The oxidative phosophorylation controversy in the light of epistemic complexity

Luigi Scorzato

Abstract The oxidative phosophorylation controversy has been analysed in great detail from the philosophy of science point of view. Hence, it represents an ideal use case to review, in detail, how theory chance can be explained in terms of accuracy and epistemic complexity alone.

1 Introduction

In a series of papers, Scorzato (2013, 2016, 2015a,b, 2024, 2025) put forward a model of theory change that relies only on the well-defined notions of *accuracy* and *epistemic complexity* (The *EC model*, for short, in the following). The EC model represents a radical challenge to the dominant view (Kuhn, 1962; Douglas, 2013; Keas, 2018) that "theory change" depends, necessarily, on a changing set of epistemic and non-epistemic values that are themselves never clearly defined. Many examples of historical scientific theory change have been provided in the aforementioned papers to support the accuracy of the EC model and no counter-example has been published to date. However, it has been argued that it is not always clear how exactly the scientists are supposed to use EC given that they never explicitly refer to it.

Because the oxidative phosophorylation (ox-phos) controversy has been analysed in great detail from the historical and philosophical point of view (Weber, 2004; Scholl and Nickelsen, 2015), it offers an opportunity to review the practical applicability of the EC model. The ox-phos controversy is particularly interesting because we need to understand not only why it was resolved, but also why it took so long to reach a resolution. In the last half century, many philosophers of science after Kuhn (1962) have used the existence of scientific controversies as evidence that scientific theory change cannot be defined by simple principles. This is neither justified nor plausible. It is not plausible because, while this idea could explain the controversies, it makes it impossible to understand their eventual definitive resolution. It is not justified because the presence of a *debate* does not imply the lack of *clear definitions*. An obvious example are the scientific measurements themselves: they can be very difficult to perform, they can take very long (sometimes decades) to complete and spur much debate. But they still aim at measuring something that is usually (although, not always) well defined. A precise definition does not eliminate the debate about the assessment of the corresponding quantity: it makes the debate rational. Couldn't the same be true for theory change? Couldn't theory selection be based on well-defined properties, whose assessment, however (like any scientific property), is always approximate and often difficult? The ox-phos controversy provides evidence of this thesis.

In the following Sec. 2 I review the EC model (Scorzato, 2013, 2025). In Sec. 3 I apply it to the ox-phos controversy. Finally, in Sec. 4, I draw my conclusions.

Accenture AG, Genève, Switzerland. E-mail: luigi.scorzato@accenture.com

2 Theory change according to the EC model

According to (Scorzato, 2025), model A is better than model B when A is either more accurate or simpler then B, without being inferior in any of these aspects. Accuracy represents a multi-dimensional measure: A is at least as accurate as B if it is so in all measured quantities and it is more accurate than B if, additionally, is more accurate in at least one quantity[1]. The *epistemic simplicity* (also referred to as *conciseness*) of a model $\mathcal{M}$ is defined (Scorzato, 2025) as the inverse of *epistemic complexity (EC)* $\mathcal{C}$:

$$\mathcal{C}(\mathcal{M}) := \min_{\mathcal{M}' \equiv \mathcal{M}} \text{length}\left[P(\mathcal{M}')\right], \tag{1}$$

where $P(\mathcal{M})$ is a string containing all the assumptions of the model $\mathcal{M}$, length is syntactic character length and the minimum is taken over all possible formulation of $\mathcal{M}$ that are logically and empirically equivalent to $\mathcal{M}$.

When scientists compare two alternative models, say $\mathcal{M}_1$ and $\mathcal{M}_2$, they first consider which further (possibly ad-hoc) assumptions are needed to make them equally empirically accurate (on the present data). Of course, each $P(\mathcal{M}_i)$ must include all the assumptions needed to derive the empirical consequences of both models and perform the comparison. Once the two (augmented) models are equally accurate, it makes sense to compare the complexity of their full assumptions. The complexity in Eq. (1) cannot be computed exactly, but it can be estimated. Because it is very challenging to discover new formulations of $\mathcal{M}$ that are logically and empirically equivalent to $\mathcal{M}$, but more concise, it is justified to use the ordinary formulation of a model as a basis to estimate Eq. (1). Like all quantities in science, also EC must be defined precisely, to be able to discuss it rationally, but it is estimated approximatively.

In many cases, even after adding all the ad-hoc assumptions necessary to make them equally accurate, we cannot conclude that either $\mathcal{M}_1$ or $\mathcal{M}_2$ is better than the other, because they both need only a few ad-hoc assumptions and it is not clear which one is more complex. This is the case, in particular, when scientific controversies are still unresolved. To solve the controversy, the scientists typically design new experiments meant to reveal those phenomena where the two models may predict different outcomes. The idea is to generate new data that fit well one model as it is, and force the other model to incorporate more and more ad-hoc assumptions. If successful, this strategy forces so many new ad-hoc assumptions predominantly in one model (say $\mathcal{M}_2$) that the difference $\mathcal{C}(\mathcal{M}_2) - \mathcal{C}(\mathcal{M}_1)$ becomes unambiguous. When this happens, the scientists often say that model $\mathcal{M}_2$ has been ruled out by the new experiments, without referring to anything like Eq. (1). But, this conclusion contradicts a vast and very convincing philosophical literature that excludes the possibility of defining the concept of *experimentum crucis* on purely empirical grounds (Ariew, 2026). However, a justified observation is that "model $\mathcal{M}_2$ requires too many ad-hoc assumptions without any empirical advantage with respect to the alternative model $\mathcal{M}_1$". This is exactly the idea that Eq. (1) is able to capture and justifies the rejection of model $\mathcal{M}_2$, consistently *to most practical purposes, even if not literally,* with the scientists' conclusions.

In the following section we will see a concrete example of this dynamics, while I refer to (Scorzato, 2025) for more details on EC.

3 The ox-phos controversy in light of the EC model

In Sec. 3.1, I briefly review the key historical events that have defined the ox-phos controversy (1961-1977), to the extent that is necessary in this paper. In Sec. 3.2, I reconstruct the debate under the light of the EC model, showing why the disagreement was initially natural and why it was eventually resolved. Finally, in Sec. 3.3, I examine previous accounts of the ox-phos controversy, highlighting how the EC model partially agrees with them, but it also overcomes their limitations.

---

[1] Because accuracy is multi-dimensional, the notion of *scope* is naturally included in the notion of accuracy (Scorzato, 2025).

3.1 Historical recapitulation of the ox-phos controversy

In this section, I briefly review the historical milestones of the ox-phos controversy, following the excellent exposition in (Weber (2004), Chapter 4), that highlights the logical steps that are philosophically relevant.

By the end of the 1950s, several key components of cellular metabolism were well-established. These included the glycolysis phase, the citric acid (Krebs) cycle, the basis of respiration and the role of ATP as universal carrier of biological energy. However, one question in particular remained prominently open among others: "how do cells convert energy from respiration into ATP?" The conflict primarily involved two competing theories: the Chemical Hypothesis and the Chemiosmotic Hypothesis.

The Chemical Hypothesis, proposed by E.C. Slater in 1953 (Slater, 1953), suggested that energy from respiration is stored in a *high-energy intermediate* (often referred to as X). The intermediate X would then acquire a phosphate group to eventually form ATP. This theory dominated the field because it was modeled after the known mechanism of glycolysis.

The Chemiosmotic Hypothesis, proposed by P.D. Mitchell in 1961 (Mitchell, 1961), argued that no chemical intermediate existed. Instead, Mitchell proposed that respiratory enzymes pump protons across the mitochondrial membrane to create an electrochemical gradient, which then powers the synthesis of ATP through an enzyme called ATPase.

When Mitchell proposed his model in 1961, the scientific community did not pay too much attention. The overwhelming majority of the biochemist community expected that the intermediate X would be eventually found. Indeed, at least seventeen claims of the isolation of X were published between 1956 and 1972, but did not survive closer inspection (Allchin, 1997).

An important event was the publication of the "Acid Bath" experiment by Jagendorf and Uribe (1966), which showed that chloroplasts could synthesize ATP in the dark, if an artificial proton gradient was created. While this supported Mitchell's model, opponents argued it could be an irrelevant side effect rather than the actual mechanism of photosynthesis. However, Jagendorf's experiment attracted attention on Mitchell's model and the community started to pay closer attention to it.

Jagendorf's experiment also triggered more experiments by Mitchell's group. The most significant ones are summarized below:

- Proton Ejection (Test 1): Mitchell and Moyle demonstrated that mitochondria eject protons during respiration (Mitchell and Moyle, 1965b, 1967). Opponents countered that this could simply be an "energy-linked function" (a secondary effect) powered by the hypothetical chemical intermediate, rather than the cause of ATP synthesis itself (Slater, 1967).
- Uncouplers (Test 2): Mitchell showed that uncoupling agents (which stop ATP synthesis) make membranes permeable to protons (Mitchell and Moyle, 1967). Opponents argued instead that these chemicals simply destroyed the hypothetical chemical intermediates (Greville, 1969).
- Intact membrane (Test 3). Mitchell's model requires an intact mitochondrial membrane. This is consistent with a number of experiments that tried to reproduce oxidative phosphorylation in-vitro: they all failed unless the membrane was intact (Mitchell and Moyle, 1965a). However, opponents argued that respiratory enzymes might also need membranes for mechanical support or structural integrity (Racker, 2012).
- Asymmetry of membranes (Test 4). Under the Chemiosmotic hypothesis the function of the membrane is asymmetric because the movement of protons has a definite direction (and is reversed if the membrane is reversed) (Mitchell and Moyle, 1965a). However, opponents argued that energy-linked functions also require an asymmetric membrane.

The controversy was finally resolved in the 1970s through the experiments led by Efraim Racker. In a first experiment (Test 5) Racker's team inserted purified ATPase, phosphate, ADP and respiratory enzymes (cytochrome c and oxidase) into artificial phospholipid vesicles. These simplified systems suc-

cessfully synthesized ATP (Racker and Kandrach, 1971). For the first time the reaction was observed outside intact mitochondria.

In a further experiment (Test 6), Racker and Stoeckenius (1974) used the same artificial vesicle, but replaced the respiratory enzymes with bacteriorhodopsin (a light-driven proton pump from bacteria). This system also synthesized ATP, when exposed to light. Because this light-driven system contained no respiratory enzymes, it could not possibly produce a chemical intermediate. This provided evidence that a proton gradient alone was sufficient to power ATP synthesis, removing the need of a chemical intermediate and leading to the acceptance of Mitchell's chemiosmotic theory.

3.2 Interpretation of the controversy in terms of the EC model

The philosophical challenge of the ox-phos controversy is to understand not only why it was resolved, but also why it took so long to reach a resolution.

In a key paper, Slater (1971) reviewed the status of the two competing models and used the table that I reproduce in Tab. 1. This is important because it uncovers the kind of reasoning that the scientists do when they discuss alternatives. It is a list of *issues* faced by each of the two alternative models. None of these issues determines the exclusion of any of these models. But each issue requires some ad-hoc correction to the model. After those corrections, both models fully agree with the empirical evidence. So what else do we need to compare? The list in Tab. 1 is relevant because it compares what we might call the *epistemic cost* of adopting either model. In the following, I will build multiple (simplified) versions of that table, to show how it evolved along time, from favouring the Chemical Hypothesis into fully supporting of the Chemiosmotic Hypothesis.

	Chemical Hypothesis		**Chemiosmotic Hypothesis**
1.	There is no evidence for the existence of the hypothetical A∼C compounds in state-4 mitochondria	1.	There is no evidence for the existence of a membrane potential of sufficient magnitude in state-4 mitochondria
2.	A high-energy compound with a $\Delta G'_0$ value of hydrolysis of 17 kcal/mole is unlikely	2.	A membrane potential of 370 mV is unprecedented in either artificial or natural membranes
3.	No explanation is given for the multiplicity of electron carriers in the respiratory chain	3.	There is no experimental support for alternate hydrogen and electron transfer in the respiratory chain
4.	An *ad hoc* hypothesis (the proton pump) is necessary to explain energy-linked cation uptake	4.	There is no experimental evidence for the translocation of H^+ in the absence of cation
5.	This hypothesis takes insufficiently into account the fact that the energy-transducing reactions take place in membranes	5.	This hypothesis takes insufficiently into account recent advances in our knowledge of the chemical properties of haemoproteins
6.	No explanation is given for the fact that uncouplers increase the electrical conductivity of artificial membranes	6.	No explanation is given for the fact that some uncouplers are not proton conductors
7.	An oligomycin- and uncoupler-sensitive ATP-P_i exchange reaction is found in pro-mitochondria lacking a respiratory chain	7.	There is no experimental support for the postulated diffusible X^- and IO^-
8.	There is no site specificity for reaction with ADP or for the action of uncouplers or inhibitors of oxidative phosphorylation	8.	No explanation is given for kinetics of ADD-induced oxidation of ubiquinone

Table 1 Redrawn from Slater (1971).

When Mitchell introduced his model in 1961, it had limited advantages: it is true that he did not need to assume any intermediate X with its properties, but he needed to assume an entirely new mechanism. A

very attractive aspect of Mitchell's model was that it might be potentially derived from more fundamental laws. However, molecular dynamic simulations are only now reaching the level of efficiency and accuracy necessary to describe significant portions of the mitochondrion. This wasn't an option in the 60s. So, Mitchell's model still required a significant amount of new assumptions about how the proton gradient is formed, its intensity and its effect on the mitochondrion.

The situation in 1961, when Mitchell's model was introduced, can be summarized in Tab. 2. Slater's model required the assumption of an unobserved intermediary molecule X possessing the necessary properties. As it was modeled on the existing example of the glycolysis, it did not require very complex specifications (although it did require ad-hoc assumptions to justify why it was not observed in the experiments that started to be performed. See Tab. 3). Mitchell's model, on the other hand, required a full model for the proton pump acting in the mitochondrion. On the basis of the EC that could be estimated in 1961, there was little reason to prefer Mitchell's model, which explains why it initially attracted limited enthusiasm[2].

Chemical Hypothesis	Chemiosmotic Hypothesis	complexity
Core assumptions (∼1961)		
ADP + P∼X → ATP + X reaction via intermediate X (1953).		S, one chemical intermediate
	Proton pump model for the mitochondria (1961)	M, a full model

Table 2 Status of the different assumptions of the two competing models as they were introduced. A crude, T-shirt size, estimate of the complexity of the corresponding assumption is added in the third column.

To resolve a controversy, the scientists design new experiments with the goal of discriminating between the different alternatives. Although the scientists often talk about "ruling-out" a model, we know that it is always possible to identify suitable ad-hoc assumptions to save any model from any evidence against it[3]. However, if we recognize that any assumption has an epistemic cost, a successful experiment might be able to impose much higher costs to one model and little (or no) cost to the other. This is how we should interpret the experiments that were designed to discriminate between the Chemical and the Chemiosmotic Hypotheses.

The experimental tests (1,3,4) performed by Mitchell's team produced outcomes that were expected under the Chemiosmotic Hypothesis, while they forced the inclusion of new ad-hoc assumptions to Slater's model[4]. That made Slater's model slightly more complex, but not significantly so.

More remarkable was the outcome of Jagendorf's experiment (Jagendorf and Uribe, 1966). It highlighted that Mitchell's model can be used to explain the production of ATP in the chloroplast. This expanded the scope of Mitchell's model and forced ad-hoc assumptions on Slater's model to keep the same scope (which is necessary to keep the same accuracy. See Footnote 1).

Furthermore, a growing amount of attempts to identify and isolate the intermediate X were failing. Each well-conducted failed attempt requires further ad-hoc assumptions about the properties of X to justify why, in each of those cases, X was not seen. This added significant epistemic cost to the Chemical

[2] One may argue that these estimates of EC are very crude, as I do not even write the corresponding assumptions in a fully formal way. As we will see, a more detailed analysis would not change the conclusions after the outcome of the experiments performed in the 70s. The question is whether a better estimate of the EC would have brought the controversy to an earlier conclusion. This is unlikely, because any new experiment would have forced a new formulation anyway, and it is arguably more efficient for the scientists to invest time and effort in new experiments rather than in the tedious endeavour of a full formalization. The formalization work can, however, be extremely important in cases where some assumptions may turn out to be unnecessary, contradicting or incomplete. But these efforts are dedicated to very fundamental assumptions that are not expected to change frequently.

[3] In the extreme case, the experimentalists might have hallucinations.

[4] As observed by Weber (2004), the outcome of Test 2 required further assumptions in both Mitchell's and Slater's model. So, it did not contribute to discriminating among them.

Hypothesis, but different estimates are still understandable, which may explain why those experiments were still not sufficient to close the controversy. The status in the late 60s is summarised in Tab. 3. Why wasn't that evidence deemed enough to settle for the Chemiosmotic Hypothesis? The rational conclusion in such situations is not to hastily declare a winner, but to focus on new experiments with the greatest chances of return of investment: if we believe that the Chemiosmotic Hypothesis is valid, exploring its consequences in full would likely force major corrections to the Chemical Hypothesis. Instead, further empty searches for X would lead only to small corrections to the Chemiosmotic Hypothesis.

Chemical Hypothesis	Chemiosmotic Hypothesis	complexity
Core assumptions (∼1961)		
ADP + P∼X → ATP + X reaction via intermediate X (1953).		S, one chemical intermediate
	Proton pump model for the mitochondria (1961)	M, a full model
Auxiliary assumptions (∼1965)		
a different model for the chloroplast (after Jagendorf, 1965)		M
	adapting the proton pump model would be also necessary	S
(1) Proton movements is a side effect of oxphos (1965).		S
(2) Uncouplers cause a proton current that destroys the chemical intermediate.	(2) Uncouplers increase the proton permeability of mitochondrial membranes.	applies to both
(3) Respiratory enzymes need membranes for mechanical support or structural integrity.		S
(4) The membrane is asymmetric.		S
X does not appear in ...		multiple, growing

Table 3 Status of the different assumptions of the two competing models by the end of the 60s.

This is indeed what happened. In the 70s, Racker and collaborators reconstructed Mitchell's model by combining its essential elements in a purified way. Reconstruction experiments are a risky bet, because they could fail for many reasons without forcing any substantial correction on any of the competing models. But the accumulated evidence in favour of the Chemiosmotic suggested that it was worth taking the risk. Indeed, the measurements reproduced the outcomes expected by Mitchell.

Importantly, this does *not* imply that Mitchell's model was *proved* and the corresponding model assumptions could be derived from more fundamental laws[5]. But, Racker reconstruction experiments provided a whole new wealth of evidence that is described by Mitchell's model without change, and it was not accounted at all by Slater's model. At that point, to enable Slater's model to cover the same scope as Mitchell's model, further small ad-hoc assumptions would not be sufficient: **the simplest way to correct Slater's model in order to cover the same empirical evidence covered by Mitchell's model consists in just including Mitchell's model as part of Slater's one**. Slater could still assume that Mitchell's mechanism has a negligible effect on the real mitochondria, but he would need to assume yet a new mechanism that suppresses the effect of Mitchell's mechanism, just to leave room to his yet to be found X. It is clearly simpler to drop the X.

---

[5] Sufficiently precise molecular dynamic simulations were still a long way to come. Such task is still very challenging today.

The pattern described above is a very common one in theory change. At some point, the ad-hoc assumptions that would be necessary to save one model become so many and/or so complex that the corresponding model is not considered an option anymore. In those cases, some scientists may say that the model was "ruled out" by a **crucial experiment**, often disregarding the well-known criticisms to the notion of crucial experiment (Duhem, 1954; Ariew, 2026). However, we have seen that Racker's experiments added the last bit of evidence that made Slater's model unambiguously more complex than Mitchell's[6]. This justifies, in this case, the intuitive idea of crucial experiment, but only under the further assumption of the EC model, that puts a limit on the admissible amount of ad-hoc assumptions.

The EC model also clarifies some old open questions about **underdetermination**. The very concrete way in which the scientists can build an arbitrary number of different—empirically accurate—models is by conjecturing the presence of procedural mistakes in every single experiment. To some extent, they actually do that each time that they find an experimental outcome surprising. But they don't do it *too much*. Philosophers have struggled to find a clear rule behind this behaviour and often resort to say that "the scientists know" or that "underdetermination is a mere theoretical possibility, but, apparently, not a serious problem". This is not satisfactory. The EC model offers a compelling explanation: *ad-hoc conjectures are possible, but they have a cost*[7].

Chemical Hypothesis	Chemiosmotic Hypothesis	complexity
Core assumptions (70's)		
ADP + P∼X → ATP + X reaction via intermediate X (1953).		S, one chemical intermediate
Proton pump model for the mitochondria (70's).	Proton pump model for the mitochondria (70's).	M, a full model
Why H pump is suppressed, and X		S
Auxiliary assumptions (70's)		
a different model for the chloroplast (after Jagendorf, 1965)		M
(1) Proton movements is a side effect of oxphos.		S
(2) Uncouplers cause a proton current that destroys the chemical intermediate.	(2) Uncouplers increase the proton permeability of mitochondrial membranes.	applies to both
(3) Respiratory enzymes need membranes for mechanical support or structural integrity.		S
(4) The membrane is asymmetric.		S
X does not appear in ...		multiple, growing

Table 4 Status of the different assumptions of the two competing models by the end of the 70s.

---

[6] In other cases, the last bit of information that determines a model selection is provided by a theoretical results, rather than an experiment.

[7] In fact, to decide whether to put the blame on the theory or on the experiment, the scientists typically design new experiments that can discriminate between the two options. The goal is to ensure that all important experimental results can be traced back to general rules and that erroneous experiments count as exceptions for all competing models.

3.3 Reassessing alternative accounts

In this section I review alternative accounts that have been proposed to explain the ox-phos controversy and I show that the EC model corrects their shortcomings.

Naive Parsimony. It has been argued that Mitchell's theory was more parsimonious but that such argument was explicitly dismissed as unimportant by the biochemists of the time. I have argued, that Mitchell's theory was, initially, *not* more parsimonious, because it could not be derived from more fundamental principles. Parsimony (or rather EC[8]) was eventually the reason to prefer Mitchell's model, but only when the difference became so large that the scientists were not even aware that they were using it implicitly for their conclusions. Unfortunately, it is very common to refer to parsimony when it is not conclusive and forget about it when it is.

Bayesianism. As explained in (Weber, 2004), Bayesianism does not explain the development of the controversy. In fact, the outcome of theory selection depends critically on the choice of the priors. The EC model offers a measure of complexity that can be used to set the Bayesian priors. But that simply means that the Bayesian approach cannot be used, without further assumptions (such as the EC model), to draw conclusions on which models are legitimate options. If we accept the EC model, though, the Bayes formula can be used to obtain averages over alternative, legitimate, models.

Social constructivism. A social constructivist account of the ox-phos controversy has been attempted by Gilbert and Mulkay (1984), who reconstructed the episode in terms of rhetorical strategies used by some of the scientists involved. However, they provide no answers to the question of why a consensus was eventually formed and why it was Mitchell's theory (rather than that of his more powerful opponent Slater) that was accepted (Weber, 2004). Today, the majority of philosophers reject a purely social constructivist account, but it remains the default option in the absence of more widely accepted alternatives.

Incommensurability. The topic of incommensurability plays an important role in the philosophical debate around the ox-phos controversy (Allchin, 1994, 2015; Weber, 2002). To be clear, there is no way to prove wrong someone who claims that her theory predicts anything and claims that she can't see any evidence to the contrary and does not even reveal what she is measuring in a way that others can also measure. But there is no need to prove her wrong: what she says is simply not relevant, because she talks about quantities to which we can't relate.

If two theories deal with the same *topic*, we should be able to identify a common basis of measurable quantities. The idea of *same quantities* which are part of *different theories* must be clarified (Scorzato, 2013): strictly speaking, any concept A within theory T (say $A(T)$) has a different meaning from $A(T')$ (no two concepts of different theories can have identical meanings[9]) but $A(T)$ and $A(T')$ can be identified as long as we agree on a one-to-one correspondence between the outcomes of $A(T)$ and the outcomes of $A(T')$. Such correspondence does require a common—although minimal—theoretical background, that covers at least a sufficient modeling of the experimental devices. We can't prove that such common background is always possible. But, in practice, I know no examples where this wasn't possible[10]. Even when a theory T' does not cover the same scope as T, we can extend them to make them equally accurate. In fact, Weber (2002) observes that the 1st, 2nd and 3rd traits of incommensurability, which were definitely present in the ox-phos debate, did not imply incomparability. What Weber calls the *causal sufficiency* of

---

[8] In fact, EC is the only *precise* definition of parsimony I am aware.

[9] This is what Hoyningen-Huene (2002) and Weber (2002) refer to as the 1st, 2nd and 3rd elements of incommensurability, which are natural aspects of competing theories and do not prevent their comparison as defined, e.g., in the EC model.

[10] If the measurement itself is contested, it is usually replaced by shared alternative measurements (Fletcher, 2024).

Mitchell's mechanism fully aligns with the EC model's selection criteria: the Chemiosmotic Hypothesis was sufficient to describe the data with little additional assumption.

To save incommensurability, Allchin (1994) goes so far as to state that the controversy had no winner: the two original models differentiated in multiple models, each with its own scope. Allchin should explain who ever used Slater's model after 1974. It is certainly possible to split models in a way that fits this narrative, but I don't see any need for doing it, except for accommodating a specific philosophical stance. What is important is that it is also possible to identify one single model (namely, Mitchell's model, with the few additional ad-hoc hypothesis recognised in the 60s and 70s) that is consistent and describes all the evidence under discussion in a more concise way than any alternative.

In a later paper, Allchin (2015) claims that the participants to the debate could not even agree on what constituted an anomaly. This doesn't explain why they kept talking to each other for so long and even less why they eventually agreed. The anomalies reported in (Allchin, 2015) represent experimental observations that can be described by both models, although, in some cases, with additional assumptions. The debate was—more or less explicitly—about the acceptability of those assumptions.

Mayo's Severe Tests and Weber's Control Experiments. The proposal of *severe tests* (Mayo, 1997; Weber, 2004) effectively tries to resurrect the idea of crucial experiments, but it cannot exclude multiple scenario in which the test passes for reasons that are different from the intended ones. It is easy to argue that Racker's tests were severe, but it is also easy to find severe tests that lead to wrong conclusions (Worrall, 2009).

To overcome some of the shortcomings of Mayo's severe tests, Weber (2004) emphasised the role of *Control Experiments*. Indeed, the experiments described in the previous section, and in particular by the group of Racker, aim at excluding potential alternative model to describe the phenomena. However, as Weber correctly notes: *Only once **all** the possible ways in which the interpretation of experimental data could go wrong have been checked with suitable control experiments will an experimental biologist announce that he or she has, for instance, "demonstrated the occurrence of oxidative phosphorylation in a reconstituted system"*. The key word is **all**. Underdetermination tells us that there are infinite ways in which the interpretation of the experimental data could go wrong. These alternative are also very concrete and easy to construct: it is sufficient to build bizarre scenario and question each and every experiment that conflicts with them[11]. The role of philosophy is to understand why some of these alternative are not an option. Philosophy cannot simply rely on *some scientists* saying so.

Epistemic Norms. My previous remarks are fully consistent with Weber (2004), when he writes that the "experimental system account" alone is not sufficient to explain why Mitchell theory was eventually accepted. Epistemic norm(s) are certainly necessary, the question is which one(s). Weber suggests that *in-vitro reconstruction experiments* assumed that role. But this opens more questions that it solves. First, this norm is very specific to a few subfields of biology. Then, one should wonder why have scientists from different disciplines accepted Mitchell theory? Secondly, why did in-vitro experiments gain such important status? It is true that they became possible only at that time, but many other techniques were enabled at the same time and afterwards. Why is Mitchell theory still the accepted explanation today, after 50 years of amazing technological progresses? Why was it not replaced by some other account?

The EC model answers all these questions, because it explains the effectiveness of in-vitro experiments on the basis of a single general rule. As noted in (Weber, 2004; Scholl and Nickelsen, 2015) in-vitro experiments allow the testing of multiple alternative scenarios where different ingredients are selectively included or different actions are performed. This provides a wealth of empirical data that are

---

[11] This is exactly the strategy taken by pseudo-science. So, arguing that the scientists do not take all those alternative seriously relies on some hypothetical clear distinction between scientists and pseudo-scientists, that, unfortunately, we don't have.

designed to be difficult to fit for either one or the other (or both) the competing models. If this strategy is successful, one of the competing models will be forced to adopt a very large amount of ad-hoc assumptions to remain accurate.

Mechanisms and the context of discovery. The analysis of Scholl and Nickelsen (2015) focuses on the context of discovery. But it is very much relevant for the context of justification as well[12]. In fact, they state: "What we expect of our strategies is not truth-conduciveness, but only that they can efficiently populate the space of possible causal and mechanistic hypotheses for the explanation of a phenomenon". But *efficiency* in which sense? If it refers to truth finding, then it does refer to a truth-conducive strategy. If not, which other criterion determines the effectiveness? But, identifying that criterion is exactly the challenge that Goodman (1983) calls the *New Riddle* (Scorzato, 2025).

Concerning the "discovery strategies" discussed in (Scholl and Nickelsen, 2015), which are an evolution of (Machamer et al., 2000), I fully agree with the analysis of Theurer and Bickle (2013), which argues that mechanisms are essentially a rediscovery of Kemeny-Oppenheim (KO) reduction[13]. I refer to (Theurer and Bickle, 2013) for the argument that I find very convincing. Here I just note that the main issue with KO reduction is the vagueness of the concept of *systematization*. Indeed, it suffers the same catastrophic language dependence that affects all non-empirical epistemic values, following (Scorzato, 2025)[14]. On the other hand, the EC model solves precisely this problem, because it requires the new $\mathcal{M}'_B$ (after reduction) to be more concise than the union of $\mathcal{M}_R$ and the old $\mathcal{M}_B$ (before reduction).

Note that the EC model also acknowledges the value of *partial* reductions. In fact, we can define *complete reductions* those cases where $\mathcal{M}_B$ is identical to $\mathcal{M}'_B$. In other words, we discover that $\mathcal{M}_R$ can be dropped completely, because the old $\mathcal{M}_B$ is just sufficient to describe everything described by $\mathcal{M}_R$. But, this is not very often the case: quite often $\mathcal{M}_B$ needs to be adjusted (hence it becomes $\mathcal{M}'_B$) with some slight extension and/or ad-hoc hypotheses. What is important, to claim any progress is that $\mathcal{M}'_B$ remains strictly more concise than the union of $\mathcal{M}_B$ and $\mathcal{M}_R$. Otherwise, any juxtaposition of non-contradicting old models would count as a reduction, which is not plausible.

In a very similar manner, the EC model also recognizes the value of *regularities*, which may not be derivable from a more fundamental $\mathcal{M}_B$, but represent a small correction to $\mathcal{M}_B$, which is still much more concise than the union of $\mathcal{M}_B$ and $\mathcal{M}_R$. This is fully consistent with the thesis of (Weber, 2008).

4 Conclusions

The oxidative phosphorylation controversy serves as a compelling case study for the Epistemic Complexity (EC) model. By re-evaluating this historical episode, it becomes clear that theory change is driven neither by pure empirical evidence, nor by empirical evidence combined with shifting, ill-defined values. Instead, it is well described by empirical evidence combined with just one well-defined and general value: conciseness of the assumptions.

The longevity of the ox-phos debate was not a sign of scientific irrationality, but a reflection of the uncertainty over the relative EC of the two competing models—the Chemical and Chemiosmotic Hypotheses, after including all the necessary additional assumptions to ensure their accuracy. The controversy reached its resolution only when the *epistemic cost* of maintaining the Chemical Hypothesis became unambiguously greater than dropping it. Racker's reconstruction experiments provided the decisive blow, not by *proving* Mitchell's theory in a naive sense, but by forcing the Chemical Hypothesis to adopt such

---

[12] One could argue that trying to understand the context of discovery without a robust model for the context of justification is like trying to understand the dynamics of turbulence without robust model for equilibrium thermodynamics.

[13] According to KO, a theory $\mathcal{M}_R$ is reduced to a reducing theory $\mathcal{M}_B$ when all phenomena described by $\mathcal{M}_R$ are also described by $\mathcal{M}_B$, and $\mathcal{M}_B$ is at least as well *systematized* as $\mathcal{M}_R$.

[14] The same problem affects Mechanisms of Machamer et al. (2000), insofar they must clarify in which sense they are *effective* strategies. Otherwise, any mechanism would go.

a large amount of ad-hoc assumptions that made it unambiguously more complex than its rival, for no empirical advantage.

Ultimately, the resolution of the ox-phos controversy demonstrates that scientific debates are settled through a rational, albeit often slow, evaluation of well-defined properties. The EC model provides the necessary framework to understand this dynamics, offering a robust account of how science moves toward greater simplicity and accuracy.

I do not argue that all scientific controversies are explainable in terms of the EC model. Also scientists happen to hold wrong prejudices, which negatively influence some debates. But, it is exactly the crucial role of philosophy of science to identify which controversies are motivated by general scientific standards and which are not.

References

Allchin, D. (1994). The super bowl and the ox-phos controversy: 'winner-take-all' competition in philosophy of science. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pp. 22–33. Cambridge University Press.

Allchin, D. (1997). A twentieth-century phlogiston: Constructing error and differentiating domains. *Perspectives on Science 5*(1), 81–127.

Allchin, D. (2015). Context-Dependent Anomalies and Strategies for Resolving Disagreement: A Case in Empirical Philosophy of Science. In *Empirical Philosophy of Science: Introducing Qualitative Methods into Philosophy of Science*, pp. 161–171. Springer.

Ariew, R. (2026). Pierre Duhem. In E. N. Zalta and U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring 2026 ed.). Metaphysics Research Lab, Stanford University.

Douglas, H. (2013). The value of cognitive values. *Philosophy of science 80*(5), 796–806.

Duhem, P. M. M. (1954). *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press.

Fletcher, S. C. (2024). On the Alleged Incommensurability of Newtonian and Relativistic Mass. *Erkenntnis 90*, 3567–3588.

Gilbert, G. N. and M. Mulkay (1984). *Opening Pandora's box: A sociological analysis of scientists' discourse*. CUP Archive.

Goodman, N. (1983). *Fact, Fiction, and Forecast* (4th ed.). Cambridge, MA: Harvard University Press.

Greville, G. D. (1969). A scrutiny of Mitchell's chemiosmotic hypothesis of respiratory chain and photosynthetic phosphorylation. In *Current topics in bioenergetics*, Volume 3, pp. 1–78. Elsevier.

Hoyningen-Huene, P. (2002). Paul feyerabend und thomas kuhn. *Journal for general philosophy of science 33*(1), 61–83.

Jagendorf, A. T. and E. Uribe (1966). ATP formation caused by acid-base transition of spinach chloroplasts. *Proceedings of the National Academy of Sciences 55*(1), 170–177.

Keas, M. N. (2018). Systematizing the theoretical virtues. *Synthese 195*(6), 2761–2793.

Kuhn, T. S. (1962). *Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Machamer, P., L. Darden, and C. F. Craver (2000). Thinking about mechanisms. *Philosophy of science 67*(1), 1–25.

Mayo, D. G. (1997). Error statistics and learning from error: making a virtue of necessity. *Philosophy of Science 64*(S4), S195–S212.

Mitchell, P. (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature 191*(4784), 144–148.

Mitchell, P. and J. Moyle (1965a). Evidence discriminating between the chemical and the chemiosmotic mechanisms of electron transport phosphorylation. *Nature 208*(5016), 1205–1206.

Mitchell, P. and J. Moyle (1965b). Stoichiometry of proton translocation through the respiratory chain and adenosine triphosphatase systems of rat liver mitochondria. *Nature 208*(5006), 147–151.

Mitchell, P. and J. Moyle (1967). Proton-transport phosphorylation: some experimental tests.

Racker, E. (2012). *A new look at mechanisms in bioenergetics*. Elsevier.

Racker, E. and A. Kandrach (1971). Reconstitution of the third site of oxidative phosphorylation. *Journal of biological chemistry 246*(22), 7069–7071.

Racker, E. and W. Stoeckenius (1974). Reconstitution of purple membrane vesicles catalyzing light-driven proton uptake and adenosine triphosphate formation. *Journal of Biological Chemistry 249*(2), 662–663.

Scholl, R. and K. Nickelsen (2015). Discovery of causal mechanisms: Oxidative phosphorylation and the Calvin–Benson cycle. *History and philosophy of the life sciences 37*(2), 180–209.

Scorzato, L. (2013). On the role of simplicity in science. *Synthese 190*, 2867–2895.

Scorzato, L. (2015a). A simple model of scientific progress – with examples. preprint: philsci-archive.pitt.edu/12024.

Scorzato, L. (2015b). Science and Illusions. preprint: philsci-archive.pitt.edu/15570.

Scorzato, L. (2016). A simple model of scientific progress. In L. Felline, F. Paoli, and E. Rossanese (Eds.), *New Developments in Logic and Philosophy of Science*, Volume 3 of *SILFS*. College Publications.

Scorzato, L. (2024). Reliability and Interpretability in Science and Deep Learning. *Minds and Machines 34*(3), 27.

Scorzato, L. (2025). A not-too-simple solution to Goodman's new riddle of induction in the age of AI. preprint: philsci-archive.pitt.edu/26915.

Slater, E. (1967). An evaluation of the Mitchell hypothesis of chemiosmotic coupling in oxidative and photosynthetic phosphorylation. *European Journal of Biochemistry 1*(3), 317–326.

Slater, E. C. (1953). Mechanism of phosphorylation in the respiratory chain. *Nature 172*(4387), 975–978.

Slater, E. C. (1971). The coupling between energy-yielding and energy-utilizing reactions in mitochondria. *Quarterly Reviews of Biophysics 4*(1), 35–71.

Theurer, K. and J. Bickle (2013). What's Old Is New Again: Kemeny-Oppenheim Reduction at Work in Current Molecular Neuroscience. *Philosophia Scientiae 172*(2), 89–113.

Weber, M. (2002). Incommensurability and theory comparison in experimental biology. *Biology and Philosophy 17*(2), 155–169.

Weber, M. (2004). *Philosophy of experimental biology*. Cambridge university press.

Weber, M. (2008). Causes without mechanisms: Experimental regularities, physical laws, and neuroscientific explanation. *Philosophy of Science 75*(5), 995–1007.

Worrall, J. (2009). Error, Tests and Theory Confirmation. In D. G. Mayo and A. Spanos (Eds.), *Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science*, pp. 125–154. Cambridge University Press.