More Telltale Signs: What Attention to Representation Reveals about Scientific Explanation

Andrea I. Woody†‡

This essay explores the connection between representation and explanation in the sciences. I suggest that scientific representation schemes be viewed as pragmatic tools for acquiring the sort of articulated awareness that is the hallmark of nontrivial knowledge. Crystal field theory in chemistry illustrates this perspective. Certain representations achieve the status of being paradigmatically explanatory, thereby shaping models of intelligibility. In turn, these explanatory preferences serve largely to define and differentiate disciplinary communities by implicitly endorsing particular epistemic aims and values. In this way, the pragmatic nature of explanatory discourse effectively grants its intellectual utility.

---

1. Starting Out. Here is a cluster of claims concerning explanation and explanatory power:

> **C1**. Explanatory power is, ceteris paribus, a theoretical virtue. Other things being equal, we are justified in preferring more explanatory theories to those that are less so.
>
> **C2**. There is no general consensus about what constitutes adequacy or success conditions for particular explanations. Nor is there a clear conception of how explanatory power is linked to the individual explanations a theory provides.
>
> **C3**. The theoretical systems that scientists themselves judge to be highly explanatory include not only a wide array of conceptual and logical structures but also a diverse set of representational schemes

---

†To contact the author, please write to: Department of Philosophy, Box 353350, University of Washington, Seattle, WA 98195; e-mail: awoody@u.washington.edu.

‡Thanks to Paul Teller, Steve Weininger, and Theodore Brown for constructive comments and discussion.

Philosophy of Science, 71 (December 2004) pp. 780–793. 0031-8248/2004/7105-0011$10.00

> and formats. Alternative representations of the same information, moreover, are often granted different explanatory status.

I consider each of these claims to be defensible, even plausible. Much less obvious is how we can make sense of the set as a whole. How do these claims relate to one another? What might we come to understand about scientific explanation if we could tell a coherent story about their interrelations?

The primary contention of this essay is that the relationship between explanation and representation in science is far more intimate than existing philosophical discussion would suggest. Ignoring the connection may hinder efforts to understand the vital role that explanatory discourse plays in the sciences, and in intellectual pursuits generally.

2. Representation and Knowledge. Issues surrounding representation in science are almost always issues of representational schemes. By this I mean something along the lines of Haugeland's (1998, 172) notion of a scheme in which a broad set of possible contents can be represented by a corresponding set of representations, themselves governed by unified sets of rules for both generation and application. Frequently, scientific representation schemes are mathematical, and their rule-based structure is explicit. Still, learning to use the representations skillfully requires instruction, training, and experience. After all, many of the rules for manipulation of mathematical structures are constraints on what is permitted. Recognizing all the moves that are forbidden is a far cry from having the capacity to make judicious choices regarding the moves that are permitted; this requires skill. Logic students must practice the construction of proofs, and no one should be surprised that, in spite of knowing the rules, I am a terrible chess player.

Likewise, when we speak of someone being a knowledgeable person, we do not mean that she can spew a barrage of facts. Someone knowledgeable is not a walking game of Trivial Pursuit. Rather, the knowledgeable individual is one who impresses us by the way her conversation displays a grasp of what is, at once, relevant and nontrivial, making connections that were not obvious before she makes them but which seem both insightful and plausible afterwards. Her talk displays, in essence, a certain *awareness*.

When claiming that science is a knowledge-seeking enterprise, this is the sort of thing I intend. Scientific disciplines strive to be aware of the structure of various empirical domains. The aim is not merely to recognize and accurately describe such structure, but to use it skillfully. We want to understand our world: to predict it, sometimes to control it, to explain it, often to change it, at least to live in it with a modicum of comfort.

But this conception of knowledge includes as much "know how" as "knowing that." Awareness of the connections between claims becomes as crucial as recognizing and justifying the claims themselves. And here, the appropriate sense of relation is one that outstrips (at least elementary) logical structure because it must incorporate notions we could label "relevance" or "significance" or "priority."

It should not be surprising then that scientific representational practices typically operate through coordinated representation schemes. Such schemes aim to focus and guide reasoning among numerous and diverse individual claims. They provide structure to the inferential landscape, enabling the otherwise daunting task of mapping some domain of knowledge. (Here I am using "inference" in a broad, nontechnical sense.) The metaphors of mapping and chess playing are each instructive. Knowledge of a landscape requires more than a map and the ability to read it. The knowledgeable soul navigates the challenges of rush-hour cross-town journeys with finesse, just as the master chess player devises strategies sensitive to the many contingencies of his opponent's moves. Likewise, the ultimate aim of most representational practices is to achieve *articulated awareness* of the nature of the objects and relations constituting that particular domain.

Such awareness encompasses recognition of the possibility space delineated by fundamental theoretical structures (the "laws" on the books) alongside strategies for undercutting the tremendous complexity of connections. Being knowledgeable often consists in knowing what to ignore. Yet to ignore something, strictly speaking, presupposes some awareness of it; ignoring is substantially different from being ignorant. What *should* be ignored, furthermore, is determined primarily by context. Sensitivities must be tuned to the relevant goals of a given knowledge-seeking enterprise. Local adequacy (or success) conditions for representational schemes necessarily are contingent upon these aims. Representation is a thoroughly pragmatic affair.

Let me discuss briefly an example from inorganic chemistry to demonstrate these sorts of representational capacities. The details of the case are of course unique and particular to this corner of chemistry, but the inferential awareness I hope to highlight seems to me a vital component of most, if not all, modern theoretical knowledge. Thus, the example is intended to display the sort of intellectual grip afforded by the representation scheme at the same time it recommends that the skills and awareness necessary to obtain such a capacity be recognized as essential constituents of substantial knowledge structures.

3. An Example: Crystal Field Theory. "Crystal Field Theory" (CFT) is a modeling strategy within coordination chemistry (a subfield of inorganic

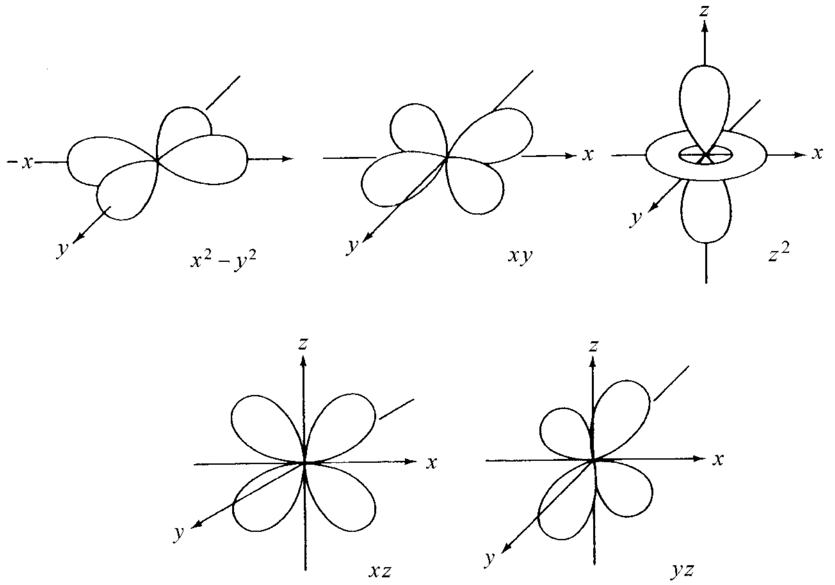

Figure 1. Spatial arrangement of the five *d*-orbitals.

chemistry) that has enjoyed great explanatory success in the last half century. (For a detailed introduction, see Huheey 1983.) I'll point out what I take to be telling aspects of this representational scheme to illustrate some general points regarding scientific representation (rather than to supply evidence for these points, which would require more elaborate argumentation).

First, an outline of the scheme: Crystal Field Theory in its pure form assumes the only interaction between a metal ion and its ligands is electrostatic; bonding has purely ionic, as opposed to covalent, character. In isolation, the five *d*-orbitals of a gaseous metal ion will be energetically degenerate. If a spherically symmetric field of negative charge was placed around the ion, the *d*-orbital energies would be raised with degeneracy maintained. We conceptualize ligand interactions with the metal ion through comparison to this spherically symmetric field. The number of ligands for most complexes will be either four or six. Let us consider the case of six, in which minimal electrostatic repulsion requires an octahedral configuration of ligands—effectively, the best possible approximation to the symmetric field. Assume the ligands approach along the x, y, and z axes, one ligand approaching from either direction of each axis. By representing the spatial arrangement of the *d*-orbitals of the metal ion in the standard way ($x^2 - y^2$, xy, z^2, xz, and yz; see Figure 1), simple geometric

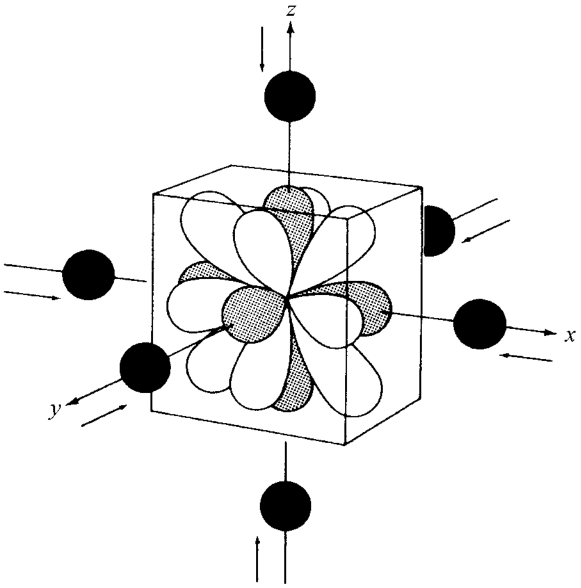

Figure 2. Complete set of *d*-orbitals in an octahedral ligand field.

reasoning allows us to conclude that two of these orbitals will experience, as a result of ligand interaction, considerably more electrostatic repulsion than the others (see Figure 2). The magnitude of the energy separation between the two sets of *d*-orbitals is fixed by convention to be 10*Dq*.[1]

We now envision metal-ligand bonding as a two-step process. First, hypothetically, the ligands approach while maintaining spherical charge symmetry. They then shift into the octahedral configuration. The total increase in the *d*-orbital energy of the metal ion occurs in the first step of this hypothetical process. The energy that would be associated with the orbitals under this circumstance determines the "barycenter." The shift to the octahedral configuration redistributes this energy among the orbitals, producing two distinct levels while maintaining the average energy of the set—that is, the barycenter. Representing the process as two steps allows one to infer the resultant energy levels of each orbital in relation to the barycenter. Moving from the spherical to octahedral configuration raises the energy of two orbitals while it lowers the energy of the remaining three. Accordingly, the two destabilized orbitals are each raised by 6*Dq* and the three stabilized orbitals are lowered by 4*Dq* (see Figure 3). One

[1] The empirical value of the quantity measured by 10*Dq* is of the same order of magnitude as a typical single chemical bond (~50–100 kcal/mol).

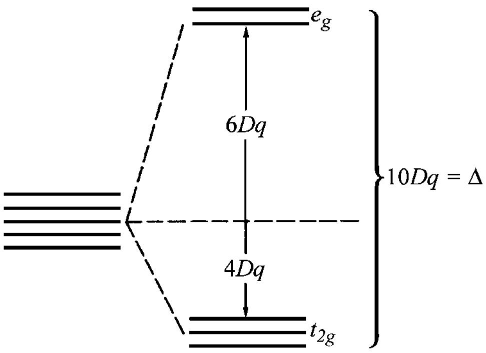

Figure 3. Energy-level diagram showing splitting of the *d*-orbital degeneracy by an octahedral ligand field.

further elaboration should be mentioned. When a large number of electrons occupy these orbitals, some electrons are forced to pair, raising the total electronic energy. A comparison of the magnitude of the orbital-splitting energy, $10Dq$, with the pairing energy, P, is required to predict the order of orbital occupation. If $10Dq > P$, we should expect one result and if $P > 10Dq$, we should expect another (see Figure 4). In either case, the resulting energy-level diagrams can be used to account for, or predict, a wide range of chemical behavior including bonding tendencies and reactivity.

Three aspects of this representational scheme seem particularly noteworthy. First, the underlying conceptual framework is shamelessly idealized and hypothetical. There is no reason to believe that coordination complex metals approximate purely ionic bonding generally; indeed, there is clear reasoning to the contrary. Certain metal complexes do have highly ionic character, admittedly, but particular instances alone cannot substantiate the *general* application of the CFT conceptual apparatus. In addition, the notion of a ligand approach that preserves spherical symmetry requires a setting aside of both logical and physical possibilities. Second, the limited quantitative structure of the representation scheme is achieved by an articulation of the hypothetical model. Introducing the "barycenter" concept produces quantitative measures, but ones far simpler than any derived from either empirical measurements (spectral analysis) or theoretical calculations (Schrödinger equation calculations). Third, the information derived from the hypothetical model is represented by a two-dimensional diagram of orbital energy. It is significant that energy is the only system property explicitly represented, and the property's quantitative structure is effectively reduced to an ordering relation.

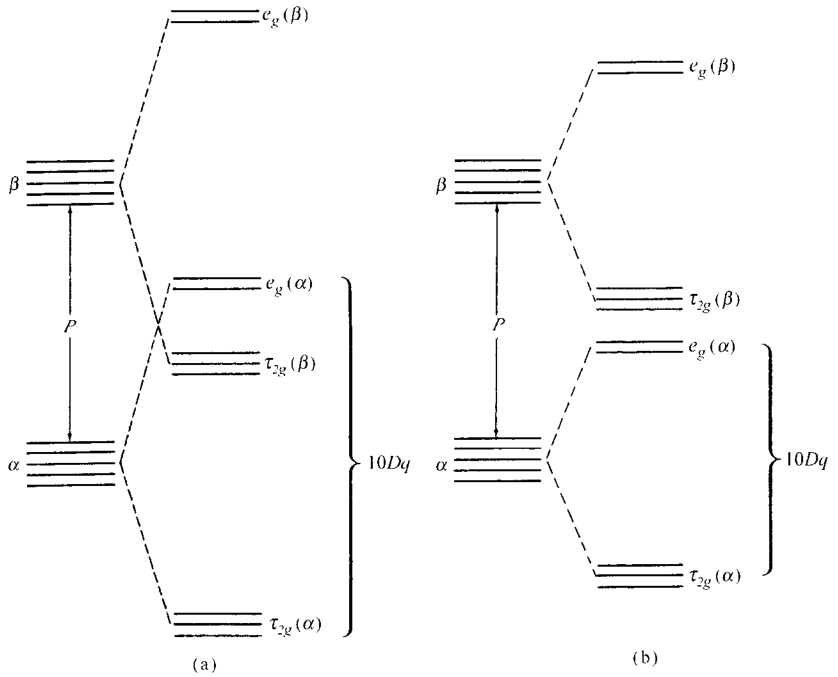

Figure 4. Comparison of (*a*) strong and (*b*) weak ligand field splittings. (Here each line represents a complete one-electron wave function, including spin.)

If we restrict attention solely to the relation between the representations and the actual coordination complexes they are intended to represent, the CFT scheme appears highly inaccurate, perhaps blatantly false, and the diagrammatic format seems arbitrary, or at least inconsequential. But as I have stressed for all representations, CFT is a pragmatic device. Its value is determined by the scheme's ability to support a specific, bounded set of inferences. Thus, CFT can only be justified, and indeed is only intelligible, within the context of a scientific practice that aims to account for and predict particular properties of metal ion complexes. We can appreciate the bold creativity of CFT only in relation to the pragmatic aims underlying its adoption and use.

By the time CFT was introduced, modern chemistry had recognized electron energy as perhaps the most informative indicator of chemical behavior. This recognition was a direct result of exploring the relevance of quantum mechanics to chemistry via the Schrödinger equation. CFT cultivates selective attention to the energetics of metal-ligand interactions by judiciously employing hypothetical reasoning. The barycenter concept both introduces and sufficiently constrains quantitative structure, while diagrammatic representation eliminates the need for explicit calculation and exploits the reliability of visual perceptions for determining the relative ordering of orbital energies—information essential for the sorts of inferences CFT typically supports. The upshot is that CFT representation achieves intellectual grip and inferential utility by emphasizing certain significant factors in actual phenomena while suppressing, or ignoring, others. Balancing the trade-offs between generalization and descriptive accuracy is always necessary, but, as they say, the proof is in the pudding.

One additional point should be made. CFT is not merely an abstract inferential device that works in some circumstances and not in others. It is part of the mapping of an entire domain of chemical knowledge. In fact, CFT can be embedded within a more articulate, comprehensive, and theoretically principled "molecular orbital" representation of metal complexes that maintains the diagrammatic format. (Huheey 1983, 420 contains a nicely integrated diagram with informative discussion.) This embedding relation allows the boundaries of the CFT application to be principled, and conceptually grounded, rather than conventional or thinly pragmatic (as would be the case were we to forbid use of CFT in exactly those instances where its results do not agree with empirical findings *simply because* they do not agree). As a result, a skilled practitioner not only can apply CFT models correctly but can be aware of the range of the model's effectiveness, including the sorts of properties it can represent and to what degree of accuracy, the sorts of reasoning reliably supported by the restricted quantitative structure, and under what circumstances the model is likely to break down. Knowing the range of effectiveness facilitates a certain conceptual understanding of the subject matter: When the model is found to be ineffective, a practitioner will infer that the representational capacities of the scheme are insufficient to capture all relevant aspects of the phenomenon under consideration. This, in turn, encourages comparison of the CFT model with the more fine-grained molecular-orbital representations, a process that leads to deeper understanding of the more principled theory.

As a representational scheme, CFT allows chemists to rationalize particular facts, to generalize about metal ion complexes, and to make theoretical connections between metals and other chemical substances. A skilled practitioner of CFT has acquired an articulated awareness that is a valuable sort of knowledge. There is much more to be said, but for now we return to the issues with which we began.

4. Return to Explanation.

Reconsider the cluster of claims concerning explanation and explanatory power introduced at the beginning of this discussion:

> **C1**. Explanatory power is, ceteris paribus, a theoretical virtue. Other things being equal, we are justified in preferring more explanatory theories to those that are less so.
>
> **C2**. There is no general consensus about what constitutes adequacy or success conditions for particular explanations. Nor is there a clear conception of how explanatory power is linked to the individual explanations a theory provides.
>
> **C3**. The theoretical systems that scientists themselves judge to be highly explanatory include not only a wide array of conceptual and logical structures but also a diverse set of representational schemes and formats. Alternative representations of the same information, moreover, are often granted different explanatory status.

These are quite different sorts of claims. C1 is a substantive normative claim about explanatory power. C2 reports the current state of affairs concerning philosophical analysis of explanatory concepts. Such analysis is frequently assumed to provide at least partial foundation for justifying C1. (Currently, there are four developed accounts of scientific explanation: Nomological, Causal, Erotetic, and Unificationist. There are stark differences in the properties of explanatory discourse each account chooses to emphasize and analyze.) C3 describes the judgments of practitioners themselves. The relevance of C3 to the other two will depend strongly upon whether one takes practitioners' judgments to be reliable.

In this discussion, I am committed to something like the following methodological principle:

> **MP1**. Assume that practitioners are generally reliable in classificatory judgments regarding the central categories of their own practice, unless there are clear and compelling reasons not to do so.

Let me briefly defend the adoption of MP1. If scientific opinion regarding explanatory power is generally reliable, then it should provide ample grist for philosophical analysis. After all, identifying members of a set often can be useful for discerning the general requirements (especially necessary conditions) for membership. Now imagine instead that scientists' judgments are out of step with philosophical judgments, such that much of what they classify as "explanatory" seems, to us, not to be so. Inferring that the scientists' judgments are incorrect, simply because they are different, presupposes the adequacy of some pre-existing general account of the nature of explanation. How could this assumption be justified? Should we not ask whether we have any reason to be more interested in the philosopher's conception of explanation than in whatever category the scientists have embraced (assuming it is coherent)? For surely the scientists' concept is likely to be operative in the practice, and understanding and justifying the "rationality" of the practice is the overarching goal of the philosophical analysis in the first place. Moreover, the lack of consensus in philosophical analysis should shake our confidence that philosophers have any independent grip on explanatory concepts. Thus, MP1 seems quite sensible.

With its adoption, descriptive claims about scientists' judgments are assumed evidential for the adequacy of conceptual analysis, making C3 highly relevant to C1 and C2. C3 asserts that explanatory frameworks across the sciences are diverse in content, structure, and format. Presumably, then, the existence of diversity can be philosophically informative. We should not, without reason, attempt to make the diversity disappear—for example, by imposing a unified analysis simply by decree. A causal model of explanation, for instance, cannot be judged sufficient if we witness considerable explanatory discourse making no reference whatsoever to causal relations. Instead, our notions both of what constitutes adequate explanations and why explanatory power is a theoretical virtue should *account* for this diversity.

Admittedly, this line of analysis greatly emphasizes the role of descriptive claims such as C3 in making conceptual progress. While I have offered a preliminary rationale for methodological reliance on such claims, I have not argued for the truth of C3 itself. Assuming it can be done (the evidence required is not hard to envision), the more significant challenge is to determine how C3 should influence our making sense of the claims in C1 and C2. In my final comments, let me sketch one possibility. Since there is little room to build up from the foundations, I will present the view rather than lead you to it.

5. A Suggestion for Making Sense of Explanatory Power. Explanatory discourse in science serves to make both the subject matter and the scientific practice itself seem intelligible. It does this by using paradigmatic examples to communicate implicitly the aims and corresponding values that constitute the practice, right alongside assertions of what sorts of information or reasoning would make the object of explanation intelligible. For example, using Newton's laws of mechanics to explain terrestrial motion amounts to an assertion of what form of awareness is appropriate and what model of intelligibility a community of practitioners has chosen to embrace. There is a normativity in our explanations that derives from the underlying values of the associated practices. Explanatory discourse serves to inculcate these values as a set of practiced skills. The student learns, through doing, how to apply the mathematical structure of the vector calculus to bodies sliding down inclined planes. In the development of these skills, we witness the embodiment of the values of the community she is attempting to join. The same can be said for the student of inorganic chemistry learning to reason with crystal field theory (CFT) diagrams.

It would be a mistake to assume, however, that the formation of a coherent scientific practice begins with a clear and determinate set of values, which are self-consciously imposed on the structure of its explanations. Rather, something akin to a feedback loop exists. A scientific community adopts a rough set of epistemic and practical aims and begins exploring some domain of inquiry. Certain theoretical structures, that is, certain representations, will be successful with respect to these rough aims; others will not. Successful representations will influence the evolution of the aims of the practice by implicitly reinforcing certain values and undermining the legitimacy of others. In the extreme, the community may come to assume that any question not successfully addressed by existing theoretical tools is even inappropriate to consider. (The stance of some early logical positivists with respect to issues in ethics and aesthetics comes immediately to mind.) Alternatively, a community may invest considerably in cultivating skills to use a particular representational tool for a certain job rather than abandon the tool or leave the job undone. (Consider the nineteenth-century development of the Newtonian framework for a kinetic theory of gases.)

Now we can envision, in crude outline, the connection between representational schemes and explanatory power: Certain representations are judged to be successful by a community of scientists. The nature of this success is pragmatic, being dependent upon the particular aims of this community. Success, in turn, encourages continued and expanded use. Practitioners attempt to make the structure of particular representations portable; in effect, they try to generalize them (e.g., CFT). If certain representational strategies continue to have great success, their status within the practice is elevated to such an extent that they become actual models of intelligibility for the community. When scientists judge a theoretical structure to be explanatory, they are either assigning or recognizing such a status. In this way, explanatory discourse, itself a subset of our most successful representational practices, influences our very notions of intelligibility, rather than simply displays them.

Given all this, it should come as no surprise when I suggest that *disciplinary identity itself is formed largely through explanatory preferences*. Though we tend to identify disciplines casually by subject matter, doing so is clearly insufficient. There currently is not, nor has there ever been, any serious chance of the disciplinary divide between chemistry and physics dissolving because of the overlap in their domains. Conversations debating the difference between the chemist's atom and that of the physicist can still be heard, or had. Such talk is not vacuous, but it is also not a claim about metaphysics. It is a claim about method. The different disciplines (and subdisciplines, for that matter) employ different representations of atoms and molecules, which in turn make different forms of investigation intelligible, coherent, and feasible. Explanations, and the representational schemes they employ, encode the aims and values of particular intellectual communities. Explanatory discourse tells us what we should want to know about the world and how we should reason to get there. It also tells us who our intellectual family is.

Does this perspective help to make sense of the connections between C1 and C2? Potentially so. We should recognize theoretical virtues hidden behind the label "explanatory power" as precisely the success-granting properties of a good representation scheme. But, the reader may object, successful representation schemes come in all shapes and sizes. Yes, indeed, good schemes must be tailored jointly to their targets, the capabilities of the intended users, and those users' aims. Consequently, the justification of a preference for explanatory theories must follow, in tandem, the lines of justification we might offer for the pragmatic virtues of powerful representation schemes.

Moreover, on this account, the lack of consensus regarding what constitutes adequacy or success conditions for particular explanations is not at all surprising. Particular explanations derive from general structures that have achieved the status of being "explanatory." But since this explanatory status is pragmatically oriented and context dependent, we should expect diversity at the level of individual explanations. At the same time, the prevalence of particular patterns of explanation—most notably those exposing causal structures (Salmon 1984) or exhibiting nomological subsumption (Hempel 1965)—can itself be understood as a consequence of the widespread sharing of certain notions of intelligibility across natural scientific disciplines.

The perspective I am recommending, because of its pragmatic orientation, will seem a natural ally to the erotetic account of explanation offered by van Fraassen (1980), and indeed it is. But while many philosophers have been dissatisfied with the apparently deflationary stance of the erotetic account, the view offered here provides both diagnosis and remedy for this seeming deficiency. The erotetic account seems thin if characterized as "explanation is any answer to a why question that the community will accept as such." But the judgment is accurate only if we assume a correspondingly weak notion of "community." A community, however, is not simply a group of people sitting together in a room. If *we* are a community, coming together periodically for conferences is not what makes it so. It is because we share things—interests, aims, values—that are sufficient for bringing us together at these times. And because what a community will accept depends upon these same defining elements, the standards for explanatory discourse can be as demanding and robust as the identity of the community itself. Furthermore, if explanatory discourse both shapes and sustains the coherence of an intellectual community by communicating and reinforcing the things that make the community what it is, then we can understand how explanatory discourse has a rich role in scientific practice.

Though the connection may be less apparent, this perspective builds also upon unification accounts of explanation. What is generally distinctive about unification accounts is their focus on the overarching explanatory power of a theoretical structure rather than on the particular explanations provided by that structure. Friedman's (1974) version finds organization in reduction of the number of brute facts a community must accept while Kitcher (1989) stresses the role of argument patterns that can be applied broadly. Each is trying to capture the aspect of savvy representation that concerns making sense by making connections. Yet by allowing logical structure to overshadow other aspects of knowledgeable awareness, neither account appears to recognize that it is not unification of claims per se that is at issue, but instead unification of practice. This methodological unification, frequently exhibited in patterns of reasoning like those Kitcher discusses, or representational schemes such as crystal field theory, facilitates the formation of coherent epistemic communities.

6. Concluding Remarks. I have suggested that sophisticated representational schemes in the sciences be viewed as pragmatic tools for acquiring the sort of articulated awareness that is the hallmark of nontrivial knowledge. In part because of the intellectual grip they provide, certain representations achieve the status of being paradigmatically explanatory, shaping models of intelligibility, which themselves serve largely to define and differentiate disciplinary communities. In this way, the pragmatic nature of explanatory discourse effectively grants its intellectual utility. The observed diversity of explanatory structures derives from this same pragmatic foundation. The disunity of philosophical discussion concerning scientific explanation seems a consequence, then, of the function of explanation within science. That it has been hard to make sense of explanation is not surprising.

REFERENCES

Friedman, Michael (1974), "Explanation and Scientific Understanding", *Journal of Philosophy* 71: 5–19.

Haugeland, John (1998), *Having Thought*. Cambridge, MA: Harvard University Press.

Hempel, Carl G. (1965), "Studies in the Logic of Explanation", in *Aspects of Scientific Explanation and Other Essays*. New York: The Free Press.

Huheey, James (1983), *Inorganic Chemistry*, 3rd ed. New York: Harper and Row.

Kitcher, Philip (1989), "Explanatory Unification and the Causal Structure of the World", in Philip Kitcher and Wesley Salmon (eds.), *Minnesota Studies in the Philosophy of Science, Vol. 13: Scientific Explanation*. Minneapolis: University of Minnesota Press.

Salmon, Wesley C. (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.

van Fraassen, Bas C. (1980), *The Scientific Image*. Oxford: Clarendon Press.