The Super Bowl and the Ox-Phos Controversy: "Winner-take-all" Competition in Philosophy of Science

Douglas Allchin

Cornell University

1. Introduction

Imagine two theories in a scientific controversy cast as competing teams in the Super Bowl, and you may get a "scoreboard of experimental evidence" such as the following, published in a review article in 1970 (after Racker 1970, 135):

SCOREBOARD OF EXPERIMENTAL EVIDENCE FOR THE CHEMICAL AND CHEMIOSMOTIC HYPOTHESIS OF ENERGY GENERATION DURING OXIDATIVE PHOSPHORYLATION AND PHOTOPHOSPHORYLATION

	Chemical	Chemiosmotic
Role of the membrane	−	+
Ion transport	+	×
Action of uncoupling agents	−	+
Isolation of high-energy intermediates	−	±
^{32}Pi-ATP exchange ADP-ATP exchange $H_2{}^{18}O$ exchanges	+	−

This figure compares two hypotheses in a debate in bioenergetics in the 1960s and 70s known as the Ox-Phos Controversy (Rowen 1986; Allchin 1990; Weber 1991). But its format, suggested by its title, bears a striking resemblance to the half-time recaps in televised football games: parallel assessments in several categories ask us to compare, say, how many yards rushed, number of first downs, passes completed, evidence for the role of the membrane, or evidence for ion transport, etc. (see also Sindermann 1982). Why did the review author—Efraim Racker, a research biochemist—borrow the scoreboard framework from sports to convey his assessment in science? Should one—can one—evaluate the performance of each scientific "team," infer a probable winner and loser from the plus-minus ratings in each column, and decide which hypothesis we should bet on or, given the final "score," which we should rationally support? Indeed, current biology textbooks would lead one to believe that the Chemiosmotic Hypothesis triumphed when its originator, Peter Mitchell, "won" the Nobel Prize in Chemistry in 1978. For those who view theory-choice in science as a matter of theory competition,

PSA 1994, Volume 1, pp. 22-33

the scoreboard may be a quite natural expression for assessing alternative hypotheses, construed (like athletic adversaries) as "rivals." Here, I explore Racker's figure, along with the Super Bowl metaphor as a model of competition, to consider more fully the nature of competition in philosophical conceptions of science.

In what follows, I analyze Racker's figure (§2) and contrast it, first, to three other comparative diagrams and tables published around the same time (§3), and then to a later "Revised Scoreboard" (§4). The most salient feature of Racker's scoreboard is, perhaps, that it frames the debate in polar, either-or, winner-take-all terms, though the outcome of the controversy suggests a pattern of differentiation or partitioning of domains among the hypotheses (§2). Incompatible or incommensurable theories, one finds, may not necessarily be mutually exclusive. There are two frameworks for interpreting competition, each applicable in separate contexts (§3). These observations further suggest strategies for scientists—to analyze their discourse in cases of disagreement and shape further research, and to bridge contexts of discovery and justification (§4).

2. A Scoreboard of Experimental Evidence?

Though Racker was a biochemist, not a philosopher, his "scoreboard of experimental evidence" above epitomized a notion of competition that has been fundamental to philosophy of science for at least the past three decades. As we came to understand the significance of alternative theories (logically, historically), we focused on ways to discriminate between them—and justification became linked to theory choice. Approaches to theory choice, however, have consistently drawn on competition as an underlying theme. Kuhn, for example, referred repeatedly to conflicting paradigms as "competitors" (1962, 147-50, 154-55) and he has suggested how the process of science may fit in a Darwinian framework (pp. 171-73; 1990 PSA Presidential address; 1992). Laudan, likewise, consistently portrays theories as rivals, scored on a scale of progressiveness (1977; 1992). For Laudan (1977), rationality itself emerges from comparing competing theories. Bayesian approaches inscribe competition in quantitative comparisons and sustain the notion of a crucial experiment as decisive between competing theories (Howson and Urbach 1989, 91-92). Even those who reorient their focus away from theory and more towards experiment (e.g., Galison 1987 and Franklin 1986) often build their accounts on episodes of competition. Finally, Hull's more social, evolutionary model adopts an explicit Darwinian metaphor with a vengeance: science is propelled by curiosity and credit and is regulated by "the visible hand" of competition (1988, Chap. 10; note also the dust jacket image alluding to "science red in tooth and claw"). For Hull, competition is a critical feature of the social structure of justification when we view "science as a process." Competition is a widespread—and often explicit—metaphor in philosophy of science, bridging Bayesian, historical, experimental, and even more sociological perspectives.

Racker's scoreboard allows one to notice a particular theme present, but not made explicit, in many of the philosophical models. Most saliently, perhaps, the scoreboard frames the scientific debate in polar, either-or, winner-take-all terms. I call this (somewhat archly) the "Super-Bowl model" of competition. The Super Bowl is *not* science, of course. But it does vividly exemplify certain *features* of competition also present in many philosophies of science (on exemplification, see Goodman 1976, 52-60; 1978, 63-65, 133-37). In the Super Bowl, as in Racker's assessment:

(1) competitors are assumed to be in the same category (are functionally equivalent or intersubstitutable) (principle of functional symmetry);

(2) competition is limited to two contenders (principle of bipolarity);

(3) only one can win (principle of either-or); and

(4) the winner wins exclusively and absolutely (principle of winner-take-all).

The winner-take-all principle (based on the other three), in particular, characterizes the rhetorical essence of the Super Bowl: the championship, the title, the best, being #1; and it may remind us of philosophical efforts to articulate a method by which we may select the single "most rational" theory (e.g., Laudan, Laudan and Donovan 1988; Niiniluoto 1992; Kukla 1992). A winner-take-all principle implies that only one theory is "right" or "completely right," and all others—even those we may call "half-right" or "partly right"—are ultimately "wrong" (see, e.g., Popper 1975). Again, while science is hardly the highly conventionalized practice of sports, when we view science through the Super-Bowl principles, we get Racker's scoreboard: either the Chemical Hypothesis is justified or the Chemi-Osmotic Hypothesis is justified, not parts of both, nor both partly (nor even some third alternative). The two solutions are exhaustive and mutually exclusive.

Participants in the Ox-Phos Controversy tended to interpret their disagreement in these "Super-Bowl" terms. In disagreeing about oxidative phosphorylation (or ox-phos)—how ATP is produced in the cell—they implicitly assumed only one theory could be correct. In the text which accompanied his table, for instance, Racker wryly characterized the then-raging debate: "In reading discussions of the proponents of the two hypotheses," he noted, "one gains the impression that the evidence against the formulations of the opponent is overwhelming." Indeed, the structure (and title) of the scoreboard implicitly invites us to compare the two columns to determine the all-inclusive "winner." The two vertical columns allow us to tally the various positive ('+') and negative ('–') scores for each hypothesis, and balance them with other scores (marked by '×') that represented "serious discrepanices" that still "need to be answered by decisive experiments" (pp. 132-37). Although the weight of the evidence might strike the casual observer as favoring the Chemi-Osmotic Hypothesis, Racker himself reached the opposite conclusion. He acknowledged that the novel chemiosmotic hypothesis was an "ingenious scheme" and evidence for it was "mounting"; nonetheless, he regarded some its assumptions as "formidable and controversial" (p. 132). He admitted, finally: "having been raised by the music of substrate-level phosphorylation [the basis of the more conventional approach] my own prejudices induce me to lean toward some aspects of the chemical hypothesis" (p. 137). Racker's assessment was governed by assumptions about dichotomous choice among mutually exclusive alternatives.

Ironically, perhaps, Racker's scoreboard also introduced information in a way that suggests another approach to competition. That is, when one views the table in terms of its horizontal rows, rather than its vertical columns, one finds the evidence divided or partitioned into separate categories. Racker identified specific sets of experiments and their corresponding phenomena: domains or perhaps sub-domains (*sensu* Shapere 1984), or data-domains (Ackerman 1986). When one sorts the observations in this way—by domain—the distribution of evidence becomes more clearly articulated. Both hypotheses claimed to describe the critical intermediate energy state in the transfer of energy to ATP and thus their domains overlapped significantly. However, the evidence for their related claims was not uniformly distributed.

The sorting of evidence is even more striking when one notes that the negative ('–') scores did not document direct counter-evidence or anomalous mismatches between predictions and observations. Rather, they indicated results that were "difficult to explain" or "more distressing to" each hypothesis. That is, there was lack of evidence or, more properly perhaps, lack of a theoretical concept through which one could even situate or address the evidence (see also Laudan, 1977, on non-refuting anomalies). The Chemical Hypothesis, for example, did not concern itself with the membrane, though data seemed to indicate that the presence of an intact, closed membrane was essential to the phosphorylation process. The Chemi-Osmotic Hypothesis, on the other hand, was ill-equipped conceptually to explain how certain atoms were transferred during the reactions (the 'exchange reactions' in the table). Plus and minus ratings were thus awkward parallels, not representing opposite evaluations of "right" and "wrong"—or even "right" versus "more right." Instead, domains were deemed more or less relevant, and thus reflected more or less favorably on each hypothesis. In an either-or, win-lose approach, of course, one disregards precisely this cross-characterization of the evidence. However, one can sort evidence or observations, not just theories. The scoreboard (as its label suggests) characterized the status of the "experimental evidence" more than of the two hypotheses.

The alternative claims about ox-phos, presumed to be incompatible, were thus found to be compatible by articulating new sub-domains and understanding how they related to each other. In a sense, differentiation dissolved the competition. But to suggest that the theories never competed at all would betray the history, here. The competition resulted precisely from differing perceptions about how one could generalize each hypothesis across the various domains (another variant of Goodman's, 1963, problem of defining induction classes; also the problem of "rightness of categorization," 1978, 127 and Chap. 7; 1976, 169-73).

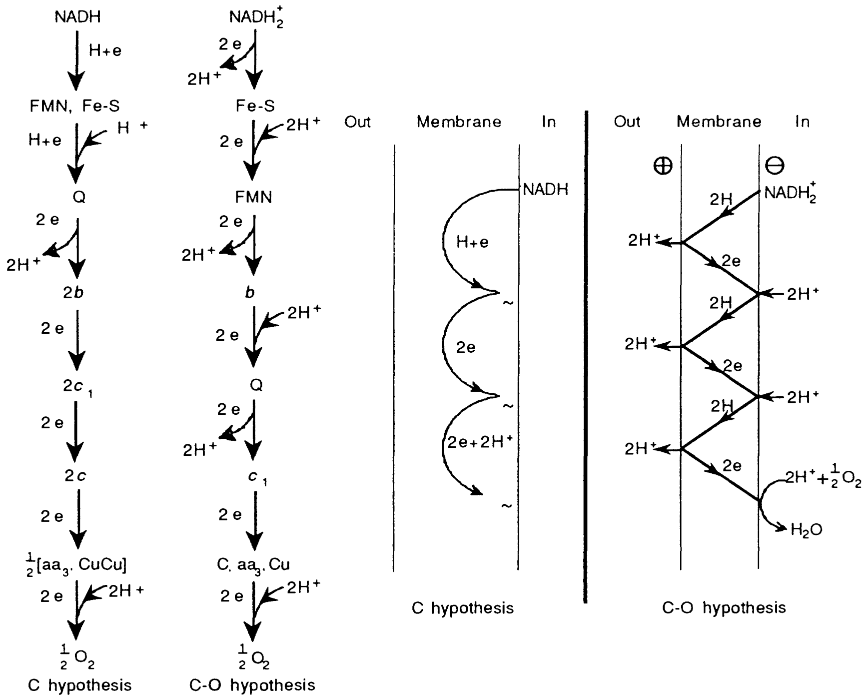
An alternative to winner-take-all competition, then, is differentiation. That is, one may differentiate the "competing" theories by sorting or partitioning the domains appropriate to each. Indeed, in retrospect, we can say that Racker foreshadowed—though surely without prescience—how the conflict or competition between the two hypotheses would eventually be resolved. When the Ox-Phos Controversy finally subsided, both hypotheses remained, though they explained different, intersecting (or adjacent) domains, as suggested in Racker's "scoreboard of experimental evidence."[1]

3. Incommensurability and the "Winner-Take-All" Principle

Racker's analysis suggests that philosophers must qualify or revise substantially many models of science and include differentiation as a possible outcome of theory competition or conceptual disagreement. But it also reminds us of the context in which scientists themselves must make these assessments. How would one know in the midst of this debate whether a winner-take-all framework was appropriate or not? The challenge introduced by Racker's analysis is to articulate the different contexts in which each framework of competition may properly apply. That is, in what particular types of occasions does each model of competition function?

One must examine the resources available in the context of the controversy. As noted earlier, participants in the ox-phos episode tended to cast the debate themselves according to implicit "Super Bowl" principles. Nowhere is this orientation more evident than in another, extraordinary review of the same two hypotheses by another prominent biochemist, E.C. Slater, in the year following Racker's (Slater 1971). In a dominant and explicitly parallel structure, each concept for one hypothesis is presented and compared against a concept for the other hypothesis. Every diagram depicting relationships for one hypothesis is carefully paired with a corresponding diagram for the rival hypothesis. The survey is systematic and thorough. Yet in terms of the outcome of the controversy, Slater's analysis was less effectively framed than Racker's: why? Here, the ineffective strategy is far more telling philosophically than the successful one. When viewed in more detail, Slater's review reveals how incommensurability itself may indicate where differentiation may be appropriate.

Consider, for example, a pair of diagrams representing alternative versions of the "Sequence of components of [the] respiratory chain" (Figure 2a). The respiratory chain is a series of proteins and other molecules where electrons (labeled "2e") cascade down energy levels to oxygen (this is precisely where the oxygen we breathe is ultimately used). Even without knowing any biochemistry, one can easily recognize the FMN or Q or c_1 in both columns and note the clearly differing sequences and varying positions of the supplementary arrows into and out of the chain. For someone familiar with the two hypotheses, however, the similarity in the diagrams disguises a fair amount of conceptual shoehorning. Proponents of the chemical (or "C") hypothesis (depicted on the left) viewed the energy conversion as a stepwise release of energy from molecule to molecule. For them, the sequence was a familiar image of energy flow, detailing each intermediate step. For the chemiosmotic (labeled "C-O") hypothesis, however, the function of each component was coupled to its position in the mitochondrial membrane, essential for understanding how an energized gradient across the membrane could be generated. The order of the components was largely incidental to how the electrons moved through space. Information about sequence alone was inadequate and perhaps peripheral or misleading. Additional domain items needed to be considered at the same time. For the chemical hypothesis, then, the sequence diagram was central to the answer about energy transfer; for the chemiosmotic hypothesis, the diagram described a state of affairs, but was far from "the" answer—and did not even



(a) (b)

Figure 2.

Comparisons of chemical (C) and chemiosmotic (C-O) hypotheses after Slater (1971, 44-45) address the significant causal questions. The sequence diagram thus represented an effort to make chemiosmotic claims conform to a conventional chemical framework—ostensibly so that the two could be compared in parallel.

Given what we know about theoretical bias, one might suspect that the (mis)interpretation was based on a singular, inflexible perspective; however, the shoehorning occurred in reverse in the very next pair of diagrams (Figure 2b). Here, the chemiosmotic hypothesis was represented in a drawing taken directly from one of its original documents (Mitchell's 1966 monograph). The diagram conveyed the fundamental chemiosmotic claim that oxidative phosphorylation was a "vectorial" process (having a spatial dimension as well as scalar magnitude). As electrons shifted to lower energy levels within the membrane (diagrammatically, downwards), hydrogen ions moved critically across the membrane ("$2H^+$," from right to left). The chemical hypothesis, on the other hand, made no essential claims about the position of ox-phos components or the physical pathway of the reactions. It assumed they were irrelevant. As a result, the diagram could depict no more than an arbitrary scalloped pathway punctuated by squiggle symbols ('~') representing the high-energy bonds of proposed intermediate molecules. Even though the membrane was where the reactions undeniably took place, there was no meaningful physical correlate to the semi-circular pathways. As in Racker's scoreboard, there was a causal category for one hypothesis with no corresponding category for the other. In this case, the claims of the chemical hypothesis were shoehorned to fit into a chemiosmotic framework. The shoehorning reflected a view that the hypotheses were funcitonally commensurable and thus could be evaluated as mutually exclusive by either-or rules.

The parallel assessment continued into a table of objections (Figure 3), another "scoreboard" of sorts. Here, the awkward matching of categories was even more striking. Slater considered eight objections to each hypothesis, some overlapping with those mentioned by Racker (C#1, C-O#4, C#6). But the objections were arranged in pairs that did not correspond directly. For example, data about the existence of high-energy intermediate compounds (C#1) differed from data about the existence of membrane gradients (C-O#1), though both related to the intermediate energy state at issue. There was no one crucial experiment that would allow one to select one alternative while simultaneously rejecting the other. Similarly, facts about how membrane conductivity was affected by uncouplers were weighed opposite facts about specific uncouplers (#6). Though both objections were about "uncoupling," one was based on membrane properties, the other on chemical properties—different categories both conceptually and experimentally. That is, the pairs were mere analogs, not functional substitutes for one another. The functionally incongruent categories inhibited commensurable comparison of the two hypotheses.

Slater's lists of objections also exhibited some of the conceptual and domain asymmetries found in Racker's scoreboard. Many objections were phrased in terms of "no evidence for" or "no experimental support for" (C#1; C-O#1,3,4,7), "no explanation is given for" (C#3,6; C-O#6,8), or "takes insufficiently into account" (C#5; C-O#5). One does not find the phrases "evidence contradicts" or "the explanation does not match available data." Each hypothesis was challenged by absence of data, not outright error. Unconfirmed, theoretically predicted results were criticized indirectly through terms such as "unlikely" (C#2) or "unprecedented" (C-O#2). As in Racker's review, evidence for or against a hypothesis was assessed in terms of whether certain domain items were effectively mapped by the concepts or demonstrated experimentally (see also Allchin 1992a). While the domains of the two hypotheses overlapped in terms of fundamental claims about energy transfer, they diverged in views of the range of relevant phenomena. Slater's assessment thus resonates strongly with

Racker's. But his table differs markedly from the scoreboard in suppressing, rather than highlighting, these distinctions so that they fit into an either-or format for winner-take-all theory choice. Slater's paired figures and his lists of objections, then, were not parallel conceptually, despite the graphic organization and commentary. The striking juxtapositions in this case—paradoxically, perhaps—simply underscore a fundamental incommensurability between chemical and chemiosmotic hypotheses.

TABLE 2. *Objections to the C and chemiosmotic hypotheses*

C 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_1 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

Figure 3. Slater's (1971) table of objections (p. 52).

Slater's analysis is remarkable because at every turn it tends to betray the presence of incommensurability while virtually refusing to acknowledge it. Slater, too, adhered to a polar, either-or, winner-take-all orientation to theory choice or theory competition. Indeed, Slater's sensitivity to the competitive framework is evident in his scrupulous fairness. The Super Bowl reminds us that the outcome of competition is legitimized in part by the "objectivity" embodied in rules of fair play. Fairness—comparing similar cases by similar standards—is part of justifying the winning team's victory. In science, similarly, each hypothesis must be given a "fair hearing" or a "fair chance" to prove itself, if a comparative assessment is to be justified. Note the role of "referees" as ajudicators in both contexts (Sindermann 1982, 3). The "'rules' governing the '*game of science*'" (Lakatos 1978, 140-43) must ensure that all evidence will be evaluated and each hypothesis weighed according to a uniform method. Slater's conceptual shoehorning can thus be viewed as an effort to fit non-corresponding concepts into parallel categories for "fair" comparison. Slater imposed symmetry where the domains were, in fact, asymmetric. His review is a tribute to fair competition between hypotheses in science framed in the Super-Bowl model. But in this case, "fairness" emerged by suppressing or avoiding the problems posed by incommensurable hypotheses.

Racker's scoreboard, by contrast, implicitly acknowledged the incommensurability by treating the hypotheses as incompatible, integrated wholes (*sensu* Kuhn 1962; see Hoyningen-Heune 1993, 220-21). The scoreboard posed possibilities about the differentiation of domains that the format of Slater's diagrams and table did not—and perhaps could not—allow. Slater's analysis, in its failure to represent the nature of the debate effectively, shows more clearly why the shift in competitive frameworks was necessary. The incommensurability—incompatible hypotheses with diverging but still overlapping domains—was itself the signal that winner-take-all assumptions were open to reassessment.

One may pause to consider this perhaps counterintuitive conclusion: *the very incommensurability of two hypotheses leads us to challenge whether they are mutually exclusive*. In conventional philosophical interpretations, of course, incommensurability is the hallmark of mutually exclusive hypotheses or paradigms. Slater's and Racker's figures in tandem, however (together with the outcome of the controversy), show us that hypotheses that are conceptually incompatible may nevertheless be "compatible." They may justifiably coexist. One must focus on domains. One must consider how empirical contexts may be differentiated.

Using Kuhn's and Hanson's gestalt metaphor, perhaps, we have become accustomed to an image of either-or competition between complex (incommensurable) wholes as winner-take-all. In a frequent gestalt example, we see either a duck or a rabbit: both cannot exist simultaneously. We sometimes assume in a competitive framework, therefore, that only one can "win." The winner-take-all principle implies that if we accept the duck image, then we accept the duck image exclusively—and no rabbit image is permissible. (Conversely, we may choose the rabbit image exclusively.) However, we may easily imagine scenarios where we may differentiate the contexts in which duck and rabbit interpretations are appropriate, even for the "same" image (far example, in a collection of rabbit drawings versus duck drawings). Either-or principles still apply, here: the final solution is not a hybrid image, half-rabbit, half-duck. The two interpretations remain distinct and incompatible (incommensurable) in the sense that their meanings are non-recombinable and resistant to hybridization. Winner-take-all assumptions break down, however, when we can specify contexts that justify each (holistic) interpretation. The whole rabbit and the whole duck may possibly both be accepted, each in clearly differentiated domains. Indeed, we often do flip happily back and forth between the two gestalts in an extended time frame (differentiating, for example—however weakly—in time). The gestalt example illustrates again that incompatible or 'incommensurable' theories need not be mutually exclusive. The either-or framework, so prominent in the Super Bowl and in Slater's review, leads us to view theories as polar opposites—as rivals—and to view choice as eliminative: only one will win. The outcome of the ox-phos debate, however, foreshadowed in Racker's scoreboard, is a stunning example of a "violation" of assuming a winner-take-all principle. In some cases—here, where domains are notably asymmetrical—two incompatible (incommensurable) hypotheses, such as the chemical and chemismotic hypotheses, may each still be justified.

To summarize, then, where alternative concepts are commensurable—that is, functionally intersubstitutable—or where whole hypotheses are incompatible but coincide in their domains, the either-or, winner-take-all framework applies well. Where hypotheses are based on incompatible concepts and, at the same time, their domains diverge, then a framework of differentiating domains is more appropriate. Incommensurability (exemplified in Slater's review) and asymmetric domains (revealed in some of Slater's phrases and captured more fully in Racker's scoreboard) are thus two contextual clues for abandoning the "Super-Bowl" mode of competition—and for focusing instead on sorting domain items. While incommensurability may tend to lead us to polarize two hypotheses in either-or terms and to regard them as mutually exclusive, it may instead be the very signal that competition is no longer winner-take-all.

4. Differentiation and Strategies for Resolution

Debate on ox-phos continued to unfold, and resolution to the controversy became clearer with additional findings. Racker was thus able to present a "revised scoreboard" one year later (after Racker and Horstman 1972, 15):

REVISED SCOREBOARD

	Hypothesis	
	Chemiosmotic	Chemical
Role of membrane	+	-
Model systems	-	-
Uncouplers and ionophores	+	-
Proton translocation	+	-
Topography of oxidation chain	+	-
K+ and Ca++ transport	-	+
The Painter and Hunter experiments	-	+
Exchange reactions	-	+

Racker still assumed an either-or, winner-take-all orientation, though now he felt that "the balance is shifting in favor of the chemiosmotic hypothesis." If one disregards the sum-total approach, though, one notices, more fundamentally, that new categories had been added to the scoreboard. The domains (or data domains) or evidence had become, literally, more finely resolved. "Ion transport" had been divided by type of ion into "proton translocation" and "K^+ and Ca^{++} transport"; this was critical because formerly ambiguous, confusing evidence now sorted itself more neatly, in Racker's view, between the two hypotheses. New categories of phenomena had also been added: topography of the oxidation chain and the Painter-Hunter experiments—again, contributing positively to each hypothesis. In Racker's revised scoreboard, one sees the growing articulation or differentiation of the evidence: the chemiosmotic hypothesis now explained the role of the membrane, ionophores and topography, for example—while the chemical hypothesis explained exchange reactions and transport of potassium and calcium ions. While the structure of the sports-like scoreboard still asked us to compare right versus left columns, the significant distinction was, in fact, between top rows and bottom rows. The revisions to the scoreboard all clarified the distributed of the data—and indicated more clearly how to reconcile the two hypotheses and to differentiate their once overlapping claims. Epistemically, then, we need to focus on how the differentiation occurs, how we move from Racker's first scoreboard to its revision, or perhaps even how we frame the first scoreboard.

Some recent philosophical approaches, focusing on experiment, have noted the significance of teasing fact from meaningless data, or fact from backgrond "noise": distinguishing fact from artifact (Galison 1986; Latour and Woolgar 1979). In the ox-phos debate, however, the significant experimental task was teasing fact from fact.

One central role of experiments in resolving the disagreement was to sort or partition the evidence. Given a Bayesian framework, for instance, one focuses less on comparing rival hypotheses and more on comparing different ("rival"?) categories of evidence. In differentiating domains, one construes the critical variable in the expression P(H/E) as the evidence (E), not the hypothesis (H). "Super-Bowl" principles compare $P(H_1/E)$ v. $P(H_2/E)$. Instead, the more important comparison may be expressed as $P(H/E_1)$ v. $P(H/E_2)$. The task is to characterize the boundaries of E properly ($E = E_1$, E_2, ... or E_n?). Experience from the ox-phos case, at least, suggests that "crucial tests"—those that embodied an either-or strategy and aimed to decide between two hypotheses—were relatively unsuccessful. The complexity and incommensurabilty evident in Slater's review may suggest why. Rather, researchers laid "claim" to certain domain items through demonstrations—that is, through concrete examples that the hypothesis "mapped" causal relationships in the appropriate domain (Allchin 1992a). Demonstrations were thus needed over a wide domain to sort the "territory." This case illustrates how the experimental resolution of domains (differentiation) is ultimately coupled to the resolution of disagreement. The controversy was resolved, in both senses of the word.

Diagnosing such occasions is crucial. The diagnostic tool as a strategy lies on the cusp between the context of discovery and context of justification. The intent is to justify, but the occasion is one where knowledge is incomplete and where discovering further information may be helpful before justification is complete or stabilized. Characterizing the proper competitive framework allows one to organize available information and ask whether differentiation may be possible—and to identify exactly where further information would be valuable. In short, a carefully framed debate articulates, and in a sense justifies, an ensuing research agenda. Strategic thinking about competition can guide us from Racker's first scoreboard to his revised scoreboard—and from there, to the resolution of controversy.

When the Ox-Phos Controversy finally ended, there was no single winner as there is in the Super Bowl. Both hypotheses had "survived" the competition, but their boundaries or scope had been dramatically reshaped. There was no "one-best" theory. There was no exclusive "winner." Rather, there were several interrelated, though still quite distinct, theories or models, each with its own scope, domain or "niche." What had once been viewed as competing, mutually exclusive, irreconcilable hypotheses became, through the sorting of evidence, complementary theories. The distribution of evidence that guides such sorting can thus be as important as the overall "score" of the evidence itself. The process of differentiating domains—vividly captured in Racker's "scoreboards" of competing hypotheses—must thus play a significant role in any complete model of theory choice in science.

Note

[1]Of course, one may generalize the pattern of differentiation beyond the ox-phos episode. Consider, for example, the conflict between phlogistonists and "anti-phlogistonists" in the late 18th century, typically cast as one of the most dramatic examples of either-or, winner-take-all competition in history (e.g., Kuhn 1962, Chap. 7). We have often assumed the concepts of oxygen and phlogiston were mutually exclusive, but late phlogistionists often accepted the discovery of oxygen. Further, their concerns about the generation of heat and light in burning, ignition, phosphorescence, electricity, and the relationship of animal heat and coal to plants and the sun as a source of light—all aspects of what we would call energy—were all warranted

(Allchin 1992b). Lavoisier, of course, focused on naming elements and studying reactions with a balance. Thus, there was a crude differentiation of the domains of matter and energy (even for combustion itself)—though the problems in the second category were largely not tractable at the time. Though the concepts of oxygen and phlogiston may have fallen in separate paradigms (see Kuhn), they could nonetheless be accommodated through differentiation.

References

Ackerman, R.J. (1985), *Data, Instruments and Theory*. Princeton: Princeton University Press.

Allchin, D. (1990), "Paradigms, Populations and Problem Fields: Approaches to Disagreement", in *PSA 1990*, Volume 1, A. Fine, M. Forbes and L. Wessels (eds.). East Lansing: Philosophy of Science Association. pp.53-66.

______. (1992a), "How Do You Falsify a Question?: Crucial Tests versus Crucial Demonstrations", in *PSA 1992*, Volume 2, D. Hull, M. Forbes and K. Okruhlik (eds.). 1: East Lansing: Philosophy of Science Association. pp.274-88.

______. (1992b), "Phlogiston After Oxygen", *Ambix* 39(3): 110-16.

Franklin, A. (1986), *The Neglect of Experiment*. Cambridge: Cambridge University Press.

Galison, P. (1986), *How Experiments End*. Chicago: University of Chicago Press.

Goodman, N. (1965), *Fact, Fiction and Forecast*, 2nd ed. Indianapolis: Bobbs-Merrill.

________. (1976), *Languages of Art*. Indianapolis: Hackett.

________. (1978), *Ways of Worldmaking*. Indianapolis: Hackett.

Howson, C. and P. Urbach. (1989), *Scientific Reasoning: The Bayesian Approach*. La Salle: Open Court.

Hoyningen-Huene, P. (1993), *Reconstructing Scientific Revolutions: Thomas S. Kuhn's Philosophy of Science*. Chicago: University of Chicago Press.

Hull, D. (1988), *Science as a Process*. Chicago: University of Chicago Press.

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

______. (1992), "The Trouble with the Historical Philosophy of Science", Cambridge, Mass.: Harvard University, Dept. of the History of Science.

Kukla, A. (1992), "Ten Types of Scientific Progress", in *PSA 1990*, Volume 1, A. Fine, M. Forbes and L. Wessels (eds.). East Lansing: Philosophy of Science Association. pp.457-66.

Lakatos, I. (1970), "Falsification and the Methodology of Scientific Research Programmes", in *Criticism and the Growth of Knowledge*, I. Lakatos and A. Musgrave (eds.). Cambridge: Cambridge University Press. pp.91-195.

______. (1978), "Popper on Demarcation and Induction", in *Philosophical Papers*, Volume 1, J. Worrall and G. Currie (eds.). Cambridge: Cambridge University Press. pp.139-67.

Latour, B. and S. Woolgar. (1979), *Laboratory Life*. Princeton: Princeton University Press.

Laudan, L. (1977), *Progress and Its Problems*. Berkeley: Univeristy of California Press.

______. (1990), *Science and Relativism*. Chicago: University of Chicago Press.

Laudan, R., L. Laudan and A. Donovan. (1988), "Testing Theories of Scientific Change", in *Scrutinizing Science*, A, Donovan, L. Laudan and R. Laudan (eds.). Baltimore: Johns Hopkins University Press. pp.3-44.

Niiniluoto, I. (1990), "Measuring the Success of Science", in *PSA 1990*, Volume 1, A. Fine, M. Forbes and L. Wessels (eds.). East Lansing: Philosophy of Science Association. pp.435-46.

Popper, K. (1975), "Rationality of Scientific Revolutions", in *Problems of Scientific Revolutions*, R. Harre (ed.). Oxford: Oxford University Press, pp.72-101.

Racker, E. (1970), "Function and Structure of the Inner Membrane of Mitochondria and Chloroplasts", in *Membranes of Mitochondria and Chloroplasts*, E. Racker (ed.). New York: Van Nostrand Reinhold. pp.127-71.

Racker, E. and Horstman, L.L. (1972), "Mechanism and Control of Oxidative Phosphorylation', in *Energy Metabolism and the Regulation of Metabolic Processes in Mitochondria*, Myron A. Mehlman and Richard W. Hanson (eds.). New York: Academic Press, pp. 1-25.

Rowen, L. (1986), *Normative Epistemology and Scientific Research: Reflections on the "Ox-Phos' Controversy, A Case History in Biochemistry*. Ph.D. dissertation, Nashville: Vanderbilt University.

Shapere, D. (1984), "Scientific Theories and Their Domains", in *Reason and the Search for Knowledge*. Dordrecht: D. Reidel, pp. 273-324.

Sindermann, C. (1982), *Winning the Games Scientists Play*. New York: Plenum.

Slater, E.C. (1971), "The Coupling Between Energy-Yielding and Energy-Utilizing Reactions in Mitochondria", *Quarterly Review of Biophysics* 4: 35-71.

Weber, B. (1991), "Glynn and the Conceptual Development of the Chemiosmotic Theory: A Retrospective and Prospective View", *Bioscience Reports* 11(6).