CHICAGO JOURNALS

The Structure of a Scientific Paper[*]

Frederick Suppe[†‡]

Department of Philosophy, Committee on the History and Philosophy of Science,
University of Maryland

Scientific articles exemplify standard functional units constraining argumentative struc-
tures. Severe space limitations demand every paragraph and illustration contribute to
establishing the paper's claims. Philosophical testing and confirmation models should
take into account each paragraph, table, and illustration. Hypothetico-Deductive,
Bayesian Inductive, and Inference-to-the-Best-Explanation models do not, garbling the
logic of papers. Micro-analysis of the fundamental paper in plate tectonics reveals an
argumentative structure commonplace in science but ignored by standard philosophical
accounts that cannot be dismissed as mere rhetorical embellishment. Papers with illus-
trations often display a second argumentative structure differing from the text's. Con-
straints on adequate testing and confirmation analyses are adduced.

"Experiments are about the assembly of persuasive arguments, ones that will stand up
in court. . . . The task at hand is to capture the building-up of a persuasive argument
about the world even in the absence of the logician's certainty."
                                              —Galison, *How Experiments End*, 277.

1. Introduction. Scientific knowledge is public, and before results are accepted they must be disseminated, evaluated, and credentialed into a scientific discipline's domain (Suppe 1993). Dissemination can be an internal technical report, a "poster session" presentation, or a talk. But the premier format is the refereed scientific journal article. Surprisingly there are no structural analyses of that ubiquitous scientific product, the journal article, in the philosophical literature.

Space is precious in scientific journals and editorial policies require truncated efficient presentations. Draft papers sometimes shrink by 80% before being published in *Nature* or *Science,* and authors often bemoan insufficient space to fully present their case or explain crucial procedures. Hefty page charges (£500 *each* for *Nature* pages having color graphics) mitigate against wasted space or gratuitously included diagrams. Thus a reasonable supposition is that everything—down to each table, graphic, paragraph and even individual sentences—makes some significant contribution to the argument whereby the paper advances its claims.[1] There now emerges a strong presumption that the "logic of testing" in fact operates at that microlevel, hence

> (1) An adequate philosophical account of testing ought to apply at the microlevel and be able to take into account everything in the write-up. The only exceptions should be items for which specific reasons for exclusion can be given.

The question becomes whether any standard philosophical accounts of testing meet this standard.

I analyze the organizational and argumentative structure of the fundamental paper in plate tectonics as a case study to show standard Hypothetico-Deductive (HD), standard Bayesian Inductive, and Inference-to-the-Best-Explanation (IBE) models of testing and confirmation fail criterion (1): Insofar as they are even applicable at the microlevel, they ignore most of the scientific substance and seriously garble the logic of scientific papers—a logic readily discernible in the paper's internal argumentative structure.

2. Architectonic of a Scientific Paper. An examination of diverse natural and social scientific papers (Suppe 1999, 2–II–4)[2] involving report of or recourse to scientific data to establish hypotheses, models, or theories revealed the following overall format:

1. This is not reasonable for scientific books.

2. References to this forthcoming work are by volume-chapter-section numbers, here Volume 2, Chapter II, Section 4.

A. Abstract
B. Introduction
C. Theoretical Background
D. Experimental or Observational Techniques
E. Samples
F. Data Analysis
G. Results or Observations
H. Discussion
I. Summary/Conclusions
J. Acknowledgments
K. References
L. Appendices

Items B and C frequently were combined, as were D–F as a "Methods" section, F–G as a "Results" section, and G-H.

Content analyses of articles (ibid.) reveals that functionally the pieces of a published observation report do the following:

Present the *data* or results of the observation (experimental or not).

Make a case for the *relevance* of the observation and its results to the concerns of the target scientific community.

Provide sufficient detail about observational setup or *method*—the design and circumstances of the observation or experiment—to facilitate methodological evaluation and possibly even *replication* of the study.[3]

Provide an *interpretation of the data* which yields the specific experimental or observational *claims* and is justified on the basis of arguments designed to *anticipate and erase specific doubts* that otherwise might be appropriately raised.

Identify and acknowledge other specific doubts that appropriately might be raised against the study's claims or affect the epistemic status accorded those claims within the discipline.

The argumentative structure of a paper is not a reconstruction of the authors' thought processes that gave rise to the results presented and defended in the paper (Medawar 1963, 377), though they often are reflections of experimental activities (Galison 1987, Ch. 4). "We must not forget to distinguish between the private and public phases of sci-

---

3. Although scientists often stress the importance of replication and claim the purpose of the methods section is to make replication possible, examination of actual scientific practice casts both claims into serious doubt. See Suppe 1993, Section 8; Lowe 1980; Collins 1975, 1985; Mulkay and Gilbert 1986; Mulkay 1984; Kitcher 1991, 9; and Galison 1987, 260.

entific work. When scientists' findings enter the public domain, they become subject to rigorous policing, to a degree perhaps unparalleled in any other field of human activity" (Mulkay 1984, 271). Sociological discourse analyses of differences between literary versions of science, including the published research report and informal discourse or "shop talk" among scientists (Mulkay 1984, 1985; Mulkay and Gilbert 1986; Gilbert and Mulkay 1984; Lynch 1985; Garvy and Griffith 1971; Gilbert 1976), find scientific research reports are "constructive in the way sequences of procedures are reported in contrast to 'actual' sequences of performance" (Lynch 1985, 150–151). It is dangerous to base accounts of scientific work on published research reports which often do not correspond to actual thought processes that produced profound and seminal scientific work (Holton 1973, 18).

The methods, data, interpretations, relevance, and replicability of observational claims are main dimensions of assessment in the credentialing process whereby knowledge claims are allowed into the scientific discipline's domain of shared putative knowledge (Suppe 1993). So, too, are the explicit justificatory arguments and acknowledgments of specific doubts included in scientific papers. The latter are selectively done, confining attention only to those specific doubts the discipline recognizes as legitimate counterpossibilities. In short, the principal tasks of a scientific paper are to present knowledge claims and support them with sufficient explicit justification to enable discipline members and gatekeepers to evaluate whether to accept those claims and admit them into the discipline's domain of putative knowledge.

The scientific paper is the vehicle whereby the scientist's private discourse and knowledge enters the *intersubjective public discourse arena* as it seeks credentialing into the body of accepted scientific fact. The argumentative structures of scientific papers thus are part of the public discourse arena. The positivist *context of justification vs. context of discovery* doctrines essentially place justification in the public discourse arena, so it is legitimate to treat the scientific paper as the vehicle of testing and confirmation, and to explore the adequacy of standard philosophical testing and confirmation analyses by exploring how well they utilize the scientific paper's richness of content.

3. Case Study: Plate Tectonics (Morgan 1968). Plate tectonics is the highly successful theory about the relative motions of semi-rigid crustal blocks or plates making up the Earth's lithosphere (upper 80–100 km.), postulating that they are bounded by three types of plate boundaries and that the motions can be described kinematically as angular velocities around an "Euler point" described by special versions of the Euler

equations for movements about a point ("pole") on a sphere (Gubbins 1990, Ch. 7; Cox and Hart 1986).

Groundwork for plate tectonics had been laid by the earlier Continental Drift Hypothesis which it supplanted,[4] the discovery of magnetic anomalies, and Harry Hess' Sea Floor Spreading hypothesis (Menard 1986). On April 17, 1967, W. Jason Morgan presented a talk at an American Geophysical Union symposium on sea floor spreading. His talk (Morgan 1967) did not correspond to the published abstract, instead deriving the main postulates of what now is known as plate tectonic theory and demonstrating their consistency with data concerning the relative motions of Africa and South America. That summer new Pacific data became available which Morgan incorporated into an expanded manuscript incorporating with minor changes his AGU talk. It was submitted August 30, 1967, revised November 30 in light of Menard's referee comments and correction of one calculation error detected by Xavier Le Pichon, and published March 15, 1968. The paper has been hailed as having "provided the foundation for all subsequent work on ancient plate tectonics and may have been the most important paper ever written in geology, and certainly in tectonics."[5] Subsequently improved data added further support for the theory which has undergone extensive further development. But the core of the theory today remains essentially unchanged from Morgan's initial formulation.

A detailed outline of Morgan's text is given below. The coding makes it clear that a scientific paper is crafted around items addressing the following concerns:

(2) a. identifying a relevant body of scientific literature and making the case that the study is an original contribution to it;
  b. methodology **[M]**;
  c. advancing specific claims **[H]**;
  d. presenting data or results **[D]** serving as evidence for its claims;

4. Plate boundaries do not correspond to continental boundaries. Since plate tectonics is in close accord with plate boundary observational data but fits continental boundary data poorly, it is a mistake to conflate plate tectonics with, or view it as vindicating, Wegener's Continental Drift. Rather it is a distinct successor theory. Plate boundaries can be identified by earthquake profiles (see Figure 1.5 in Gubbins 1990, 14–15). Continental boundaries are identified by offshore shelves using statistical regression techniques of Bullard, Everett, and Smith 1965.

5. Menard 1986, 293. Menard chaired the AGU session where Morgan first presented his work but did not recall even hearing it. As the *Journal of Geophysical Research* referee Menard initially was quite unsympathetic, perhaps even hostile, to it and delayed its publication (Le Pichon 1991, 4). This makes his later assessment all the more impressive.

    e.   arguing for a preferred interpretation of those results by
         i.   specifying and motivating the preferred interpretation **[I]**
         ii. consideration of specific doubts **[Q]** which provide alternative interpretations of results;
         iii. impeaching **[R]** as many of the competing interpretations as possible.

In typical scientific papers each paragraph contributes to at least one task in (2).

4. The Argumentative Structure of a Scientific Paper. We next examine and code Morgan's text using (2)'s designators, then extract its argumentative structure which is a reflection of (2) and typical of scientific papers advancing data-based claims.

The paper begins with a long two-paragraph "Abstract" detailing the various sub-hypotheses H_1, H_2, H_3, H_4, H_5 and corollaries C_1, C_2, C_3 (see below) that together constitute the paper's result **H**.

Paragraphs 1–8 present the main postulates of plate tectonics in essentially their present-day form, although Morgan terms them "blocks" instead of "plates". In this section, Morgan often uses 'hypothesis' interchangeably with 'assumption', yet one can distinguish empirical hypotheses to be tested from assumptions introduced to "give . . . this model mathematical rigor" (paragraph 3). Care must be taken deciding what to construe as the "hypotheses" to plug into philosophical models of testing or confirmation.

Paragraph 1 introduces two main hypotheses:

> **[H_1]** The earth's surface is made of 20 units or blocks
>
> **[H_2]** Three types of block boundaries (ridges with symmetrical spreading motions normal to their strike, transforms with sheer motions but not convergence or spreading, and trenches with convergent motions) determined by present day tectonic activity:
>> rise type (new crustal material formed)
>> trench type (crustal surface destroyed) including other compressive systems
>> fault type (crustal surface neither created nor destroyed)

which provide a "general framework" for extending the transform fault concept to a spherical surface for describing present-day continental drift and gives examples of blocks & boundaries **[D_1]**.

Paragraph 2 anticipates the objection that slow compressive boundaries are hard to delineate **[Q_1]**, acknowledges that he has freely placed boundaries at likely places **[R_1]**, gives examples **[R_2]**, then indicates that

neither the exact numbers of blocks nor their precise boundaries are crucial to the paper's claims **[R₃]**.

Paragraph 3 lists two key "assumptions" (also called "hypotheses")

> **[H₃]**Each crustal block is perfectly rigid;
> **[H₄]**distances between islands do not change; no stretching, large dike injection, thickening, distortion

which "gives . . . mathematical rigor" and indicates that the model will be correct if "in accord with observation" and if only partially valid the enterprise will be to assess the extent of distortion by comparing the model with observations. Popper-like falsification is not presented as an option.

The derivation of the plate tectonic model occurs in two stages: First a planar version is introduced, objections are anticipated

> **[Q₂]**Will vector addition via different routes yield same relative velocity?
> **[Q₃]**Non-parallel (45°) transform fault strikes are allowed;

and the latter rebutted with

> **[R₄]** Examples where this happens.

Specific methods that carry over to the spherical generalization are given for key calculations.

> **[M₁]**Magnetic anomaly pattern projected along line parallel to direction of relative motion enables determining velocity of relative motion.
> **[M₂]**If ridge pattern remains symmetric, axis of ridge has drift velocity = vector average of the velocity of the two sides.

Then in paragraphs 7–8 the model is generalized to a sphere

> **[H₅]**The relative motion of two rigid blocks on a sphere may be described by an angular velocity vector by using three parameters—two to specify pole location, one the angular velocity magnitude.

and three corollaries drawn. One of these

> **[C₁]** All the faults common to two blocks must lie on small circles concentric about the pole of rotation.

is an object of test. The next **[C₂]** concerning velocity of one block relative to the another varying along a common boundary is dealt with by the "half-velocity perpendicular to ridge strike" data-analysis *convention.* The remaining **[C₃]** that is an object of empirical test carries

over from the planar version and concerns vector additivity of relative velocities and the congruence of relative velocity estimates via different routes. Equations are given, and $[M_1]$ and $[M_2]$ are generalized to the spherical case:

> $[M_3]$Relations given for projecting magnetic anomaly patterns to a line perpendicular to ridge strike that enable calculating spreading velocities given latitude, longitude, and strike of a ridge crest point.

We are told the "critical test" of plate tectonics will be via *congruence* between observational data and these corollaries. This will be demonstrated for the *four* boundaries occurring between the postulated 20 plate blocks for which any data were available.

Such demonstration is complicated by the fact $[Q_{10}]$ that for the Pacific/North American boundary, Menard 1967 failed to satisfy $[C_1]$. More generally, for each boundary there were published data which were incompatible with $[C_1]$ or calculated block spreading rates. The only hope for establishing the compatibility of plate tectonics with available observations was to impeach anomalous cases. Thus the main argumentative structure of Morgan's paper is to analyze the extant data sets relevant to a given plate boundary, impeach selected observations to remove them from consideration, and then show that there is adequate fit among the unimpeached observations.

The impeachment arguments are not straight-forward. The best available data concerned the relative motion of the African and the South American blocks. Paragraph 9 determines that plate boundary data (Table 1 = $[D_2]$) can be fitted to small circles rotating about

> $[I_1]$ pole location 58°N ± 5° latitude and 36°W ± 2° longitude.

The determination counterfactually assumes North and South America are a single block. In paragraph 10 transcurrent earthquake mechanism solutions (Table 2 = $[D_3]$) are used to obtain an alternate pole location solution $[I_2]$ which has a poorer common intersection fit $[Q_4]$. That fit then is dismissed on grounds of inaccuracies inherent in the data $[R_6]$. Paragraph 11 shows $[R_7]$ that the conflict between these two methods of calculating pole locations is not an artifact of treating North and South America as a single block $[Q_5]$.

Paragraph 12 then compares Atlantic Ocean spreading rates $[D_5]$ and values obtained from data $[D_4]$ via the model to calculate an alternative pole point $[I_3]$ 62°N ± 5°, 36°W ± 2° that is incompatible with paragraph 9 results $[Q_6]$. Paragraph 13 shows that if one uses this latter pole position to *recalculate the data* used in paragraph 9 $[I_4]$, one also gets a good fit of the paragraph 9 data. So rejecting the original par-

agraph 9 fit, excellent agreement between theory and data is obtained. Remaining doubts $[Q_7]$, $[Q_8]$ about the fit (known problems with Vine's data $[D_5]$) are countered with quite specific instructions for additional data $[R_8]$, $[R_9]$, $[R_{10}]$ that would settle things (paragraphs 14–15). Paragraph 16 anticipates a potential confusion $[Q_9]$ between average Africa/North American relative motion since splitting which requires a pole of rotation of 44.0°N, 30.6°W (Bullard et al. 1965 data = $[D_6]$) and present relative motion with pole 62°N, 36°W that was used in paragraph 9 and argues the two poles are compatible under plausible scenarios where half the motion lies around the present pole $[R_{11}]$.

Paragraph 17 deals with Menard's 1967 determination that not all Pacific block fracture zones are on great circles $[Q_{10}]$. Morgan rebuts this $[R_{12}]$ by showing that great circles concentric on pole 79°N, 111°E coincide with all but two fracture zones that most likely are due to North America overriding and interfering with flow of northernmost end of rise at early date, indicating the Pacific block once moved away from the North American toward trenches off New Guinea and Philippines, with the pattern changing about 10 m.y. to present orientation.

The summer before submitting his paper Morgan obtained access to new Lamont data on Pacific spreading rates and from them constructed great circles perpendicular to the fault segments having common intersection near 53°N, 53°W $[I_5]$. These and other data (Table 3 = $[D_7]$) concerning the Pacific/North American block boundary were not fully compatible with plate tectonics. In particular, paragraph 18 notes the Juan de Fuca region $[Q_{12}]$, Northern California portions of the San Andreas Fault, and the Elsinore Fault in Southern California are anomalous and determination of plate boundaries is problematic in two other regions $[Q_{11}]$. Paragraphs 19–23 deal with these problems. For the ambiguous boundary regions he notes that with hindsight a different choice of faults could have been made, hence the apparently significant shift in lines 1, 2, 3, 4 is not significant $[R_{17}]$.

For the anomalous boundary regions Morgan mounts arguments to the effect that *since these regions violate the assumptions of plate tectonics, they are irrelevant to the testing of plate tectonic theory* $[R_{13}]$– $[R_{16}]$. He then shows that if one excludes data for anomalous regions one obtains satisfactory fit between plate tectonics and observation with pole position of 53°N ($\pm$ 3°), 53°W ($\pm$ 5°) $[I_5']$.

Philosophers of science tend to be scandalized by this argumentative form. However, the fact that this is one of the most thoroughly examined and evaluated papers in the history of science and that *scientists* have not seen fit to challenge this argumentative strategy suggest that Morgan's moves here are legitimate. Thus they should call into question conventional philosophical understanding of the relations between

data and theory, suggesting that adequate of philosophical analyses of testing and confirmation should accommodate such reasoning.

Key to understanding Morgan's moves here is the fact that theories need not accommodate cases outside their scope (Suppe 1989, 141–145). The scope of plate tectonics is those block or plate boundaries which, to the first approximation, act as if the plates were fully rigid, with clean boundaries of the three postulated sorts, etc. But plate tectonics is a kinematic theory, and various mechanisms interfere with the rigid plate motions. In particular some of these produce plate boundaries that do not conform to the idealized assumptions of plate tectonics. In laboratory experiments we use experimental control to minimize the influence of factors not taken into account by the theory under test. And when such attempted control fails, we ignore those experimental results when evaluating the theory. Plate tectonics must be tested in the "natural laboratory" where we lack potential for experimental control of extraneous factors. Instead we exclude observations contaminated by intrusion of *identifiable* extraneous factors—influences not taken into account by the theory—and assess the theory only in light of reasonably pure cases. This is exactly what Morgan does.[6]

Paragraph 24 also anticipates doubts [Q_{13}] over fit scatter being greater than the Atlantic Ocean case, countering by discussing limitations of poorer fit and how scatter affected determination of pole position [R_{18}]. Then paragraphs 25–26 show that excluding problematic earthquakes [Q_{14}] in Syke's data [D_8], justified by prior exclusion of Juan de Fuca and other anomalous regions [R_{19}], yields a congruent pole position of 53°N ± 6°, 53°W ± 10° [I_6]—which, when certain doubts [Q_{15}] are countered [R_{20}], yields a congruent estimate [L_7] of motion of the Pacific block relative to North America of 4.0 ± 0.6 cm./yr. based on data [D_9].

In paragraph 27 pole locations [I_8] and spreading rates [I_9] for the Antarctic block relative to the Pacific block are calculated from data [D_{10}] and shown to agree with LePichon's 1968 result [R_{21}]. Implicit doubts [Q_{16}] about the determinations are countered with paragraphs 28–30 discussing controls over latitude and longitude fits [R_{22}], effects

6. If one maintains that theories provide counterfactual descriptions of how real phenomena *would behave were they isolated* from influence by factors not taken into account by the theory (Suppe 1989), then experimental confirmation depends on data obtained under controlled circumstances where extraneous influences are eliminated by control. (For the logic whereby this is epistemologically efficacious see Suppe 1999, 2–IV–3). The Morgan case shows that sensitive evaluation of data enables one to sort pure and impure cases, confining evaluative attention to pure cases, as an alternative to experimental control.

of spreading rate errors [R₂₃], and how calculations were done without knowing ridge strikes [R₂₄].

For the Antarctic/African block relative movements, there was insufficient data to fix pole locations. However, surrounding data enables different *predictions* of pole location via different vector addition paths. Morgan does this [M₄] in paragraph 31 and finds relatively close agreement [I₁₀]. Despite lack of data for direct determination of pole location or check of such estimates, such consonance of predictions is sufficient to show that vector addition [C₃] is not an unrealistic consequence of plate tectonic theory [Q₂]. Thus the data here suffices to test a further implication of the theory. Paragraphs 31–33 anticipate doubts about the determination [Q₁₇] involving other data [D₁₁] and rebut them [R₂₅] [R₂₆] [R₂₇].

"Conclusions", paragraph 34, weakens claim [H₁] of the Abstract to [H₁*] "favoring the existence of large 'rigid' blocks" describable in accordance with the Euler equations and then counters any remaining residual doubts [Q₁₈] with further remarks, speculations, and conjectures—including postulating hot spots—for subsequent test when data is available [R₂₈].

Now we turn to what Morgan 1968 in fact established. This paper is so fundamental that it has been the subject of retrospective assessments, and there is scientific consensus as to exactly what this paper established. Le Pichon (1977, 367) summarizes that consensus: Morgan's 1968 paper established

> H₂*:"that plates can be considered in the first approximation as rigid."
> H₅*:"that plate kinematics can be described in a satisfactory fashion" by the Euler equations and other main assumptions of plate tectonics.

It follows from **H₂*** that

> Morgan did *not* establish the truth of the hypothesis that plates are rigid.

(Indeed, Morgan never did believe they were rigid and still doesn't.[7])

> Morgan did *not* establish the truth of the 20 plate model.

(Le Pichon's 6 plate model (1968) also works, and today it is common to suppose 23 plates.)

7. Personal communication.

> Morgan did *not* provide an explanation *why* the plates move or
> *why* they move with the velocities they do.

(Plate tectonics is a kinematical theory, not a dynamical one.).

These observations, aided by (2) enable us to extract the argumen-
tative structure for the text of Morgan 1968 which is summarized in
Model 1. Note that each of Morgan's paragraphs contributes to the
argument.In summary,

> (3) the basic argumentative structure is to present data, give a fa-
> vored interpretation of the data that involved throwing out
> some of the data, show the remaining data compatible with the
> plate tectonics hypotheses, anticipate doubts or objections that
> favor alternative interpretations of the data, rebut as many of
> those alternative interpretations as one can, slightly downgrade
> the strength of claims to reflect any unresolved interpretative
> doubts, and conclude the downgraded version of the claims.

When a text is graphically replete (as Morgan's is with 22 figures or
subfigures, 5 tables, and a set of displayed equations) scientists often
do not read the entire text. Instead they read the title, abstract, tables,
figures and captions, and conclusions. In a well-crafted paper this
"graphic" path through the manuscript presents a second interpreta-
tive argument distinct from that of the text. Model 2 presents the
graphic argumentative structure of Morgan 1968. It is less complex
and rebuts fewer specific doubts about interpretations than the text
argument in Model 1.

Although all but five paragraphs discuss specific figures, Figure 14-c
never is referred to in the text. Sociologists of science previously have
noted and correctly analyzed this phenomenon (Bastide 1990, 209–
210). The argumentative structure of a paper in large part is concerned
with blocking alternative interpretations of one's results. Textual and
graphic accounts sometimes differ in the alternative interpretations
they invite, and where they do text and figures may require non-parallel
"blockers" to prevent undesired interpretations.

Finally, I note that although Morgan 1968 depends heavily on pre-
viously published data—especially sea floor spreading rates and mag-
netic anomaly data, and associated theories or models—it is the first
articulation *and* test of plate tectonic theory. Thus its argumentative
structure is essentially self-contained and is especially suitable for eval-
uating philosophical testing and confirmation models.

5. Failure of the Hypothetico-Deductive Method.
Positivists distin-
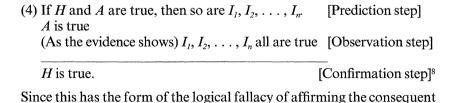guished hypotheses about things that could be directly perceived and

Hypothesis	Data	Associated Interpretation	Associated Doubts	Rejoinders
H			Q_{18}	R_{28}
H_1, H_2	D_1		Q_1	R_1, R_2, R_3
H_3-H_5			Q_3	R_4
C_1: Am-Af	D_2	I_1	I_2	R_6
			Q_6	$I_4,$
	D_3	I_2	Q_4	R_6
			Q_5	R_7
	D_2, D_3	I_4	Q_8	R_9, R_{10}
	D_6		Q_9	R_{11}
C_2: Am-Af	D_4	I_3	Q_6, Q_7	I_4, M_3
	D_5		Q_7, Q_8	$R_8, R_9,$ R_{10}, M_3
C_1: Pac-Am			Q_{10}	R_{12}
	D_7	I_5	Q_{12}	R_{13}
			Q_{11}	$R_{14}, R_{15},$ R_{16}, R_{17}, I_5'
		I_5'	Q_{13}	R_{18}
	D_8	I_6	Q_{14}	R_{19}
C_2: Pac-Am	D_9	I_7	Q_{15}	R_{20}, M_3
C_1: Ant-Pac	D_9	I_8	Q_{16}	R_{21}, R_{22}
C_2: Ant-Pac	D_{10}	I_9	Q_{16}	$R_{21}, R_{22},$ R_{23}, R_{24}, M_3
C_1: Ant-Af	I_3, I_7, I_9	I_{10}	Q_{17}	R_{25}
C_2: Ant-Af	I_3, I_7, I_9	I_{10}	Q_{17}	R_{25}, M_4
C_3	I_3, I_7, I_9	I_{10}	Q_2	R_{25}, R_{26}, R_{27}

Model 1: *Model of Interpretative-Argument Structure of Morgan 1968.* Italicized rejoinders do not fully remove associated doubts, accounting for the weakened conclusion **H*** that the evidence presented here "favors" the hypothesis **H**. Coding Key: Hypothesis; Corollary; Data; Interpretations of data; Queries or doubts about interpretations of data (often in the form of competing interpretations or hypotheses); Methods used in manipulating or interpreting data; Rejoinders to queries or doubts. *R*ejoinders leaving unrebutted residual doubts or that call for better data are italicized. Corollaries are separated for different boundaries between the **Am**erican, **Af**rican, **Pac**ific, and **Ant**arctic blocks or plates.

those that could not. The *HD method* (or *method of hypotheses*) was designed for testing hypotheses that cannot be confirmed by direct observation. It is common to render the method schematically:

Hypothesis	Figure/display	Data (displayed or identified)	Associated Interpretation	Associated Doubts	Rejoinders
H_1, H_2, H_3	Figure 1			Q_1	R_2, R_3
	Figure 2				
	Figure 3				
C_1	Figure 4				
C_2	M_3 Figure 5 Figure 6				
C_1: Am-Af	Figure 7	D_4	I_4		
	Figure 8a		I_1	I_4, Figure 7, Figure 9	I_3
	Figure 8b	D_2 (Table 1) D_3 (Table 2)			
C_2: Am-Af	Figure 9	D_5	I_4		
	Figure 10	D_6		Q_9	R_{11}
C_1: Pac-Am	Figure 11			Q_{10} (tacit)	
	Figure 12	Tobin & Sykes		Q_{10} (tacit)	R_{12}
	Figure 13	Raff & Mason, Vine	R_{12}		
	Figures 14a-d	D_7 (Table 3)		Q_{10} (tacit)	R_{12}
	Figure 14e	Figures 14a-d	I_5' I_7	Q_{11}	R_{14}-R_{17}
	Figure 14f	D_8	I_6		
C_2: Pac-Am			I_7		
C_1: Ant-Pac	Figure 15		I_8		
C_2: Ant-Pac	Figure 16		I_9		
C_1: Ant-Af C_2: Ant-Af	Table 5 (prediction)	I_4, I_7, I_9	I_{10}		
H (= H_1 & H_2 H_3 & H_4 & H_5 & C_1 & C_2)				Q_{18} (tacit)	R_{28}, H_1^* (hence H^*)

Model 2. *Model of Graphic/Tabular Interpretative-Argument Structure of Morgan 1968.* Italicized rejoinders do not fully remove associated doubts, accounting for the weakened conclusion H^* that the evidence presented here "favors" the hypothesis **H**. For the most part, associated doubts are tacit or "expected" given the readership rather than explicitly raised. Coding Key is the same as for Model 1.

(4) If H and A are true, then so are I_1, I_2, . . . , I_n.   [Prediction step]
   A is true
   (As the evidence shows) I_1, I_2, . . . , I_n all are true   [Observation step]

   H is true.   [Confirmation step][8]

Since this has the form of the logical fallacy of affirming the consequent not just any instance of this schema will do. So lots of qualifications

---

8. This is Hempel's 1966 schema [2c] (p. 8) augmented by the truth of A per Hempel's discussion in Section 3.2. Notational changes have been made.

are put in about requiring independence of observations from predictions, excluding ad hoc hypotheses, variety of evidence, the need for the predictions to be "new," many replications, simplicity, etc. While some have wanted to link this scheme to a process for assigning explicit probabilities as measures of degree of confirmation, Hempel (1966, 45–46) expressed doubts this probabilistic twist would work. Regardless, it is clear that many replications are presumed necessary to strongly confirm a hypothesis.

There is a large body of literature debating the merits of this treatment of confirmation, but I intend neither to trace nor enter into those controversies. I simply want to know how applicable the analysis is to the microstructure of hypothesis-testing studies such as Morgan 1968.

Applying (4) to Morgan 1968 it is natural to let H and its constituents $[H_1]$–$[H_5]$ be H and $[M_1]$–$[M_4]$ be A. The basic HD picture is that one predicts values which then are compared with measurements. Nowhere does Morgan use H to predict specific values for Euler points. And it is only used to predict spreading rates in the African/Antarctic boundary case. But those predictions cannot be accommodated by the HD method since there is no comparison of predictions with values determined independently of the hypothesis under test. The only other predictions in the paper are that the data will be "compatible" with $[C_1]$ and $[C_3]$[9] which we take as I. The best fit of HD model (4) to Morgan's paper involves using items in paragraphs 4, 7, and 8 to derive the "predictions." Results pitted against the predictions are found in paragraphs 13, 19, 24, 25, 26, 27, and 31. Thus the HD model takes into account only 30% of the paper's paragraphs.[10]

But that fit does not instance (4): The actual data are incompatible with $[C_1]$ and so via (4) *disconfirm* it. Coherence is obtained only by ignoring incompatible portions of the data and nothing in (4) makes any provision for such exclusion. In one case the same data provided both a confirming and a disconfirming pole position. Further, the conclusion drawn is not H but the weaker $[H_2^*]$ and $[H_5^*]$, to which $[C_1]$ and $[C_3]$ are *not* deductive corollaries. Finally, those 30% of paragraphs are accommodated incoherently at the price of ignoring the main argumentative form of the paper—"If we throw out all this data we can make it fit well enough"—involving *all* of the paragraphs the HD analysis ignores.

9. Recall that $[C_2]$ is satisfied by a convention.

10. One might want to include the results in paragraphs 9 and 12 which are impeached. Then the HD model utilizes 35% of all paragraphs and 29% of paragraphs concerning data and their analysis.

6. Failure of Bayesian Inductive Models. Bayesian approaches construe confirmation and testing probabilistically and begin with:

(5) *Bayes' Theorem:* If $H_1, \ldots, H_n$ are mutually exclusive and jointly exhaustive on condition B and $\mathbf{P}(R, B) > 0$, then

$$\mathbf{P}(H_i, R \ \& \ B) = \frac{\mathbf{P}(H_i, B) \bullet \mathbf{P}(R, H_i \ \& \ B)}{\displaystyle\sum_{j=1}^{n} \mathbf{P}(H_j, B) \bullet \mathbf{P}(R, H_j \ \& \ B)}$$

for $i = 1, \ldots, n$.

The $\mathbf{P}(H_i, B)$ are called the *prior probabilities*, the $\mathbf{P}(R, H_i \ \& \ B)$ are called the *degrees of prediction*, and the $\mathbf{P}(H_i, R \ \& \ B)$ are called the *posterior probabilities*. Intuitively B will consist of information from the domain and background, details of the observational set-up and the like.

The following result can be proved about (5):

(6) *Confirmation Theorem:* Let R be some prediction based on hypothesis H and B in accordance with prior probabilities $\mathbf{P}(H, B)$ and degrees of prediction $\mathbf{P}(R, H \ \& \ B)$, and let R_m be the relative frequency of correct predictions in m tests. Then the posterior probability $\mathbf{P}(H, R_m \ \& \ B)$ will converge on the true probability as m approaches infinity.[11]

If hypothesis testing is understood as confirmation construed as determining the posterior probability that the hypothesis is true, (6) says that repeated testing will give you the correct confirmation probability in the limit and that doing so does not depend on how accurate the estimates of prior probabilities were.[12]

The relevance of (6) to Morgan 1968 is unclear. Perhaps one might treat each of the polar point determinations as a test, or perhaps even each of the points calculated in determining the great circle pole positions. In any case (6) can be construed as relevant when subsequent tests of plate tectonics did occur.

Like (4), the Bayesian account involves making predictions then observing whether the predicted outcome obtains. The primary difference here is that the predictions are probabilistic, not deterministic. That difference has no bearing on what in Morgan's paper counts as the

---

11. For a rigorous statement, see Burks 1977, 84; for the proof, see Savage 1954, 46–50.

12. Good accounts of the Bayesian hypothesis testing are Burks 1977, Ch. 2; Horwich 1982; Howson and Urbach 1989; and Salmon 1967. Within Statistics, more sophisticated Bayesian approaches to scientific inference have been developed. See Lindley 1965; Box 1980, 1983; Box, Leonard, and Wu 1983; and Suppe 1999, 2–VIII–5.

hypotheses, predictions, and results, so they are as in the previous section. Again, only 30% of the paragraphs are accommodated.

Let **M** be [**M₁**] & . . . & [**M₄**], **D** be *previously-published* data [**D₁**] & . . . & [**D₁₁**], and **I** be the various polar points Morgan determined. On the Bayesian account, Morgan should confirm [**H₂***] and [**H₅***] by calculating the prior probabilities

$$P(H_2^*, D \& M)$$
$$P(H_5^*, D \& M),$$

and then the degrees of prediction

$$P(I, H_2^* \& D \& M)$$
$$P(I, H_5^* \& D \& M).$$

Finally these would be plugged into (5) to calculate the posterior probabilities

$$P(H_2^*, I_m \& D \& M)$$
$$P(H_5^*, I_m \& D \& M)$$

where I_m is the relative frequency of great-circle fits of spreading rate fits in *m* trials.

This is precisely what we do not find Morgan doing: Instead of simply noting the percentage of boundaries that can be fit into a great circle, Morgan impeaches then ignores those that do not fit. At no point is his argument probabilistic. This is particularly telling since paragraph 24 is concerned with assessing the adequacy of fit and Morgan could have used and standard goodness-of-fit tests to give probabilistic estimates. Indeed Bullard et al. 1965 had developed such methods for assessing the closeness of continental fit. Morgan could have calculated prior and posterior probabilities and degrees of prediction but did not do so. Bayesians often claim that the probabilistic reasoning is only implicit. But this rings hollow in cases like this where the quantitative data and hypotheses especially lend themselves to probabilistic evaluation. Bayesians owe us explanations both why their analysis ignores 70 percent of the paragraphs concerned with data but also why those they do fit to their analysis do not *overtly* conform to it.

7. Inference to the Best Explanation. The challenge to philosophy of science is making sense of the interpretative argumentation science *actually* deploys to establish theories such as plate tectonics.

(7) HD, Bayesian, and other analyses that construe hypothesis testing as repeatedly predicting results, comparing them with

> independent data, and then using successful predictions as the
> basis for confirmation do serious violence to Morgan's paper.

More broadly there is mounting evidence that the importance such
methods accord to replication or repeated testing is at odds with actual
scientific practice.[13]

IBE avoids those fundamental difficulties (7). Harman introduced
the notion:

> one infers from the fact that a certain hypothesis would explain the
> evidence, to the truth of that hypothesis. In general there will be
> several hypotheses which might explain the evidence, so one must
> be able to reject all such alternative hypotheses before one is war-
> ranted in making the inference. Thus one infers, from the premise
> would provide a 'better' explanation for the evidence than would
> another hypothesis, to the conclusion that the given hypothesis is
> true. (1965, 89)

At first blush IBE seems to conform rather well to Morgan's sorts of
interpretative arguments (Models 1, 2)—whose form (3) roughly is "If
we can find good reasons to throw out enough of the data we can
obtain adequate fit of theory to the data." For hypotheses are inter-
pretations of results. Doubts against a favored interpretation of one's
results translate into competing interpretations, hence yielding alter-
native hypotheses. Thus comparative evaluation of interpretations in
the attempt to impeach all plausible alternatives to the favored inter-
pretation becomes the rejection of alternative hypotheses.

IBE can be summarized schematically as follows

(8) Given the present state of science, $H_0, H_1, \ldots, H_n$ are the only
plausible hypotheses for explaining phenomenon P.
For such and such reasons, H_0 provides a better explanation
of P than do $H_1, \ldots, H_n$.
Therefore,
H_0 probably is true.

It is clear that Harman understands IBE as an ampliative inductive
principle. van Fraassen (1980, 19–22; 1989, Ch. 6) objects to IBE on
the grounds that the premises at most establish that H_0 is the best of
the lot which might be a wretched lot—in which case inferring the con-
clusion would be unwarranted. Fairly "soft" claims often are sufficient
to impeach specific doubts against an hypothesis H even though they

13. See note 3.

are insufficient to establish H (Suppe 1999, 2–II–8, 2–VI–2). This suggests another potential line of argument against IBE:

> (9) Reasons given in rebuttal of alternative hypotheses/interpretations often are too soft or weak to warrant concluding that the unimpeached hypothesis/interpretation is probably true. Thus while IBE avoids some of the fundamental defects of the HD and Bayesian models, it nonetheless is problematic as an analysis of testing and confirmation.

How well does IBE accommodate Morgan's interpretative argument? We noted the consensus that Morgan established not the truth of **[H$_2$]** but rather **[H$_2$*]** "that plates can be considered in the first approximation as rigid." This is not equivalent to establishing that **[H$_2$]** is probably true as required by (8). Indeed, although analyzing the relationships between truth and being "considered in the first approximation as rigid" is a knotty philosophical problem, **[H$_2$]** almost certainly is false. We do not find here the sort of conclusion IBE would lead us to expect.

IBE proponents say surprisingly little about the sorts of explanations schema (8) requires. Harman 1965, for example, never says.[14] He offer a number of examples all of which appear equivalent to explanations *why*. Most philosophical accounts of explanation tacitly or explicitly assume all explanations are equivalent to explanations *why* (Hempel 1965, 334). But that is false since many explanations *how* cannot be reduced to explanations *why*. Indeed, explanations *why* prove to be a proper sub-class of *how possible* explanations (Suppe 1989, Ch. 6). Plate tectonics is a kinematical theory and so does not purport to explain *why* the plates move as they do.

Insofar as *consistency claims* such as Morgan repeatedly makes are *explanations* of data, they are *how possibly* explanations. Thus application of IBE to Morgan 1968 requires the following version :

> (10) Given the present state of science, $H_0, H_1, \ldots, H_n$ are the only plausible hypotheses for explaining phenomenon P.
> For such and such reasons, H_0 is consistent with [explains how possibly] data P whereas $H_1, \ldots, H_n$ are not.
> Therefore,
> H_0 probably is true.

---

14. He does require that to be the best explanation H_0 be "considerably more probable than" $\sim H_0$ given P and that H_0 make P more probable than $\sim P$ (1968, 169). Invoking this requirement makes the applicability of (8) to Morgan's case even more problematic while undercutting none of my objections. So it will be ignored.

As a general principle of ampliative inference, this is not terribly plausible. When we note that the data sets Morgan uses often are sparse or inadequate, the application of (10) to his study is even more problematic. To the extent that (10) is defensible as an ampliative inference principle, it will be when the data sets P are rich and where consistency claims are enhanced by truth-constraining theory.[15] That is not the case in Morgan 1968 where *simple compatibility* of data and theory are being claimed. And the specific Euler points which do fit *were not predicted* in advance by the theory.

IBE is an ampliative inference principle where, presumably, the ability of H_0 to explain P augments the evidential basis B for H_0 enough to warrant its probable truth where B alone is insufficient. Thus in application to Morgan 1968, the detailed argumentative strategies Morgan pursues must be the ampliative source. But those activities consist in throwing out data, ignoring data from boundary regions incompatible with the assumptions of plate tectonics, recalculating results to improve agreement, and then claiming the resulting data are consistent with the theory. It is difficult to see how any of this can be ampliative. For Morgan is making recourse to exactly the sorts of "soft" claims that enable one to impeach competing interpretations of one's data but do little by themselves to establish the correctness of the favored interpretation. To the extent that such "soft" arguments legitimate inferring the probable truth of H_0 under (10) it will be in virtue of the *true* hypothesis being among the $H_0, H_1, \ldots, H_n$ or other inductive warrant. And it is this fact, not the impeachment of competing hypotheses, that shoulders the ampliative burden (Achinstein 1991). But (10) only requires that these be the only plausible hypotheses— which is no guarantee that the true hypothesis is among them. Here "softness" considerations (9) join force with van Fraassen's objections that H_0 may just be the best of a wretched lot.

But suppose we accept (10) for the sake of argument. Crucial to the logic of (10) is having a plurality of hypotheses one brokers amongst. One thing notably absent from Morgan 1968 is consideration of *any* hypotheses alternative to plate tectonics, so the very application of (10) to it is problematic. One might attempt an analysis where $n = 2$ and the competing hypotheses are H_0 and $\sim H_0$. However that does violence

15. For example, production of geological crosssections using fault-bend folding theory involves showing a candidate cross section consistent with well-log and seismic data. However part of that consistency involves the solution being retrodeformable to a pristine state with conservation of various parameters such as bed-thickness. The notion of "consistency" with data here is so complex and constraining that, in practice, consistent solutions tend to be approximately true or true in clearly identifiable aspects. (See Suppe 1999, 2–VII–3 for detailed discussion.)

to paragraph 3, where Morgan describes his enterprise: "If this hypothesis is true, our conclusions will be in accord with observation. If this hypothesis is only partially valid, perhaps we will be able to assess the extent of such distortion by comparing observations with this model" (pp. 1960–1961). There is no mention of falsifying the plate tectonics hypothesis. So we are left with the fact that Morgan is concerned with the evaluation of exactly one hypothesis and engages in none of the comparative evaluation (8) leads us to expect.

Nevertheless, the basic argumentative structure in Morgan 1968 is a dialectic of impeachment. He impeaches Menard's competing interpretation of the African/South American data. Repeatedly he considers data sets D that seem to falsify plate tectonics. Then he analyzes D, raising specific considerations that impeach selected data, resulting in a restricted data set $D' \subset D$, and shows that D' is consistent with **[H]**. The suppositions that D vs. that D' are the data **[H]** should accommodate are *different interpretations* of the data set D. The dialectic of Morgan's argument repeatedly is putting forth a favored interpretation D' of the data and impeaching competing ones such as D.

> (11) It is at the data *interpretation level*, not the *hypothesis level*, that we find dialectic most resembling IBE pattern (8). But interpreting data is *nonampliative.*

The maximal relevant evidence for establishing **[H]** is D. The sorts of considerations used to decide among different interpretations of D typically do not augment the evidential basis for **[H]**—even when additional facts are adduced in the impeachment of items in D.[16]

The import of (11) thus is this:

> (12) Once the data have been collected the evidential basis for establishing H is complete. All subsequent statistical or other data analysis and interpretative argument concerns the interpretation of that data and in no way augments the evidential basis for **H.**

The dialectic of argument we find in Morgan concerns evaluating competing interpretations of data, hence is nonampliative. Thus whatever contribution interpretation and interpretative arguments make to scientific knowledge, it is nonevidential. Morgan 1968 is typical of sci-

---

16. To be sure, additional evidence is adduced. But the evidence is negative evidence against the competing interpretations and ordinarily is not positive evidence in favor of the preferred interpretation of **[H]** unless it functions as a premise in a *modus tollendo ponens* type of inference—a notably rare occurrence in science. In many cases interpretations augment the data with additional assumptions—typically ones that are not well established. Such augmentation usually does not add to the evidential basis.

entific papers in that its argumentative structure largely concerns interpretation of data and proceeds via a dialectic of impeachment. The fundamental mistake of IBE is that it conflates such nonampliative interpretative argument with ampliative inductive arguments. It conflates interpretations of data with the hypotheses being tested.

8. Discussion. Morgan 1968 is atypical of scientific papers in its scientific merit and importance, but it is paradigmatic of how science argues its claims. In the five years I have been studying the structure of data-based papers, I have read over 1000 papers and heard hundreds of scientific talks. With rare exceptions they all evidence the same logic of interpretation and impeachment. Other microanalyses spanning the social to the natural sciences show similar violence done when one attempts to force arguments into the forms of standard philosophical accounts of testing and confirmation. For instance an outstanding psychology of personality paper had two hypotheses. For the first, the HD and Bayesian models could not accommodate what on independent grounds was the most crucial evidence; for the second, they could accommodate all the evidence only by so construing the hypothesis that the evidence was inadequate (Suppe 1999, 2–VI–2). The percentages of paragraphs that must be ignored are similar over a wide range of literature.

One might try to dismiss my challenge (1) on grounds that I am studying the *rhetoric* of papers, not their testing and confirmation *logic* or how they yield *knowledge*. The role of rhetoric is to persuade. One function of the argumentative structure of a paper *is* to persuade others to accept one's interpretations and claims over competitors'. One can either claim the means of persuasion are *epistemic* in the sense of playing a role in establishing scientific knowledge claims or one can relegate them to the *nonepistemic* (Pera and Shea 1991, Kitcher 1991).

Suppose the argumentative structure in Morgan 1968 is dismissed as nonepistemic rhetoric. Presumably the seventy percent of paragraphs the HD and Bayesian models ignore would be consigned to rhetorical embellishment serving no epistemic function. But then advocates of the HD and Bayesian models owe us an explanation why seventy percent of precious journal space is devoted to modes of persuasion having no bearing on establishing the claimed results. IBE advocates need to explain why Morgan chose to devote the *entire* paper to a nonepistemic rhetorical argument rather than the unmentioned IBE argument that "really" establishes the paper's claims.

It is more promising to give rhetoric an epistemic function—especially since the argumentative structures identified here bear close resemblance to the "construction of arguments . . . cutting the data . . .

to isolate causes and eliminate alternatives" Galison (1987, 258) shows is at the heart of experimentation underlying papers. But then it remains problematic why the overt rhetorical arguments of the paper *and* experimentation do not conform to standard philosophical models.

If knowledge is justified true belief, and the justification need not exhaust the grounds for belief, then the argumentative structure of a paper could be primarily rhetorical in the sense that it is designed to convince others to accept or believe the author's interpretations of data and associated claims, and given (12) the data or results *alone* are the evidential basis in virtue of which those beliefs are justified and become knowledge. The argumentative structure thus is instrumental to producing knowledge without being ampliative. I think that is in fact what is going on and elsewhere have presented a non-inductive epistemology and associated credentialing analysis that so accommodates in fine detail the argumentative structure of papers here identified (Suppe 1993, 1997).[17]

9. Conclusions. Scientific papers from diverse disciplines display a common organizational structure and exhibit similar argumentative structures. Once a scientific paper has located its place in intellectual space (the discipline and domain to which it intends to contribute) and presented its results, the main business of the paper is putting forth a favored interpretation of the results and impeaching competing interpretations. Considerations used to impeach interpretations often are soft, and generally do not add to the evidential basis supporting the favored interpretation. When papers involve substantial recourse to graphics there typically are distinct textual and graphical argumentative structures which, while different, evidence the same basic logic of brokering among competing interpretations of data or results.

Interpretative arguments are non-ampliative. The evidential basis for the paper's claims consists of the paper's results or data augmented by background knowledge. Interpretative arguments have as their focus finding a correct interpretation of the results, not a strengthening of the evidential basis. In effect the situation is that the results are evidence for *something* and the argumentative task of the paper is to determine what that *something* is—calibrating one's conclusions to one's findings.

An adequate philosophical account of testing and confirmation must accommodate every paragraph in a scientific paper and make sense of

17. Suppe 1999 makes the case for: the mentioned psychology study (2–VI–2); J. J. Thomson's cathode ray experiments (2–IV–2); Weinberg/Salam electro-weak theory (2–IV–3); and other examples.

the argumentative structures such papers actually exemplify. Standard HD, Bayesian inductive, and IBE models fail this test, presenting garbled accounts of testing and confirmation that render the actual arguments put forth in scientific papers unintelligible. These philosophical analyses are fundamentally inadequate.

REFERENCES

Achinstein, P. (1991), *Particles and Waves: Historical Essays in the Philosophy of Science*. New York: Oxford University Press.

Bastide, F. (1990), "The Iconography of Scientific Texts: Principles of Analysis", in M. Lynch and S. Woolgar (eds.), *Representation in Scientific Practice*. Cambridge, MA: MIT Press, pp. 187–230 .

Box, G. (1980), "Sampling and Bayes' Inference in Scientific Modelling and Robustness", *Journal of the Royal Statistical Society* 243/4: 383–404.

———. (1983), " An Apology for Ecumenism in Statistics", in Box, Leonard, and Wu (eds.) 1983, pp. 51–84.

Box, G., T. Leonard, and C. Wu (1983), *Scientific Inference, Data Analysis, and Robustness*. New York: Academic Press.

Bullard, Sir E., J. Everett, and A. Smith (1965), "The Fit of the Continents Around the Atlantic", *Philosophical Transactions of the Royal Society of London*, Series A, 258: 141–151.

Burks, A. (1977), *Chance, Cause, Reason*. Chicago: University of Chicago Press.

Collins, H. (1975), "The Seven Sexes: A Study in the Sociology of a Phenomenon, or the Replication of Experiments in Physics", *Sociology* 9: 205–224.

———. (1985), *Changing Order: Replication and Induction in Scientific Practice*. Beverly Hills: Sage.

Cox, A. and R. Hart (1986), *Plate Tectonics: How it Works*. Boston: Blackwell Scientific Publications.

Galison, P. (1987), *How Experiments End*. Chicago: University of Chicago Press.

Garvey, W. and B. Griffith (1971), "Scientific Communication: Its Role in the Conduct of Research and Creation of Knowledge", *American Psychologist* 26: 349–362.

Gilbert, G. (1976), "The Transformation of Research Findings to Scientific Knowledge", *Social Studies of Science* 6: 281–306.

Gilbert, G. and M. Mulkay (1984), *Opening Pandora's Box: A Sociological Analysis of Scientists' Discourse*. Cambridge: Cambridge University Press.

Gubbins, D. (1990), *Seismology and Plate Tectonics*. Cambridge: Cambridge University Press.

Harman, G. (1965), "The Inference to the Best Explanation", *Philosophical Review* 74: 88–95.

———. (1968), "Knowledge, Inference and Explanation", *American Philosophical Quarterly* 5: 164–173.

Hempel, C. (1965), *Aspects of Scientific Explanation and Other Essays in Philosophy of Science*. New York: Free Press.

———. (1966), *Philosophy of Natural Science*. Englewood Cliffs: Prentice Hall.

Holton, G. (1973), *Thematic Origins of Scientific Thought: Kepler to Einstein*. Cambridge, MA: Harvard University Press.

Horwich, P. (1982), *Probability and Evidence*. Cambridge: Cambridge University Press.

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

Kitcher, P. (1991), "Persuasion", in Pera and Shea 1991, pp. 3–28.

Le Pichon, X. (1968), "Sea-floor Spreading and Continental Drift", *Journal of Geophysical Research* 73: 3661–3697.

———. (1977), "A Personal and probably Biased View of Plate Tectonics Hypothesis Formulation and Evolution" (abstract), *EOS* 58/6: 367.

———. (1991), "Introduction to the publication of the extended outline of Jason Morgan's

April 17, 1967 American Geophysical Union paper on 'Rises, Trenches, Great Faults, and Crustal Blocks' ", *Tectonophysics* 187/1–3: 1–23.

Lindley, D. (1965), *Introduction to Probability and Statistics From a Bayesian Viewpoint, Volume I: Probability.* Cambridge: Cambridge University Press.

Lowe, E. (1980), "Sortal Terms and Natural Laws: An Essay on the Ontological Status of Laws of Nature", *American Philosophical Quarterly* 17: 253–60.

Lynch, M. (1985), *Art and Artifact in Laboratory Science: A Study of Shop Work and Shop Talk in a Research Laboratory.* London: Routledge and Kegan Paul.

Medawar, P. (1963), "Is the Scientific Paper a Fraud?", *The Listener & BBC Television Review* 70, no. 1798, Thursday, Sept. 12, 1963, 377–378.

Menard, H. (1967), "Extension of Northeastern-Pacific Fracture Zones" *Science* 155: 72–74.

———. (1986), *Ocean of Truth: A Personal History of Global Tectonics.* Princeton: Princeton University Press.

Morgan W. J. (1967), "Rises, Trenches, Great Faults, and Crustal Blocks", manuscript presented April 17, 1967 in a symposium, "Sea Floor Spreading", American Geophysical Union Meeting, Washington. D.C.; manuscript later published in photo-reproduction form as pp. 6–22 of Le Pichon 1991.

———. (1968), "Rises, Trenches, Great Faults, and Crustal Blocks", *Journal of Geophysical Research* 73/6 (March 15): 1959–1982.

Mulkay, M. (1984), "The Scientist Talks Back: A One-Act Play, with a Moral about Replication in Science and Reflexivity in Sociology", *Social Studies of Science* 14: 265–282.

———. (1985), *The Word and the World: Explorations in the Form of Sociological Analysis.* London: George Allen and Unwin.

Mulkay, M. and G. Gilbert (1986), "Replication and Mere Replication", *Philosophy of Social Science* 16: 21–37.

Pera, M. and W. Shea (1991), *Persuading Science: The Art of Scientific Rhetoric.* Canton, MA: Science History Publications.

Salmon, W. (1967), *Foundations of Scientific Inference.* Pittsburgh: University of Pittsburgh Press.

Savage, L. J. (1954), *Foundations of Statistics.* New York: Wiley.

Suppe, F. (1989), *The Semantic Conception of Theories and Scientific Realism.* Urbana: University of Illinois Press.

———. (1993), "Credentialing Scientific Claims", *Perspectives on Science* 1/2: 153–203.

———. (1997), "Science Without Induction", in John Earman and John Norton (eds.), *The Cosmos of Science.* Pittsburgh: University of Pittsburgh Press and Konstanz: Universitätsverlag Konstanz, 1997, pp. 386–429.

———. (1999), *Facts, Theories, and Scientific Observation, Volume 1: A Posteriori Knowledge and Truth, Volume 2: Scientific Knowledge.* Forthcoming (under review).

van Fraassen, B. (1980), *The Scientific Image.* New York: Oxford.

———. (1989), *Laws and Symmetry.* Oxford: Oxford University Press.