Reprinted from SOCIAL STRUCTURES: A Network Approach,
Edited by Barry Wellman and S.D. Berkowitz. 1988. New York: Cambridge University Press

14

Cites and fights: material entailment analysis of the eighteenth-century chemical revolution

Douglas R. White and H. Gilman McCann

The structural-analytic literature of the past decade indicates that transitivity in social networks is an important concern. However, existing structural analytic methods are frequently unable to deal directly with the substantive and methodological problems posed by transitive relations in social structures (Berkowitz, 1982). These difficulties typically surface in one of two contexts.

First, those who model structures through multiple graphs defined onto the same set of elements tend to assume either (1) global transitivity or intransitivity – that is, the same degree of transitivity obtains throughout the network (Johnsen and McCann, 1982), or (2) that there is a sharp and arbitrary limit to the graph theoretic distances over which effects travel (Berkowitz, Carrington, Kotowitz, and Waverman, 1978). Neither strategy allows analysts to examine transitivity, itself, empirically. Second, techniques for examining relations among sets of *overlapping* attributes or ties are still in their infancy. Once again, transitivity among these sets is typically treated a priori rather than as a substantive or empirical problem.

A new technique – *Material entailment analysis* – allows researchers to model concrete situations in which the degree of transitivity present within social structures defined in either of these ways may be investigated empirically. This chapter describes this new technique and outlines ways in which it can be applied to a variety of structural problems in the social sciences.

We address the transitivity problem as a special case of the more general issue of orders and partial orders of variables or attributes. Following Nadel (1957), sociologists and anthropologists are often concerned with clusters of attributes in which the presence of one implies the presence of others. To the extent that such an implication is not reciprocated (symmetrical), we have an ordering of the attributes or cultural values. Consequently, a variety of social and cultural domains may be modeled in terms of "if...then" or set-subset relationships among cultural items. One study, for instance, found a partial ordering of basic color terms of the form "if a language has color term *X*, then it also has color term *Y*" (Berlin and Kay, 1969); and another found a similar ordering in the sexual division of labor such that if women (men) perform certain tasks they also tend to perform specific other tasks (Burton, Brudner, and White, 1977). Similar orderings have been found by other anthropologists, sociologists, linguists, and psychologists (Nadel, 1957; Gagne, 1965; Greenberg, 1966; D'Andrade, 1976). Most of the methodological work on orderings has been done by educational statisticians who utilize only a single overall cutoff level for strengths of ties within a system and an overall test of significance; that is, they tend to focus only on global orderings (Bart and Krus, 1973; Bart and Airasian, 1974; Baker and Hubert, 1977).

A more general approach, which examines both local and global orderings – and which can be used with relational as well as attribute data – is necessary for structural analysis. From a methodological point of view there is a need for a method that can address both morphology (the overall structure) and attributes using the same rules, or language. Material entailment analysis is such a method. It models patterns of either structure (relational data) or attributes in terms of entailment chains or hierarchies. These structures reflect "if . . . then" relations: *If* some attribute or connection (tie) is present, *then* some other attribute or tie is also present. The fact that such implications can be extended to three or more attributes, thus implies transitivity. Entailments extended to further attributes, ties, or nodes form entailment *chains* (partial orders) or hierarchies.

In the following sections of the chapter we briefly explain the theory and methods of entailment analysis and then illustrate its applications through an examination of citation networks among eighteenth-century chemists. We also draw some general conclusions about its applications to a range of social scientific problems.

Entailment analysis

Entailment analysis detects tendencies toward set-subset relationships among binary variables.[1] Given a set of variables, if some attribute (or some score on a dichotomized variable) logically implies (entails) the presence of some other attribute – or if *cases* with one attribute *(X)* are a subset of *cases* with another *(Y)* – then *X* implies *Y* ("If *X* then *Y*" is true). If the cases having *Y* as an attribute are also a subset of cases having *Z*, then *Y* implies *Z and,* quite interestingly, *X* also implies *Z*, since the set-subset relation is transitive. Formally: if *X* implies *Y,* and *Y* implies *Z*, then *X* implies *Z.* Transitive relations of this general kind may extend to any number of attributes. Other logical relations are also possible between two variables: for instance, the presence of one implying (being a subset of) the *non*-presence of another (*X* implies not-*Y*) or the absence of one implying the presence of a second (not-*X* implies *Y).* The relations are more complex among more than two variables.

In logic, set-subset relations are necessarily categorical: *All* of the cases with *X* also have *Y*. With real data, however, there may be some exceptions. That is, some cases may have *X* as an attribute, but not *Y*. The possibility of such exceptions to purely logical relations makes it necessary to examine empirical implications – that is, "material entailments" – statistically. We call an analysis that takes these factors into account *"material* entailment analysis." The consideration of higher-level entailments, those involving three or more attributes or ties, raises the problem of transitivity and complicates the statistical analysis since exceptions may cumulate among the chain of entailments.[2]

More formally, suppose we have *n* observations on *m* variables where observation N_i on variable M_j may take one of two values, *X*, or its complement *X*. The *frequencies* of observations with values *X; X* and *Y; X* and *Y* and *Z;* for example, are respectively designated *X*, *X*; *X.Y, X.Y, X.Y, X.Y; X.Y.Z*, and so forth. An implication is a statement of the form[3]

"If *X* then *Y*,"

which has exceptions designated *X.Y*. "If *X and Y then X*" and "If *X then Y or Z*" are the canonical forms of higher order and entailments, that is, they are standard forms or models to which other types can be reduced. Other forms, such as "If *X or Y then Z*," can be reduced to first-order entailments – in this case, "If *X* then *Z*" and "If *Y* then *Z*."

A system of entailments is a set of implications that meet the following criteria:

> **Criterion 1** (exceptions). For "If *X* then *Y*" to be true, but not its converse ("If *Y* then *X*"), there must be fewer exceptions to the former than to the latter. This criterion induces asymmetry (ordering) in the resulting system. In the case that the statements "*X* implies *Y*" and "*Y* implies *X*," both have the *same* number of exceptions, *X* and *Y* are considered equivalent (and they are set-theoretic equal if the number of exceptions is zero).

In addition to assuming the form "*X* implies *Y*," an entailment should give us some confidence that *X* and *Y* are *in fact* related, that is, that *X* has "relevance" (Salmon, 1971) for the existence of *Y*. This consideration leads to

> **Criterion 2** (material relevance of material correlation). For "If *X* then *Y*" and its converse to be true, *X* and *Y* must be positively correlated ($\varphi_{xy} > 0$, for the phi coefficient[4]).

Table 14.1, using fictitious data, provides us with an example of a proposition with few exceptions, but in which the two variables are statistically independent. There are only 5% (5/100) exceptions to the

Table 14.1. *Corporate size and centrality*

	Corporate size		
Centrality	Large	Small	Total
Low	95	9,405	9,500
High	5	495	500
Total	100	9,900	10,000
φ (= r) = ∞			

proposition that "large corporations have low centrality," and a miniscule 1% (5/500) to the (logically equivalent) proposition that "corporations with high centrality are small." However, being large does not entail having low centrality since the association between size and centrality is zero.

A material entailment thus not only requires relatively few exceptions, but also a positive correlation between variables.

> **Criterion 3** (transitivity). For the statements "If *X* then *Y*" and "If *Y* then *Z*" both to be true, the statement "If *X* then *Z*" must be true, and the partial correlation between *X* and *Z* given *Y* must be nonnegative ($\varphi_{xz.y} \geq 0$).[5]

The requirement of transitivity in a chain of entailments is motivated by the tendency for sets of observations defined by particular values to form *transitive subsets*. "If *X* then *Y*" and "If *Y* then *Z*" are propositions about the tendency for observations with value *X* to form a subset of those with value *Y*, and those with *Y* to form a subset of those with value *Z*. If there were no exceptions, we could infer from these relations that those with *X* form a subset of those with *Z*; so we require such transitivity in cases with exceptions as well.

Figure 14.1 shows a case in which there are no exceptions in a three-set chain: One set (with 3 elements) has attribute *X*; another (with 6 elements) has attribute *Y*; the third (with 9 elements) has attribute *Z*. The set {*X*} is contained in {*Y*}, {*Y*} is contained in {*Z*}, and it follows that {*X*} is contained in *{Z}*. The correlations between *X* and *Y* and *Y* and *Z* are equal, φ_{xy} = φ_{xz} = 1/√3. The correlation φ_{xy} =⅓. The partial correlation $\varphi_{xz.y} = 0$ is given by the formula:

$$\varphi_{xz.y} = \frac{\varphi_{xz} - (\varphi_{xy}\varphi_{yz})}{(1-\varphi_{xy}^2)(1-\varphi_{yz}^2)}$$

[formula corrected from original]

Thus condition (3) is satisfied.

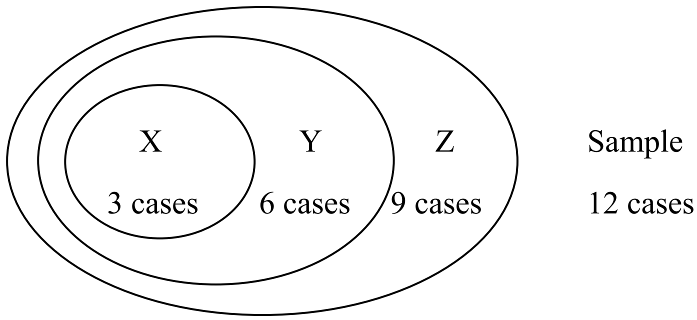

Figure 14.1. A three-set chain with no exceptions.

Condition (3) is *always* satisfied when set-subset relations or entailment chains contain zero exceptions. For those containing one or more exceptions, it is not necessarily satisfied, and potential entailments are regarded as valid in a system of entailments only if it is; that is, a tendency toward transitivity exists for all entailment chains in which the entailment is embedded.

These are the central criteria of entailment analysis. In addition, there are decision criteria for comparing a given *system* of observations to an expected distribution of entailments by levels of exception and degrees of correlation (under the null hypothesis) and for testing assumptions of no higher-order interactions.

> **Criterion 4** (rejection of total independence hypothesis). The probability that an observed entailment with a given level *of* correlation and exceptions will occur by chance under the assumption of total independence of the variables must be lower than some preset level for rejection.

This criterion is implemented through an application of signal detection theory (Coombs, Dawes, and Tversky, 1970) to the problem of whether observed entailments are likely or unlikely to be due to chance. This approach is explained later.

> **Criterion 5** (replication – lack of interaction). For entailments to be valid, interaction should not be present; that is, the measure of relevance (correlation coefficient) must not differ significantly across the categories of control variables.[6]

The significance of the difference can be expressed probabilistically by generalizing Fisher's exact test from the 2 x 2 case to the 2 x 2 x 2 case under the (null) hypothesis of no trivariate interaction with given bivariate distributions (White and Pesner, 1983). White, Pesner, and Reitz (1983) show how to derive a group significance for the hypothesis of no interaction for a system of binary variables.

Material and probabilistic entailments

The entailments we have been examining are material entailments associated with specific exceptions. These exceptions can also be expressed as *conditional probabilities* of exceptions to an entailment ("If X then Y") given either (a) the antecedent, $P(Y \mid X)$, or (b) the complement of the consequent, $P(X \mid Y)$.[7] The same pair *of* conditional probabilities will obtain whether we consider the entailment or its *contrapositive.*[8] To obtain a single such probability for the comparison of entailments, we take the larger value of this pair. That is, in Table 14.1 we take .05 (5/100) rather than .01 (5/500). Given this, we can now reformulate criteria 1-3 for *probabilistic entailments:*

1. For "If X then Y" (but not its converse, "If Y then X") to hold, there must be a lower probability *of* exceptions to the former than to the latter.

2. For "If X then Y" and/or "If Y then X" to hold

$d1 = \max [(P(Y \mid X) - P(Y \mid X)), (P(X \mid Y) - P(X \mid Y))]$ and/or
$d2 = \max [(P(X \mid Y) - P(X \mid Y)), (P(Y \mid X) - P(Y \mid X))]$

must be positive.

3. For both "If X then Y" and "If Y then Z" to hold, "If X then Z" must hold (by 1 and 2) and

$P(Z \mid X) \leq P(Z \mid Y)P(Y \mid X) + P(Z \mid Y)P(Y \mid X)$
$P(X \mid Z) \leq P(X \mid Y)P(Y \mid Z) + P(X \mid Y)P(Y \mid Z)$

must both be true.

These criteria are logically equivalent to the original criteria 1-3.

Several interesting consequences flow from these criteria. Most basic is that for an entailment to be valid, both antecedent and consequent must vary; that is, both column marginals and both row marginals in our basic 2 x 2 table must be greater than zero, so that we must work with variables (not constants). Second, a material entailment logically implies its contrapositive, which is important since in ordinary logic an implication is equivalent to its contrapositive. Third, we cannot obtain contradictory conclusions;[9] that is, if "X entails Y" is valid, then neither "X entails not-Y" nor "not-X entails Y" can be valid. Finally, two technical consequences follow that are important for the logical status of entailment analysis: "X entails X" (given that X varies), and "X entails Y" logically implies both "X entails X" and "Y entails Y."

Material or probabilistic entailment analysis, therefore, has the properties of a formal logic (noncontradiction, restricted identity, transitivity, contraposition, etc.) comparable to structures of logical entailment (Anderson and Belnap, 1975).[10]

Methods

Material entailment analysis of a system of binary variables is dependent upon two *constructive* procedures. One involves signal detection (Coombs et al., 1970): the comparison of the potential entailment relationships within an actual data set to a simulated (Monte Carlo) distribution of potential entailments.[11] This serves to separate those entailments that might occur randomly from those that ate not likely to occur by chance. Once entailments that are considered to be signal are determined, they are then *ordered* by level of exception and, within each exception level, by strength of correlational relevance.

The second procedure begins by accepting the strongest entailment (fewest exceptions, strongest relevance) and then adds successive entailments to the structure only if they satisfy criteria 1 through 5, including transitivity with respect to entailments previously admitted to the entailment structure. For example, if "*X* entails *Y*" is the strongest entailment, it is selected first. "*Y* entails *Z*" will be added to the structure – assuming it has passed the signal detection test (which implies that it passes criteria 1, 2, 4) – only if "*X* entails *Z*" also passes and all three entailments satisfy criteria 3 and 5 as a system. Furthermore, the investigator can place additional constraints on admissible entailments, such as a maximum (number or percentage) of exceptions or a minimum level for correlational relevance.

Representing entailment structures

There are three types of forms of entailment between two variables (sets of ties).

1. *Inclusion* refers to the inclusion of one set in another; that is, ties to *X* are a subset of (included in) ties to *Y* (cases with *X* are a subset of those with *Y*) or vice versa, or both: "If *X* then *Y*" or "If *Y* then *X*," or both. The inclusion with the *lesser* number of exceptions is called a "strong" inclusion and that with the greater number of exceptions is called a "weak" inclusion.
2. *Exclusion* refers to the exclusion of one set by another; that is, the presence of one attribute (set of ties) entails the absence of another: "If *X* then not-*Y*" or "If *Y* then not-*X*" (which are contrapositives and, therefore, equivalent).
3. *Coexhaustion* refers to the situation in which two ties (attributes) exhaust the possibilities; that is, if a case does not have a tie to *X* then it will have a tie to *Y*: "If not-*X* then *Y*" (equivalent to "If not-*Y* then *X*").

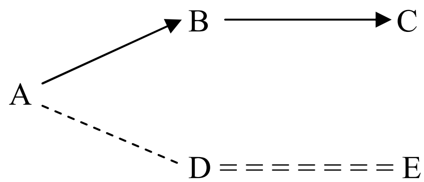

Figure 14.2. An entailogram.

Inclusion, exclusion, and coexhaustion are entailment relationships between antecedents and consequents that we can designate by different types of symbols for pictorial representation. Inclusion is an asymmetric relationship unless accompanied by its converse. Exclusion and coexhaustion are symmetric relationships:[12]

1. Inclusion *X* ⟶ *Y*
2. Exclusion *X* - - - *Y*
3. Coexhaustion *X*= = =*Y*

"Complementation" provides a number of ways of expressing the same entailment; and it is always possible to express an entailment by complementing the antecedent, the consequent, or both (the contrapositive). In general, either the three types of entailment or two of them plus complementation of some elements are necessary and sufficient to represent all possible entailment structures in a graphic form called an entailogram. Let us say that we have found that if people choose *A* they also choose *B;* that is, *A* entails *B,* and that *B* entails *C* (this implies that *A* also entails *C*). Also, those who choose *A* do not choose *D (A* entails not-*D*), and those who do not choose *D* choose *E* (not-*D* entails *E).* We can represent this pictorially (Figure 14.2).

An illustration: structure among citations of revolutionary chemists

The chemical revolution that we focus on here is a classical example of a "paradigm shift" (Kuhn, 1962; McCann, 1978). Kuhn argues that science may assume two forms: "normal" and "revolutionary." Normal science is based upon a *paradigm:* a matrix of accepted "beliefs, values, techniques," as well as concrete examples, which bind together a scientific community. The paradigm guides scientists in the community in choices of problems and methods and provides clear expectations about solutions. Normal science, thus, takes on the character of puzzle solving: data that apparently conflict with the reigning theory are not viewed as counterexamples, but as anomalies or puzzles that a scientist can solve (Kuhn, 1962: 5).

In contrast, a revolution, in which a new paradigm replaces an old one, occurs at times when such anomalies are incorrigible, and new or different explanations of some observed phenomena are advanced by one or more scientists. The result is a crisis in the community that can only be resolved by the defeat of one of these alternatives.

The issues raised during the chemical revolution centered around experiment and theories concerned primarily with the chemistry of gases ("airs") and the phenomena surrounding burning and other forms of what is now called oxidation. The existing paradigm, which had dominated for some 25 to 30 years in France and Great Britain (Guerlac, 1959; Rappaport, 1961; Schofield, 1970) and even longer in Germany, was based on the phlogiston theory of combustion and related phenomena. Bodies, phlogistonists thought, were combustible because, and to the extent that, they contained phlogiston, the matter of fire. They contended that when something burns, it gives up its phlogiston. Substances such as coal or oil were thought to be full of phlogiston; and hydrogen, when it was discovered, was thought by many phlogistonists to be this substance itself.

During the 1770s the discovery of "airs" led to the rapid growth of chemistry (McCann, 1978: chap. 3), intensive work on combustion (Perrin, 1969), and to widespread interest in it within fashionable circles in Paris and London. Thorough examination of the weight relations during combustion led Antoine Lavoisier to question the phlogiston explanation.[13] During the 1780s, Lavoisier developed a countertheory and undertook supporting experiments to a point where he was able to convert the leading chemists of France (including most of his colleagues in the Paris Academy of Sciences, the world's leading scientific institution at the time) to it. By the end of the 1790S, most British and European chemists had converted as well.

The context in which these changes were taking place becomes clear when we examine citations, that is, references that one scientist makes to another's work in his or her published writings. Citations reflect their purpose[14] and the structure of the disciplines in which scientists are embedded. We expect, for instance, that subject areas will hang together such that if one member of a specialty is cited, then others in the same group will also be cited.

These specialties may be thought of as information pools; Some members of a given pool will contribute more to it than others and, hence, will be cited more often by both those in the pool and those outside it. If, for instance, *A* and *B* are two members of a given pool, and *A* contributes more (is more "important"), then those who cite *B* should also tend to cite *A,* whereas those who cite *A* may not cite *B.*

Thus, specialties tend to have hierarchical structures; For one thing, more productive, prestigious, and visible scientists are likely to be cited more often. These scientists would then appear at the "top" of entailment chains, and those who are less productive, prestigious, and visible would only appear at the bottom. Thus, a scientist's location the bottom. Thus,

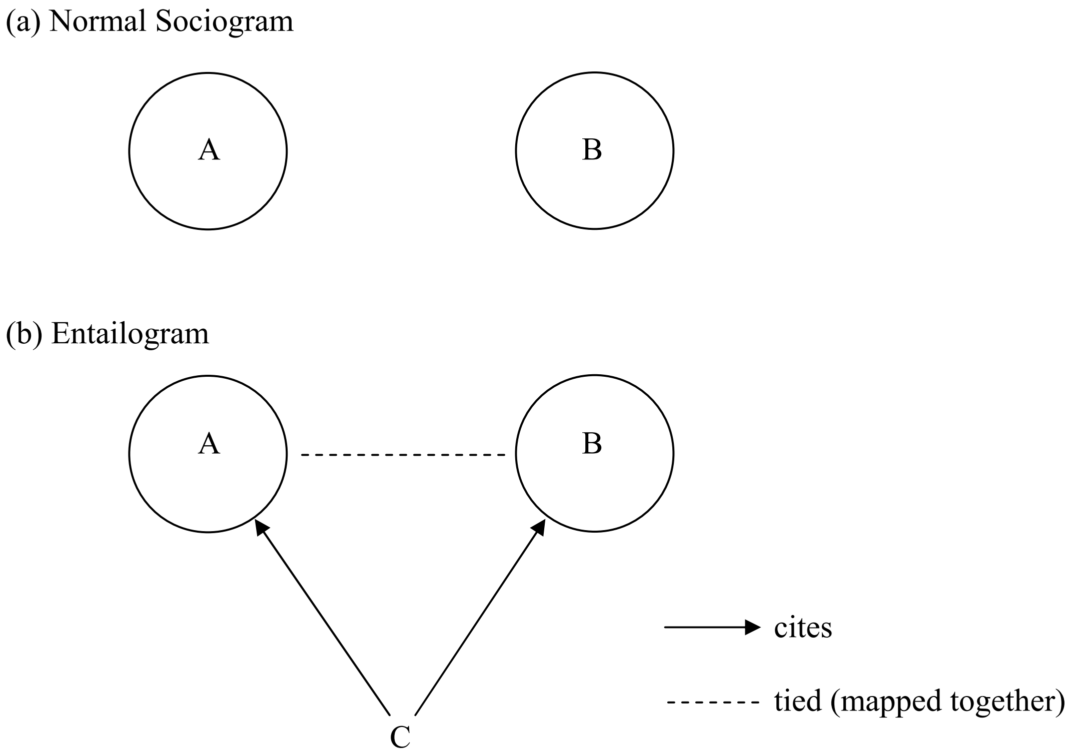

Figure 14.3. Comparison of a normal sociogram and an entailogram.

a scientist's location within an entailment chain of citations is a good indication of his role within the overall structure.

Because entailment analysis involves set-subset relations, if a given discipline or specialty is structured in this fashion (i.e., citing of one person leads to the citing of another), the citations to the first will form a subset of those to the second, and so on. Further, to the extent that the second is cited more often, his or her citations can be only a superset of those to the first. Therefore, entailment analysts view the data somewhat differently from the usual sociometric conception (Figure 14.3). A "link" exists between *A* and *B* not when *A* cites *B* but when one or more chemists who cite *A* also cite *B*. Since there may be more than one problem area in a discipline, we would expect more than one entailment chain to surface.

Data

The data we examine here come from a study by McCann (1978) and consist of *all* references to chemists in scientific papers published by British and French nationals in Great Britain and France from 1760 through 1795. There were a total of 3311 citations in 858 papers in which 219 authors referenced 591 chemists. Elimination of those not cited or not citing (including all scientists outside the sample owing to nationality) yields a structure containing 758 links (multiple citations ignored) among 115 authors.

In the eighteenth century, references rarely specified particular works (papers, books, talks) but merely mentioned names. Consequently all citations take the form "person *A* cites person B." Thus, our data reflect only the number of persons who cite a given person; that is, a citation from *A* to *B* exists if there is at least one paper in which *A* cites *B*.

For purposes of illustration we consider the sets composed of links to the 48 chemists most often cited between 1760 and 1795. We use the following cutoff values: (1) two exceptions – the first because we do not include self-citation, and the second to allow for random variation or noise; (2) φ (relevance) of .3 to eliminate weak relations (l or 2 exceptions but a small set); and (3) a reasonably conservative signal-to-noise ratio of 1 to 1 (White, 1980; White and McCann, 1981).

We consider all ties to the top 48 chemists among chemists publishing in each period. We first figure out the sets of citations – that is, the subset of citers that is tied to each top man. We then look for set-subset relations by considering each of the *[N(N - 1)/2]* possible pairs among the 48 and calculating both the percentage of exceptions and the correlation. For example, suppose chemists *A, B,* and *C* cite 1; *A, B, C*, and *D* cite 2; and *B, C*, and *D* cite 3. Then 1 entails 2 (since those who cite 1 are a subset of those who cite 2) and 3 entails 2 – both with no exceptions.

This case shows a convergence of two subgroups. A chain may also split if those who cite 1 are a subset of those who cite 2 and are also a subset of those who cite 3, but some of those who cite 3 (and do not cite I) do not cite 2.

Once all of the coefficients and exception levels have been calculated for each of the pairs, the entire system of exception levels and the coefficients are compared against a randomly generated set. Then we start with the strongest one and add those that meet all of the criteria, in particular, those that satisfy the transitivity criterion. For example, assume that 4 entails 5 with no exceptions and a correlation coefficient of 1.0, and 5 entails 8 with no exceptions and a correlation coefficient of 1.0. Then 4 must entail 8 (since there are no exceptions). If 8, in turn, entails 20 with 2 exceptions and a coefficient of .8, for example, then before 8 entails 20 can be added to the structure 4 and 5 must also entail 8 (pass the signal-detection test), and the appropriate partials must be positive (e.g., $\varphi_{4,8.5} \leq 0$).

Thus, in the following analysis, a chain leading up to Priestley, for instance, will mean that some group of chemists who cited a man lower in the chain also cited every person higher, with less than the cutoff level of exceptions, and that all of the partial correlations were positive. Priestley will not only be cited by everyone who cites anyone else in the chain (not counting exceptions), but he will also have the highest total number of citations.

Results

Figures 14.4 through 14.7 depict the entailment structure of the chemists' network over time. Each entailogram corresponds to an important phase in the chemical revolution. In general we see a picture of growth in structure: from almost none in the earliest years (Figure 14.4), to great complexity during the period of dispute and conversion (Figure 14.6) to a simpler pattern in the last, consolidation phase (14.7)). In detail, these entailograms both support and add important dimensions to the discussion in McCann (1978).

Figure 14.4 shows the entailment structure for 1760-71, a period during which chemistry was in its infancy and in which, consequently, there were relatively few journals or chemists. There is no evidence of an overlap in citations: Although 24 chemists cited one or more of the 48 leaders during these years (16 of 48), the only set overlap that met the criteria was the citing of Lassone and Montet by Cadet.

The period 1772-84 was the one in which the chemistry of airs became popular, and Lavoisier recognized various anomalies and then launched his attack on phlogiston theory. The entailogram (Figure 14.5) shows a distinct structure: a small number of generally short chains leading from less prominent, primarily young, chemists to recognized leaders in the community (Priestley, Black and Cavendish in England, the leaders in gas chemistry; and Baume, Lassone, and Lavoisier in France). Priestley clearly links two large groups whereas Lavoisier appears somewhat peripheral, a result that reflects the roles of Priestley as the primary exponent of the phlogiston theory – vocal, prolific, working in the "hot" area (airs) – and Lavoisier as a critic attacking the accepted model. We also note that, with the exception of Cavendish, the eminent men tend to head separate subgroups rather than being cited by identical followers (or outsiders).

In 1785 Lavoisier won his first major convert, Claude Berthollet, who was followed by Antoine Fourcroy in 1786 – both members of the Paris Academy of Sciences. By 1789 the battle had been won in France, and a new journal, devoted to the oxygen paradigm (the *Annales de Chimie),* was established. The entailogram for this period (Figure 14.6) exhibits the sort of structure one would expect: a complex and dense network with many chains, representing the great publishing activity resulting from paradigm dispute and the shifting alliances and arguments (McCann, 1978: chap. 3). The structure is complicated by the fact that both positive and negative citations are included,[15] so that Priestley, for example, appears at the top of both French and British phlogistonist and oxygenist chains. New French chemists, notably Berthollet, Morveau (coauthor of the new chemical nomenclature based on oxygen and cofounder of *Annales de Chimie),* and Macquer (older, author of the dominant phlogiston textbook, and defender of phlogiston), rise to the top, or close to the top, of entailment chains. As in the preceding period, we find that the most prominent men

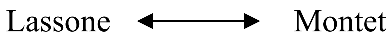

Figure 14.4. Entailment structure of network of chemists, 1760-71. $\varphi \geq .3$; 0 exceptions; signal-to-noise ratio, 1:1.

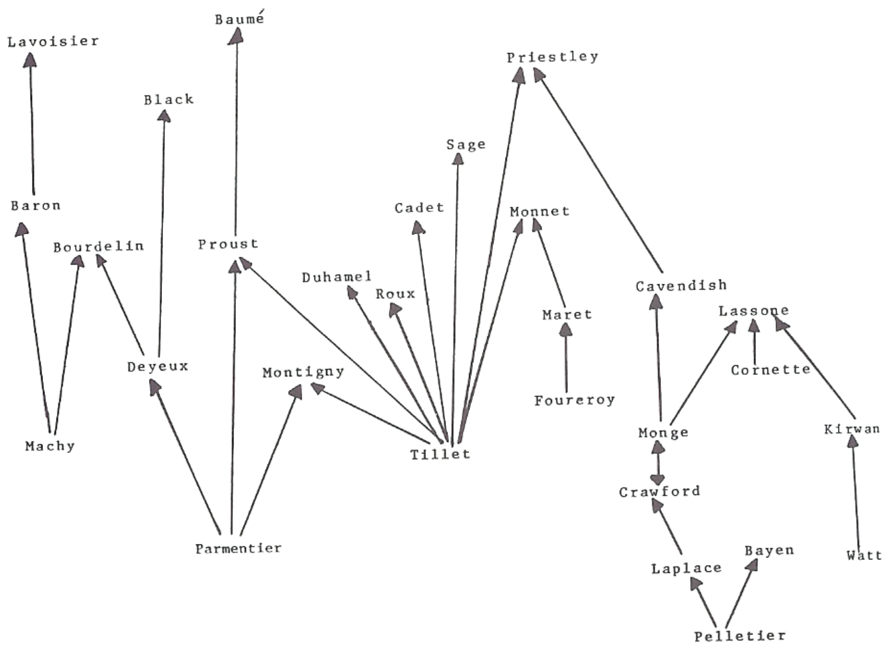

Figure 14.5. Entailment structure of network of chemists, 1772-84. $\varphi \leq .3$; exceptions ≤ 2; signal-to-noise ratio, 1:1.

tend to fall into distinct, although often highly connected, chains; that is, they specialize enough to remain the heads of slightly different subgroups. This separation may help to account for the tendency of disputants to talk past one another (Kuhn, 1962; Hufbauer, 1982) as well as reflect, in part, the tendency for paradigms under attack to fragment (Kuhn, 1962; McCann, 1978).

Finally, in the last entailogram (Figure 14.7) we see a period of consolidation in which the new paradigm has succeeded and the battle is over (at least in France). Almost all French chemists have converted or have stopped publishing: Only two men, one a mineralogist (Sage) and one an editor (Delametherie), published pro-phlogiston articles in France after 1790 (McCann, 1978: 81). This period exhibits a clear and relatively simple structure: a major chain leading to Lavoisier, the instigator of the revolution, with lesser and connected chains leading to other oxygen chemists.[16] The picture is one of normal

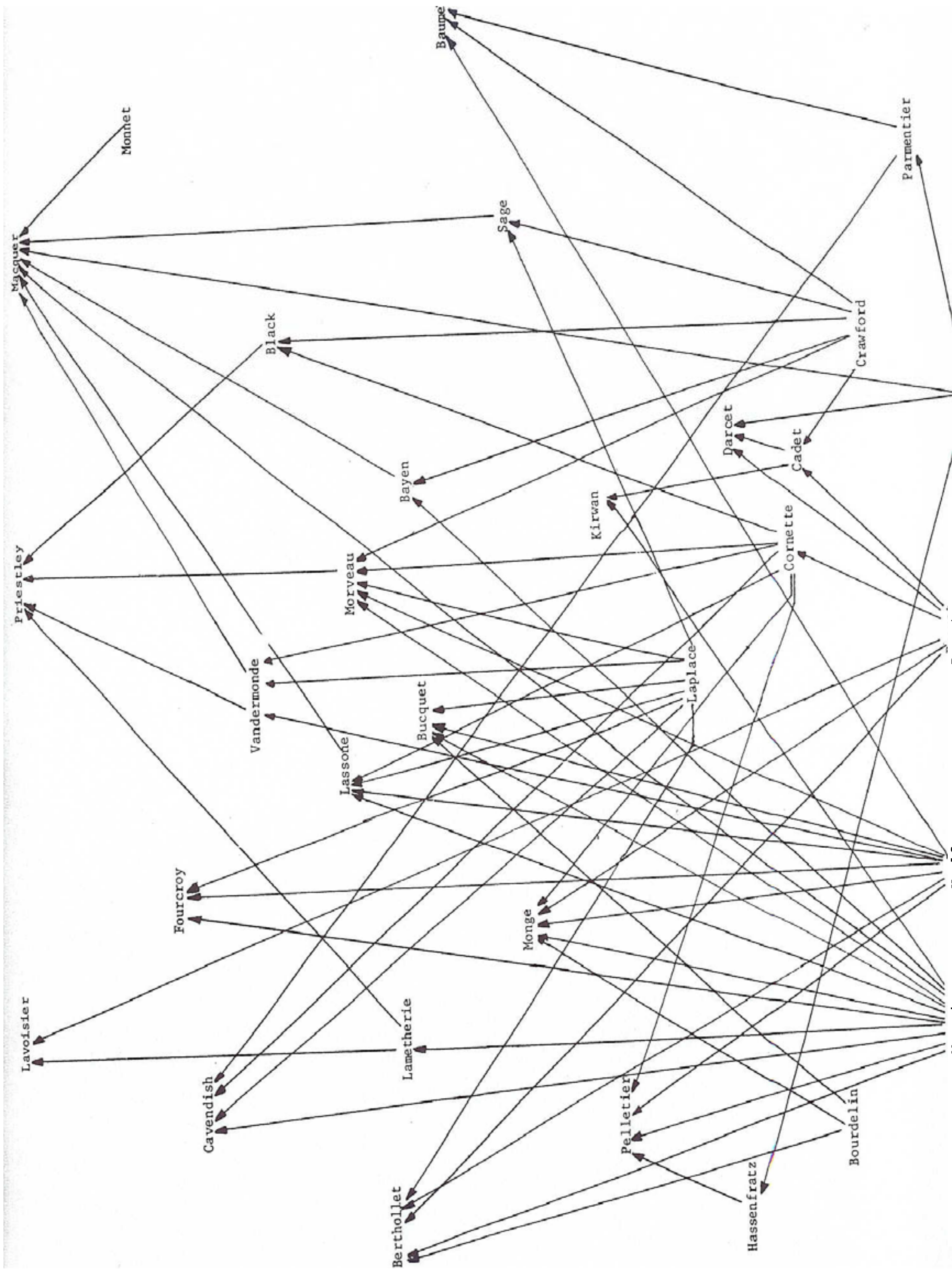

Figure 14.6: Entailment structure of network of chemists, 1785-8, ≥ .3; exceptions; signal-to-noise ratio, 1:1.

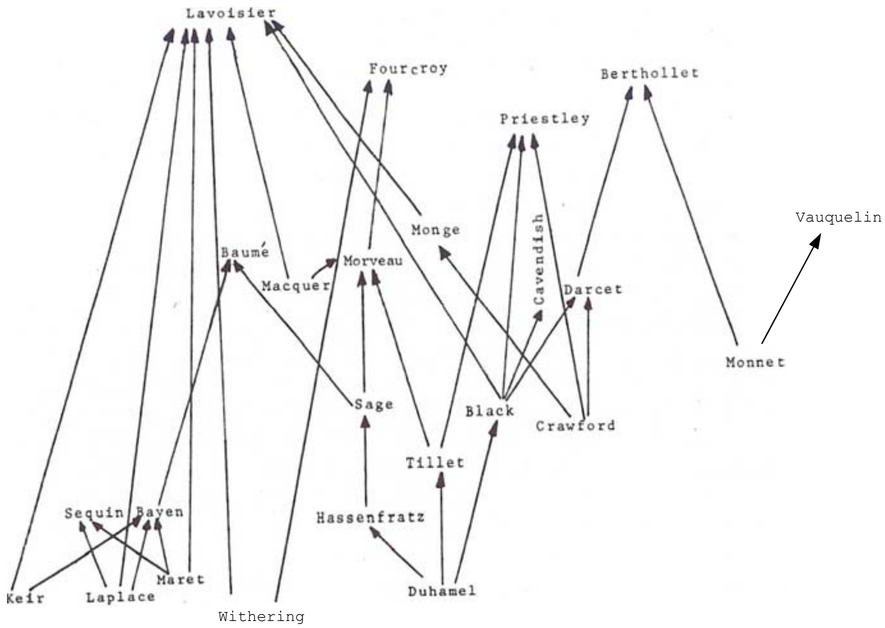

Figure 14.7. Entailment structure of network of chemists, 1791-5. $\varphi \geq .3$; 1 exception; signal-to-noise ratio, 1:1. [Note: Vertical placement determined by number of citations, horizontal by ease of reading (i.e., by "eyeball").]

science: the leading expert of the paradigm (Lavoisier) at the top, with other experts at the tops of their subspecialties.

In sum, the changes in structure over time show an early period with little coherence,[17] followed by a period in which clear subareas appear and experts take their positions at the heads of these subareas and in which a young Turk begins the fragmentation process with an attack on the guiding paradigm. During the period of most intense conflict, the structure is highly interconnected, exhibiting overall coherence, but with internal divisions reflecting various camps in the dispute; that is, both the prime revolutionary and the major defenders of the faith take their positions at the top. Finally, as the conflict subsides and the research questions guided by the new paradigm come to the fore, the structure becomes simpler, more orderly, and closer to a single paradigmatic hierarchy, with the leading revolutionary at its head.

The analysis refines Kuhn's (1962, 1970) broad description of the process of paradigmatic change: breakdown in a given paradigm, a period of crisis and conflict during which the supporters of each view contend with one another, and the triumph of the new paradigm. It elaborates his view that revolutions will be reflected in shifting distributions of cited literature (1962: ix), which themselves reflect changes in the formal and informal communication networks of scientific communities (1970: 178) by providing details of the structure of the communities and revealing changes in the structures that accompany revolutionary change. Consonant with previous literature on science, the analysis shows a center-periphery pattern (McCann, 1974; Burt, 1978), or what has been called an "invisible college" (Crane, 1972), with the experts who are most often cited sitting at the heads of entailment chains. An interesting refinement, however, is the discovery that the most prominent scientists do not appear in a single chain or as structurally equivalent (in the sense of being linked to the same others), but each seems to carve out his own niche.

In addition we get a dynamic picture of the changing roles of important scientists as the process of revolution unfolds. Older scientists fall away and newer ones take their place, a process that is, of course, inevitable but that is undoubtedly speeded up by a revolution, during which role players rapidly change.

Conclusion

This chapter has described a new technique that allows researchers to model concrete situations in which the degree of transitivity present within social structures may be investigated empirically. As we have seen, the transitivity problem is a special case of the more general issue of orders of variables or attributes. Entailment analysis represents a very general approach that simultaneously examines both local and global orderings present within relational and attribute data. Entailment structures reflect "If ... then" relations that can be extended to three or more attributes; the fact that they can be expended implies transitivity.

We have discussed and illustrated the use of five criteria to examine the empirical relations in a social network: number of exceptions, relevance, transitivity, independence, and interaction. The number of exceptions in set-subset relations and the relevance of one set for another were used to define local structure. The criteria of transitivity, independence, and lack of interaction were used to determine global structure. The use of these criteria, with varying values or cutoffs, permits us to look at the global and the local structure simultaneously and to discover the empirical connections in the network in a subtle, rather than heavy-handed, manner. The exceptions to entailments can also be expressed as conditional probabilities, and as a result, the first three criteria can be viewed probabilistically.

We then described a means of representing these entailment structures in a diagram, the entailogram. Entailograms were constructed for four periods representing a paradigmatic shift from normal to revolutionary and then again to normal chemistry: what is known as the chemical revolution of the eighteenth century. This revolution, as we have seen, was primarily concerned with the chemistry of gases and the phenomena now known as oxidation.

We found that the entailment structure of citation groups during the chemical revolution revealed new information that extended the historical and sociological investigation carried out by McCann (1978). The new analysis presented here revealed details of the process of structural development of a scientific specialty and the varying roles played by members of its community during a period of revolutionary change. It further showed that entailment analysis, while preserving a global center-periphery pattern (McCann, 1974: chap. 7; Crane, 1972: chap. 3) was also able to depict the internal structure of specialty groups in fine detail.

From a theoretical standpoint material entailment analysis, because of its set-subset orientation, may force us to reconceptualize the way networks and similar phenomena have traditionally been represented. Further, entailment analysis focuses on relations and demands transitivity, and consequently is able to make fine separations and exhibit both orders and partial orders. Thus, it incorporates hierarchical and quasi-hierarchical structures that may be implicit in the data.

More generally, material entailment analysis can be used to complement other structural analytic techniques. It provides, for instance, a fine-grained image of potential roles played by actors in a concrete system that could be used in constructing more precise blockmodels. Similarly, the structure represented by an entailogram might be clustered or the entailogram itself could be superimposed upon a clustering (White, 1981), providing internal structure to clusters. Other methods of network analysis, such as multidimensional scaling or cliquing, may also be profitably supplemented in this fashion.

In sum, material entailment analysis promises to add to our "bag" of methodological equipment and to enrich substantive thinking. It is useful for detecting ordering or transitivity in both relational and attribute data and may therefore help integrate the two.

NOTES

1. For a discussion and example of nonbinary variables, see Burton et al. (1977).
2. The transitivity problem for material entailment analysis may be stated as follows: If we allow a given number (level) of exceptions to an entailment (i.e., if we assume the validity of an entailment although there are (some small number of) exceptions to it), then if *X* entails *Y* and *Y* entails *Z*, it may not be the case that *X* entails *Z* because the exceptions to *Y* entails *Z* may be different from those to *X* entails *Y* so that their sum, which is the number of exceptions to *X* entails *Z*, may exceed the allowed cutoff level. In other words, the exceptions may cumulate along the chain of entailments; and al-

though transitivity always holds for purely logical entailments, it may not hold for material (empirical) ones.

3. [The original of the sentence in the text has been corrected] The element "*X*" may be replaced by a *conjunctive* series (e.g., *A and B and C*) and the element "*Y*" may be replaced by a *disjunctive* series (e.g., *D or E or F).* It is sometimes necessary for convenience of discussion or representation to refer to the complement of an attribute or to look at contrapositives. To form the contrapositive of an implication (which is always equivalent to the given implication and has the same exceptions), reverse the antecedent and consequent, interchange "ands" and "ors," and complement all elementary terms. For example, the contrapositive of "If *X* then *Y*" is "If not-*Y*, then not-*X*."
4. Although we use the phi coefficient in this discussion because of its symmetric character and its relation to conditional probabilities, other coefficients can obviously be used.
5. In longer chains, such as "if *W* then *X* then *Y* then *Z*," the fact that all first-order partials ($\varphi_{WY.X}$, $\varphi_{WZ.X}$, $\varphi_{WZ.Y}$, etc.) are positive does not imply that second-order partials will be positive. This may be required in stronger tests of transitivity.
6. Failure to satisfy the criterion of no interaction is not necessarily a serious problem in entailment analysis. If only first-order entailments (two place: "If *X* then *Y*," for example) are examined where higher order interactions are present, the common effect is not to create spurious entailments, but simply to miss the higher-order ones. However, higher-order entailments may be theoretically important.
7. This notation is a standard way of stating conditional probabilities, where "*X*" is the presence of value (event), *X*, as above, and "*X*" represents the absence of *X* (nonoccurrence of *X).* Thus, P(*Y* | *X*) is to be read as "the probability of not-*Y* (or of not observing *Y*) *given X,"* which is equivalent to the probability of an exception, *Y*, to the entailment "If *X* then *Y*."
8. Consider "*X* entails *Y*." *X* is called the antecedent (because it logically comes before *Y*) and *Y* the consequent. [Again,] the contrapositive of an entailment is obtained by negating (taking the complement of) both the antecedent and the consequent and interchanging them. For example, the contrapositive of "*X* entails *Y*" is "not-*Y* entails not-*X*" and the contrapositive of "*X* entails not-*Y*" is "*Y* entails not-*X*."
9. The first and third consequence (theorems 1 and 3) rule out the type of possible contradiction where "If *X* then *Y*" and "If *X* then not-*Y*" both could occur with zero exceptions due to the nonoccurrence of *X* (*X* = 0). Similarly, "If *X* then *Y*" and "If not-*X* then *Y*" could otherwise occur owing to the universal presence of *Y* (*Y* = *N).* These kinds of contradictions are standard in ordinary implicational logic, where the falsity of an antecedent implies the truth of contradictory consequences and the truth of the consequent implies the truth of propositions with contradictory antecedents (Anderson and Belnap, 1975).
10. The strengthening of material implication by the criterion of relevance directly parallels Anderson and Belnap's (1975) restriction of logical entailment to a stronger case of logical implication with the added criterion of logical relevance (derivation by proof). The axioms of logical entailment are identity, transitivity, restricted assertion, and self-distribution (Anderson and Belnap, 1975: 24).
11. [In principle, the hypergeometric distribution could be used to compute expected distributions.] The decision to accept or reject a given entailment is based on a signal-to-noise ratio used in this comparison. For details, see White and McCann, 1981.

12. There are further logical combinations possible, some of which may be useful in simplifying the presentation of complex systems.
13. There were a very few others (e.g., Bayen, an apothecary with the army, and Turgot) who raised doubts about phlogiston, but only Lavoisier created an alternative.
14. We realize that reasons for citations are diverse and that their exact import is problematic. Scientists presumably cite one another to acknowledge influence or ideas, and they also use citations to indicate knowledge of a field. For these reasons citations usually lead to recognition and prestige for those cited (Hagstrom, 1965; Blume and Sinclair, 1973; Gustin, 1973). Since citations lead to recognition, scientists may also cite their friends or others they wish to promote (or denigrate in the case of negative citations). The place of friendship ties in science is little studied, although some work has been done by net workers on networkers (Freeman and Freeman, 1979, 1980). Nevertheless, most theory and findings support the view that scientists tend to cite disproportionately the leaders of a field or specialty.
15. Negative citations, those in which a scientist disagrees with the person being cited, are almost entirely ignored in the literature. Their recognition and coding might clarify and extend results of citation analyses - the blind counting of citations without taking into account their valence may easily bias interpretations. For the case here, we note that Priestley, the leader of the phlogistonists, received more citations from oxygen chemists than did Lavoisier. Lavoisier and Priestley also, not surprisingly, cited each other frequently.
16. Among these other oxygen chemists were Fourcroy and Berthollet, the most eminent, who dominate French chemistry after Lavoisier's death by guillotine in 1793, and Baume, an older man and a late convert. Another chain leads (through phlogiston chemists only) to Priestley.
17. The lack of structure is not due to the lack of a paradigm. However, there were many reactions for which phlogiston was not relevant, and chemistry as a profession was only weakly institutionalized (McCann, 1978: chap. 3).

LITERATURE CITED

Anderson, A. R., and N. D. Belnap. *Entailment: The Logic of Relevance and Necessity.* Princeton, N. J.: Princeton University Press, 1975.

Baker, F. B., and L. Hubert. "Inference Procedures for Ordering Theory." *Journal of Educational Statistics* 2 (1977): 217-33.

Bart, W. M., and P. W. Airasian. "Determination of the Ordering among Seven Piagetian Tasks by an Ordering Theoretic Method." *Journal of Educational Psychology* 55 (1974): 277-84.

Bart, W. M., and D. J. Krus "An Ordering-Theoretic Method to Determine Hierarchies Among Items." *Educational and Psychological Measurement* 33 (1973): 291-300.

Berkowitz, S. D. *An Introduction to Structural Analysis.* Toronto: Butterworths, 1982.

Berkowitz, S. D., P. J. Carrington, Y. Kotowitz, and L. Waverman. "The Determination of Enterprise Groupings through Combined Ownership and Directorship Ties." *Social Networks* 1 (1978): 291-413.

Berlin, B., and P. Kay. *Basic Color Terms: Their Universality and Evolution.* Berkeley: University of California Press, 1969.

Blume, S. S., and R. Sinclair. "Chemists in British Universities: A Study in the Reward System of Science." *American Sociological Review* 38 (1973): 126-38.

Burt, R. "Stratification and Prestige among Elite Experts in Methodological and Mathematical Sociology Circa 1975." *Social Networks* 1 (1978): 105-58.

Burton, M., L. Brudner, and D. White. "A Model of the Sexual Division of Labor." *American Ethnologist* 4 (1977): 227-51.

Coombs, C. H., R. M. Dawes, and A. Tversky. *Mathematical Psychology: An Elementary Introduction.* Englewood Cliffs, N. J.: Prentice-Hall, 1970.

Crane, D. *Invisible Colleges: Diffusion of Knowledge in Scientific Communities.* Chicago: University of Chicago Press, 1972.

D'Andrade, R. G. "A Propositional Analysis of U.S. American Beliefs about Disease." In K. H. Basso and H. A. Selby (eds.), *Meaning in Anthropology.* Albuquerque: University of New Mexico Press, 1976.

Freeman, L., and S. Freeman. "A Semi-Visible College: Structural Effects on a Social Networks Group." In M. Henderson and M. J. McNaughton (eds.), *Electronic Communication: Technology and Impacts.* Boulder: Westview Press, 1980.

Freeman, S., and L. Freeman. "The Networkers' Network: A Study of the Impact of a New Communications Medium on Sociometric Structure." University of California, School of Social Sciences, Research Reports 46. Irvine, 1979.

Gagne, R. M. *The Conditions of Learning.* New York: Holt, Rinehart, and Winston, 1965.

Greenberg, J. H. *Language Universals, with Special Reference to Feature Hierarchies.* The Hague: Mouton, 1966.

Guerlac, H. "Some French Antecedents of the Chemical Revolution." *Chymia 5* (1959): 73-112.

Gustin, B. H. "Charisma, Recognition, and the Motivation of Scientists." *American Journal of Sociology* 78 (1973): 1119-34.

Hagstrom, W. O. *The Scientific Community.* New York: Basic Books, 1965.

Hufbauer, K. G. *The Formation of the German Chemical Community.* Berkeley: University of California Press, 1982.

Johnsen, E., and H. G. McCann. "Acyclic Triplets and Social Structure in Complete Signed Digraphs." *Social Networks* 4 (1982): 251-72.

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

The Structure of Scientific Revolutions. 2d ed. Chicago: University of Chicago Press, 1970.

Lorrain, F., and H. C. White. "Structural Equivalence of Individuals in Social Networks." *Journal of Mathematical Sociology* 1 (1971): 49-80.

McCann, H. G. *The Development and Reception of the Chemical Revolution by the Chemical Communities of France and Great Britain,* 1760-1795. Ph.D. diss., Princeton University, 1974.

Chemistry Transformed: The Paradigmatic Shift from Phlogiston to Oxygen. Norwood, N. J.: Ablex, 1978.

Nadel, S. F. *The Theory of Social Structure.* London: Cohen and West, 1957.
Perrin, C. E. "Prelude to Lavoisier's Theory of Calcination: Some Observations on *Mercurius Calcinatus Per Se." Ambix* 16 (1969): 140-51.
Rappaport, R. "Rouelle and Stahl- The Phlogistic Revolution in France." *Chymia* 7 (1961): 73-102.
Salmon, W. C. *Statistical Explanation and Statistical Relevance.* Pittsburgh: University of Pittsburgh Press, 1971.
Schofield, R. E. *Mechanism and Materialism: British Natural Philosophy in an Age of Reason.* Princeton, N.J.: Princeton University Press, 1970.
White, D. R. "Material Entailment Analysis: Theory and Illustrations." University of California, School of Social Sciences, Research Reports 15. Irvine, 1980.
"American Beliefs About Disease: Material Entailment Analysis of D'Andrade's Data." University of California, School of Social Sciences working paper. Irvine, 1981.
White, D. R., and H. G. McCann. "Material and Probabilistic Entailment Analysis: Methods for Multivariate Analysis." University of California, School of Social Sciences working paper. Irvine, 1981.
White, D. R., and R. Pesner. "Internal Replication, the Systems Concept, and Sources of Validity in Nonexperimental Research." *Behavior Science Research* (1983): 26-44.
White, D. R., R. Pesner, and K. Reitz. "An Exact Significance Test for Three-way Interaction." *Behavior Science Research* 18 (1983): 103-22.