded by the Swiss National Science Foundation grant n.176153 "Basics of visual processing: the first half second".

Funding Open access funding provided by University of Lausanne

Declarations

Conflict of Interest The authors declare to have no conflict of interest (Disclosure of potential conflicts of interest).

References

Aaronson, S. (2013). Why philosophers should care about computational complexity. *Computability: Turing Gödel Church and Beyond*, *261*, 327.

Barak, B. (2017). The complexity of public-key cryptography. *Tutorials on the Foundations of Cryptography: Dedicated to Oded Goldreich* (pp. 45–77). Springer International Publishing.

Ben-Mehanem, Y. (2018). *Causation in science*. Princeton University Press.

Chalmers, D. J. (1996). *The conscious mind. In search of a fundamental theory*. Oxford University Press.

Craver, C. (2006). When mechanistic models explain. *Synthese*, *153*, 355–376.

Esfeld, M., & Sachse, C. (2011). *Conservative reductionism*. Routledge.

Fodor, J. A. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, *28*, 97–115.

Hemmo, M., & Shenker, O. (2022). Flat physicalism. *Theoria*, *88*, 743–764.

Jonas, E., & Kording, K. P. (2017). Could a neuroscientist understand a microprocessor? *PLoS Computational Biology*, 13(1), e1005268.

Kearns, M., & Valiant, L. (1994). Cryptographic limitations on learning boolean formulae and finite automata. *Journal of the ACM (JACM)*, *41*(1), 67–95.

Kearns, M. J., & Vazirani, U. V. (1994). *An introduction to computational learning theory*. MIT Press.

Kim, J. (1998). *Mind in a physical world*. MIT Press.

Kim, J. (2005). *Physicalism, or something near enough*. Princeton University Press.

Krakauer, D. (2015). Cryptographic Nature. *arXiv preprint arXiv:1505.01744*.

Ladyman, J., Lambert, J., & Wiesner, K. (2013). What is a complex system? *European Journal for Philosophy of Science*, *3*(1), 33–67.

Lazebnik, Y. (2002). Can a biologist fix a radio? Or, what I learned while studying apoptosis. *Cancer cell*, *2*(3), 179–182.

Lewis, D. (1994). Humean supervenience debugged. *Mind* 103, 473–490. Reprinted in D. Lewis (1999). *Papers in metaphysics and epistemology*, 224–247. Cambridge University Press.

Oppenheim, P., & Putnam, H. (1958). The unity of science as a working hypothesis. In H. Feigl, G. Maxwell, & M. Scriven (Eds.), *Minnesota Studies in the philosophy of Science, II* (pp. 3–36). The University of Minnesota Press.

Papineau, D. (2001). The rise of physicalism. In C. Gillett, & B. Loewer (Eds.), *Physicalism and its discontents* (pp. 3–36). Cambridge University Press.

Pitt, L., & Valiant, L. (1988). Computational limitations on learning from examples. *Journal of the ACM*, *35*, 965–984.

Polger, T., & Shapiro, L. (2016). *The multiple realization book*. Oxford University Press.

Putnam, H. (1975). The nature of mental states. In: H. Putnam: *Mind, language and reality. Philosophical papers. Volume 2*, 429–440. Cambridge University Press. First published as "Psychological predicates" in W. H. Capitan and D. D. Merrill (eds.) (1967): *Art, mind and religion*. University of Pittsburgh Press.

Simon, H. A. (1991). The architecture of complexity. *Facets of systems science* (pp. 457–476). Springer.

Stoljar, D. (2023). Physicalism. *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), forthcoming https://plato.stanford.edu/archives/sum2023/entries/physicalism/.

Van Riel, R., & van Gulick, R. (2019). Scientific reduction. *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/spr2019/entries/scientific-reduction/.

Walker, F. O. (2007). Huntington's disease. *The Lancet*, *369*(9557), 218–228.

Weber, M. (2017). What kind of causal specificity matters biologically? *Philosophy of Science*, *84*, 574–585.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanations. *Biology and Philosophy*, *25*, 287–318.