Epistemology Futures, Stephen Hetherington, ed.
Oxford: Clarendon, 2006, pp. 199-215.

From Knowledge to Understanding

Catherine Z. Elgin

Science, Spencer contends, is organized knowledge.[1] No doubt science is organized. Nevertheless, epistemologists speaking ex cathedra should deny that it is knowledge. 'Knowledge' is a factive. An opinion is not knowledge if it is not true. But even the best scientific theories are not true. Although science may produce some justified or reliable true beliefs as byproducts, for the most part, the deliverances of good science are not knowledge.

The analysis of 'knowledge' that yields this untoward verdict accords with our intuitions about the proper use of the term. We do not consider false beliefs knowledge, no matter how well grounded they may be. Once we discover that a belief is false, we retract the claim to know it. So we ought to deny that our best scientific theories are expressions of knowledge. Nevertheless, good science affords some sort of worthwhile take on nature. Epistemology should explain what makes good science cognitively good. It should explain why it is correct to say that we learn science in school rather than just that we change our minds about scientific matters. Its current focus on knowledge, being too narrow, stands in the way.

My goal in this paper is to show how epistemology's emphasis on knowledge constricts and distorts its purview, and to begin to sketch an epistemology capable of accounting for the cognitive contributions of science. Although I concentrate on science, the epistemological factors I foreground figure in other disciplines as well. My focus on science is mainly strategic. Science is undeniably a major cognitive achievement. It would be implausible in the extreme to contend that science's claim to epistemic standing is suspect. Moreover, science is methodologically self-reflective. So epistemically significant factors may be easier to recognize in science than in other disciplines. The epistemology of science then can serve as an

entering wedge for a broader reconsideration of the nature and scope of human cognitive achievements.

Good science, as I use the term, is science that affords epistemic access to its subject matter. A good theory is a theory underwritten by good science. A central ambition of this paper is to begin to characterize that mode of epistemic access. For now, all that is necessary is to concede that some science is cognitively good, and that scientists often can tell what science is good. Although I will offer a sketch of how I think epistemology should approach the issue, my main purpose is to make a convincing case that it should – that something of major significance is omitted if our understanding of our epistemic condition does not account for the contributions of science.

Knowledge, as epistemology standardly conceives of it, comes in discrete bits. The objects of knowledge are individual facts, expressed in true propositions and/or stated in true declarative sentences. Judy knows (the fact) that the bus stops at the corner. Suzy knows (the fact) that ripe strawberries are red. These discrete bits are supposed to be what is justified or what is generated and sustained by reliable mechanisms. We can readily identify the evidence that supports Judy's belief, and the perceptual mechanisms that sustain Suzy's, and we can explain how they secure the beliefs in question. What emerges is a granular conception of knowledge. A subject's knowledge consists of discrete grains, each separately secured. She amasses more knowledge by accumulating more grains. Goldman labels such truth-centered epistemology *veritism*.[2] Whether or not veritism is plausible for mundane knowledge, I contend, it is clearly inadequate for science.

Science is holistic. It is not an aggregation of separate, independently secured statements of fact, but an integrated, systematically organized account of a domain. Let us call such an account a theory.[3] There is no prospect of sentence by sentence verification of the claims that comprise a theory, for most of them lack separately testable consequences. In Quine's words, they 'confront the tribunal of sense experience not individually but only as a corporate body'.[4] Independent of a theory of heat transfer, nothing could count as evidence for or against the claim that a process is adiabatic. Independent of an evolutionary theory, nothing could count as evidence for or against the claim that a behavior manifests reciprocal altruism. Together the sentences of a theory have testable implications; separately they do not. Indeed, it is not even clear that all

scientific statements have truth values in isolation. If the individuation of the items they purport to refer to -- a species, or a retrovirus, for example -- is provided by a theory, there may be no fact of the matter as to whether they are true independent of the theory.

Such holism might seem epistemologically innocuous. One way to accommodate it would be to take the bulk of a theory as "background knowledge" and then ask whether, together with the empirical evidence, it affords sufficient grounds to underwrite a particular claim. Given the theory and the empirical evidence, does this food sharing manifest reciprocal altruism? Although this reveals whether a theory supports a claim, it plainly does not solve our problem. For the assumption that the "background knowledge" is genuine knowledge cannot be sustained. There is no viable non-holistic explanation of how the individual sentences of the theory serving as background could have obtained the support they require to qualify as knowledge. Scientific theories are not granular in the way that epistemology takes knowledge to be.

Another, perhaps more promising strategy is to take holism at its word. The simple sentences that comprise a theory cannot be separately justified. Evidence always bears on a theory as a whole. So evidence for the claim that a given process is adiabatic is evidence for an entire theory of heat transfer, which is tested along with the claim. This is in principle epistemologically unproblematic. The contention that knowledge is propositional says nothing about the length of the propositions that constitute knowledge. We can accommodate scientific holism by treating a theory as a conjunction of its component propositions and saying that the evidence bears on the truth or falsity of that long conjunction. If the conjunction is true, is believed, and is justified or reliably produced, it is known.

This may be as good a schema for scientific *knowledge* as we are likely to get. But it sheds little light on the cognitive value of science, for its requirements are rarely met. In particular, the truth requirement is rarely satisfied. As will emerge, theories contain sentences that do not even purport to be true. For now, however, this complication will be ignored. Still there is a problem. For even the best scientific theories confront anomalies. They imply consequences that the evidence does not bear out. Since a conjunction is false if any of its conjuncts is, if a scientific theory is a conjunction, an anomaly, being a

falsifying instance, tells decisively against the theory that generates it. Since a theory that generates an anomaly is false, its cognitive deliverance is not knowledge.

Perhaps we can evade this predicament. The characterization of a theory as a conjunction might seem to offer some hope of isolating anomalies and screening off their effects.[5] All we need to do is identify and expunge the troublesome conjuncts. Consider the following conjunction:

(1) (a) Sally is in Chicago & (b) Sam is in New York

If Sally is in fact in Detroit, (1) is false, even though Sam is in New York. If we lack adequate evidence that Sally is in Chicago, (1) is unjustified, even though we have ample evidence that Sam is in New York. If our source of information about Sally's whereabouts is suspect, (1) is unreliable, even though our source of information about Sam's location is impeccable. (1) then is not something we are in a position to know. Still, we can rescind (a), leaving

(b) Sam is in New York

which is true, justified, and reliable. Since neither (a) nor the evidence for (a) lends any support to (b), (b)'s tenability is not undermined by the repudiation of (a). On standard accounts of knowledge, we are in a position to know that (b). If the components of a scientific theory were related to one another as loosely as (a) and (b) are related in (1), we could simply rescind the anomalous sentences and be left with a justified, reliable truth – something that could be known.

But the components of a theory lack the requisite independence. A theory is a tightly interwoven tapestry of mutually supportive commitments. Simply excising anomalous sentences would leave a moth-eaten tapestry that would not hang together. Before Einstein, physicists devised a variety of increasingly drastic revisions in their theories to accommodate the perturbation in Mercury's orbit. But even at their most desperate, they did not suggest simply inserting an exception into the theory. Although 'All planets except Mercury have elliptical orbits' is apparently true, justified, reliably generated, and believed, it pulls so strongly against the ideal of systematicity that scientists never considered incorporating it into astronomy. Temporarily bracketing anomalies may be a good tactic in theory development, but simply discounting them as exceptions is not. The reason is not merely aesthetic. An anomaly might be just a pesky irritation that

stems from undetected but ultimately insignificant interference, but it might also, like the perturbation in Mercury's orbit, be symptomatic of a subtle but significant misunderstanding of the phenomena. Science would lose potentially valuable information if it simply dismissed its anomalies as exceptions that it need not explain. There is then no hope of simply extracting anomalous sentences without undermining the epistemic support for the rest of the theory. The theory rather than the individual sentence is the unit we need to focus on.

These points are familiar and uncontroversial, but their epistemological consequences are worth noting. A theory can be construed as a conjunction of the sentences that appear in it. But science does not yield knowledge expressed by such conjunctions. For the conjunction of the sentences that constitute a good scientific theory is apt to be false. The unavailability of sentence by sentence verification discredits the idea that science delivers knowledge of each component sentence. The hopelessness of selectively deleting falsehoods in and false implications of a theory undermines the plausibility of claiming that scientific knowledge is what remains when a theory's falsehoods have been expunged. Knowledge requires truth. And there seems to be no feasible way to get good scientific theories to come out true. So knowledge is not the cognitive condition that good science standardly engenders. We seem forced to admit that scientific accounts that contain falsehoods nonetheless constitute cognitive achievements. If so, to understand the cognitive contribution of science, knowledge is not the epistemic magnitude we should focus on.

Much good science falls short of satisfying the requirements for knowledge. But the problem is not just a shortfall, it is a mismatch. For mere knowledge does not satisfy the requirements of good science either. Science seeks, and often provides, a unified, integrated, evidence-based understanding of a range of phenomena. A list, even an extensive list, of justified or reliably generated true beliefs about those phenomena would not constitute a scientific understanding of them. Veritism, in concentrating on truth, ignores a host of factors that are integral to science. These factors cannot be dismissed as just instrumentally or practically valuable. They are vital to the cognitive contributions that science makes. In assessing a theory, we should not ask, 'Does it express knowledge?' Rather, we should ask, 'Does it convey an understanding of the phenomena? Is it a good way to represent or think about a domain if our goal is to

understand what is going on in that domain?'

Representation depends on categorization, the division of a domain into individuals and kinds. The members of any collection, however miscellaneous, are alike (and unlike) one another in infinitely many ways. So in seeking to devise a taxonomy, we cannot hope to appeal to overall likeness. Nor is it always wise to group items together on the basis of prescientifically salient similarities. Different diseases, such as viral and bacterial meningitis, often display the same symptoms, and a single disease, such as tuberculosis, can manifest itself in different clusters of symptoms. A science requires a taxonomy or category scheme that classifies the items in its domain in a way that furthers its cognitive interests – discovery of causal mechanisms, functional units, widespread patterns, overarching or underlying regularities, and so on. Science regularly reveals that things that are superficially alike are deeply different and things that are superficially different are deeply alike. Without an adequate system of categories, significant likenesses and differences would be missed.

Scale is critical. As Nancy Cartwright's discussion of Simpson's paradox shows, factors that are salient or important at one level of generality can be unimportant at another.

> The graduate school at Berkeley was accused of discriminating against women. . . . The accusation appeared to be borne out in the probabilities: The probability of acceptance was much higher for men than for women. Bicknell, Hammel, and O'Connell looked at the data more carefully, however, and discovered that this was no longer so if they partitioned by department. In a majority of the eighty-five departments, the probability of admission for women was just about the same as for men, and in some even higher for women than for men. . . . [W]omen tended to apply to departments with high rejection rates, so that department by department women were admitted in about the same ratios as men but across the whole university considerably fewer women, by proportion, were admitted.[6]

Admissions rates calculated department by department show one pattern; overall rates show another. The point is general. At different scales, the same data display different patterns. It is not unusual in biology for subpopulations to display one pattern and the larger population to show another. Each pattern is really

instantiated. But to understand what is occurring in the domain requires knowing which pattern is significant.

Both categorization and scale involve selection. The issue is what factors to focus on. The problem is that there are too many epistemically accessible facts about a domain. To obtain any sort of systematic understanding requires filtering. Science has to select, organize and regiment the facts to generate such an understanding. It needs criteria for selection, organization and regimentation. Veritism does not supply them.

Such criteria are far from arbitrary. It is possible to make mistakes about them. If we choose the wrong scale, we miss important patterns. We wrongly decide that Berkeley is, or that it is not, discriminating. We wrongly conclude that a genetic trait is, or that it is not, widespread in a species. If we draw the wrong lines, we miss important similarities and differences. We wrongly conclude that rabbits and hares are, or that they are not, the same sort of thing. In such cases, we fail to understand the phenomena, even if our account consists entirely of justified true beliefs.

Science places a premium on clarity. It favors sharply differentiated categories whose members are readily distinguished. One reason is that science is a collaborative enterprise grounded in shared commitments. Because current investigations build on previous findings, it is imperative that scientists agree about what has been established and how firmly it has been established. Clarity and definiteness foster intersubjective agreement and repeatable results. Repeatability requires determinacy. Unless it is possible to tell what the result of a given investigation is, it is impossible to tell whether a second investigation yields the same result or a different one; whether it yields a cotenable result or a noncotenable one. Vagueness is undesirable then, since within the penumbra of vagueness there may be irresolvable disagreements about what situation obtains.

The requisite clarity and determinacy can sometimes be achieved by fiat. We eliminate vagueness by stipulating where sharp lines will be drawn. But even if lines are sharp, instances may prove irksome. The sharp criteria for distinguishing mammals from birds may leave us bewildered or dissatisfied about the classification of the platypus. Sometimes, regimenting familiar categories does not yield a partition of the

domain that suits scientific purposes. Either the lines seem arbitrary or they do not group items in ways that disclose the regularities or patterns the science seeks. 'Weight' for example, is a familiar and easily regimented category. It is of relatively limited scientific interest, though, since it is a function of gravity, which varies. 'Mass', although less familiar, is a more useful category, for it remains constant across variations in gravity. Where gravity is constant, weight may be a fine magnitude to use. Where differences in gravity matter, science does better to measure in terms of mass. To the extent that systematicity is of value, this is a reason to favor mass over weight across the board. A critical question then is what modes of representation foster the realization of scientific objectives. Phenomena do not dictate their own descriptions. We need to decide in what units they should be measured and in what terms they should be described.

Rather than characterizing familiar items in familiar terms, science often construes its phenomena as complexes of identifiable, even if unfamiliar, factors. Frequently the factors are not assigned equal significance. Some are deemed focal, others peripheral. The liquids that fall from the skies, that flow through the streams, that lie in the lakes contain a variety of chemicals, minerals and organic material. Nonetheless, we call all these liquids 'water', acknowledging only when necessary, that there are chemical, mineral, and biological ingredients as well. Tellingly, we call such ingredients 'impurities'. H_2O then is taken as the focus, and the other components are treated as peripheral. Most of the liquid we call 'water' does not consist wholly of H_2O. To obtain pure samples of the focal substance requires filtering out impurities. The justification for calling the liquids 'water' and identifying water with H_2O is not fidelity, but fruitfulness. Our scientific purposes are served by this characterization. Sometimes, the effects of the impurities are negligible, so we can treat the naturally occurring liquid as if it were H_2O. In other cases they are non-negligible. Even then, though, H_2O serves as a least common denominator. We compare divergent samples in terms of how and how far they differ from 'pure water' -- that is, H_2O. There is nothing dishonest about using a description that focuses on H_2O. But it would be equally accurate to simply describe the liquid in the rain barrel, the lake and the river more fully. Instead of characterizing them as impure water, we could simply supply the chemical, biological and mineral profile of the liquid in Walden Pond, the

liquid in the Charles River, and the liquid that fell in today's storm. Although the latter descriptions would be accurate, they would mask the common core. Treating the three samples as instances of a single substance differing only in impurities highlights features they share. And by seeing what they share we can begin to investigate their differences. Why are the impurities in one sample, e.g., the water from Walden Pond, so different from the impurities in another, the water from the Charles River?

This pattern is widespread. Astronomers describe the motions of the planets in terms of regular geometric orbits with perturbations. Linguists describe verbal behavior as rule-based competence overlaid with performance errors. Engineers describe the output of a sensor as a combination of signal and noise. In all such cases the focal concept serves as a point of reference. What occurs in the domain is understood by reference to, and in terms of deviations from, the focus.

Although these examples exhibit the same conceptual configuration, the differences between them are significant. Where it is a matter of signal and noise, only the focal element – the signal -- is important. It is often both possible and desirable to sharpen the signal and eliminate or dampen the effects of the noise. We fine tune our measuring devices or statistical techniques to eliminate static and highlight focal features. In cases where noise is ineliminable, it is simply ignored. What counts as signal and what counts as noise varies with interests. Ordinarily, when someone answers questions, the content of the answers is the signal. But in some psychology experiments, content is mere noise. The signal is reaction time. Psychologists want to ascertain not what a subject answers, but how long it takes her to answer, for reaction time affords evidence about psychological and neurological processes. The choice of a focus is thus purpose relative.

We cannot always ignore complications. If we want to understand language acquisition, we cannot simply overlook performance errors. We need to see how or whether they affect what is learned. If we want to send a probe to Mars, we cannot simply ignore the planet's deviation from a perfect elliptical orbit. We must accommodate it in our calculations. In such cases, we employ a schema and correction model. We start with the focal concept and introduce elaborations to achieve the type and level of accuracy we require.

All these cases involve streamlining the focus and sidelining or downplaying complexities. Sometimes, as in the model of signal and noise, the complexities are permanently sidelined. As much as

possible, we sharpen the signal and eliminate static. We have no reason to reintroduce the static we have removed. In other cases, when the model of schema and correction is appropriate, complexities may be set aside only temporarily. They may need to be reintroduced at a later stage.

Focal points are readily defined. The choice among them turns on utility, not just accuracy. Three points described by Dennett illustrate this: The center of gravity is 'the point at which the whole weight of a body may be considered to act, if the body is situated in a uniform gravitational field'.[7] The center of population of the United States is 'the mathematical point at the intersection of the two lines such that there are as many inhabitants north as south of the latitude and as many inhabitants east as west of the longitude'.[8] Dennett's lost sock center is 'the center of the smallest sphere that can be inscribed around all the socks' that Dennett has ever lost.[9] All three points are well defined. Each is as real as any of the others. If points are real, all three exist; if points are unreal, none of the three exists. If points are constructed through stipulative definition, all three points are equally constructs. Whatever their ontological status, all are devices of representation. We represent portions of reality in terms of them. Still, they are hardly on a par.

Gravity is a fundamental force whose effects are uniform, law governed, and ubiquitous. It is often simpler, both conceptually and computationally, to represent an extended body as a point mass located at the body's center of gravity, and to calculate, predict, and explain gravitational effects of and on the body as though it were a point mass located at the center of gravity. The center of gravity is a manifestly useful device of representation.

Dennett's lost sock center is inconsequential. It does not engage with any significant questions, even if one happens to care about Dennett's propensity for losing socks. Conceivably a biographer or psychologist might take an interest in the distribution of his lost socks. But exactly where the midpoint lies makes no difference. Dennett's lost sock center is a well-defined, utterly trivial point.

The center of population of the United States is an intermediate case. It changes over time, and its changes display both short term fluctuations and long term trends. It shifts, day by day, even minute by minute, as people move about, some of them crossing the crucial lines, now this way, now that. The fluctuations are insignificant. But through the fluctuations we can discern a trend. If we look at the change

in the population center, not by day but by decade, we see that US population has moved westward. This is a significant demographic change. It engages with other sociological information and figures in a broader understanding of American society. So the center of population is not, like Dennett's lost sock center, a useless point. But it is not, perhaps, as useful as it might be. To discern the demographic trend, we need to see past the noise generated by the small scale fluctuations. We might do better to devise a different device of representation. Rather than an instantaneous measure, perhaps we should concentrate on longer periods of time. The representation might still take the form of a point, but it would not represent a position at an instant. A better focus could readily be devised.

It is critical that the focus need not occur naturally. Laboratory processes may be required to obtain a refined, pure sample of a focal substance like H_2O. Computational processes may be required to fix the population points that best display important demographic trends. Sensor readings are subjected to statistical analyses to synthesize the information we seek. In yet other cases conceptual processing is called for. To understand grammatical errors it may be helpful to subject an utterance to a sort of conceptual factor analysis, construing it as consisting of invariable grammatical rules overlaid with idiosyncratic applications. The focus of representation may be fairly distant from the robust phenomena it bears on.

We construct devices of representation to serve certain purposes and can reconstruct them both to enable them to better serve their original purposes and to serve other purposes that we may subsequently form. We can revise the scope, scale, and content of our representations to improve their capacity to promote our evolving cognitive ends. In such matters there are feedback loops. As we come to understand more about a domain we refine our views about what kinds are significant, at what level of generality they should be investigated, in what terms they should be represented.

Ecologists sampling the water in Walden Pond ordinarily would not just extract a vial of liquid from any convenient place in the pond. They would consider where the liquid is most representative of the pond water, or is most likely to display the features they seek to study. If they seek a representative sample, they would not take it from the mouth of the stream that feeds the pond, nor from the shore right near the public beach, nor from the area abutting the highly fertilized golf course. They might draw their sample from the

middle of the pond. Or they might take multiple samples from different areas and either mix them physically or generate a composite profile based on them. Their sampling would be guided by an understanding of where in the pond the features they are interested in are most likely to be found. This means though that even if the water in the sample occurs naturally, data collection is driven by an understanding of the domain, the way it is properly characterized and the way it is properly investigated. All these go into determining what makes a sample a representative sample.

A sample is not just an instance. It is a telling instance. It exemplifies, highlights, displays or conveys the features or properties it is a sample of. No sample exemplifies all its features. Exemplification is selective. The sample drawn from Walden Pond is (a) more than 1000 kilometers from the Parthenon, (b) taken by a left handed graduate student, (c) obtained on the second Tuesday of the month. It also (d) contains H_2O, (e) contains E. coli bacteria; (f) has a pH of 5.8. In a suitable scientific context, it may well exemplify any or all of (d), (e), and (f). Although it instantiates (a), (b) and (c), it is unlikely in normal scientific contexts to exemplify any of them.

A sample then is a symbol that refers to some of the properties it instantiates. It thereby affords a measure of epistemic access to these properties. Epistemic access can be better or worse. One reason for careful sampling is to insure that the sample has the properties of interest; another is to obtain a sample that affords ready epistemic access to them. Some factors occur only in minute quantities in pond water, so although a liter of water drawn from the pond exemplifies them, they may still be hard to detect. Moreover, such a sample may include confounding factors, which although unexemplified and (for current purposes) irrelevant, impede epistemic access to exemplified properties. So instead of working with samples drawn directly from nature, scientists often process samples to amplify features of interest and/or remove confounding factors. In the lab, the water sample undergoes purification processes to remove unwanted material. What results is a pure sample in which the features of interest stand out. Scientists then experiment on this sample, and devise explanations and predictions based on its behavior. Although the lab specimen does not occur naturally in the form in which it is tested, the tests are not a sham. For the features the specimen exemplifies do occur naturally. The lab specimen's divergence from nature in exemplified

features is neglible; its divergence in other respects is irrelevant.

Different sorts of samples are suited to different experiments. Scientists might experiment on a random sample of a substance, a purposeful sample, or a purified sample. In all such cases, the goal is to understand nature. An experiment is designed to reveal something directly about the sample, which can be projected back onto the natural phenomena it bears on. Just how to project from the lab to the world depends on the sort of sample used, and the operative assumptions about how it relates to the phenomena whose features it exemplifies. The extrapolation is not always strightforward. A good deal of interpretation may be required to effect the projection.

To determine whether a substance *S* is carcinogenic, investigators place genetically identical mice in otherwise identical environments, exposing half of them to massive doses of *S* while leaving the rest unexposed. The common genetic endowment and otherwise identical environments neutralize the vast array of genetic and environmental factors that are believed to standardly influence the incidence of cancer. By controlling for genetics and most aspects of the environment, scientists insure that these factors, although instantiated by the mice, are not exemplified. They arrange things so that exposure or non-exposure to *S* is the only environmental feature exemplified, thereby enabling the experiment to disclose the effects of *S*. The use of mice is grounded in the assumption that, in the respects that matter, mice are no different from humans. Given this assumption, the experiment is interpreted as exemplifying *the effect on mammals*, not just on mice. The mice are exposed to massive doses of *S*, on the assumption that the effect of lots of *S* on small mammals over a short period is reflective of the effect of small amounts of *S* on larger mammals over a long period. So the experiment is interpreted as exemplifying *the effect of S* rather than just the effect of high doses of *S*. To make its cognitive contribution, of course, experiment must be properly interpreted. If we took the experimental situation to replicate life in the wild, we would be badly mistaken. But if the background assumptions are sound, then we understand the ways the experiment is and is not representative of nature – that is, we understand what aspects of the experiment symbolize and how they do so. That enables the experiment to advance understanding of the effect of *S* on mammals.

The experiment is highly artificial. Even the mice are artifacts, having been intentionally bred to

exhibit a certain genetic structure. The exposure is to a vastly higher dose of *S* than would occur in nature. The environment is rigidly controlled to eliminate a huge array of factors that normally affect the health of mice. The experiment eliminates some ordinary aspects of mouse life, such as the dangers to life and limb that predators pose. It nullifies the effects of others, such as the genetic diversity of members of a wild population of mice. It exaggerates others, exposing the mice to much higher doses of *S* than they would be exposed to naturally. Rather than rendering the experiment unrepresentative, these divergences from nature enable the experiment to reveal aspects of nature that are normally overshadowed. They clear away the confounding features and highlight the significant ones so that the effects of *S* on mammals stand out.

Science distances itself even further from the phenomena when it resorts to models, idealizations, and thought experiments. Scientific models are schematic representations that highlight significant features while prescinding from irrelevant complications. They may be relatively austere, neglecting fine grained features of the phenomena they concern. They may be caricatures, exaggerating features to bring subtle but important consequences to light.[10] They may be radically incomplete, representing only selected aspects of the phenomena.[11] Strictly and literally, they describe nothing in the world. For example, although financial transactions are complexes of rational and irrational behavior, economics devises and deploys models that screen off all factors deemed irrational, regardless of how large a role they play in actual transactions. Such models would provide nothing like accurate representations of real transactions, but would not be defective on that account. They operate on the assumption that for certain purposes irrationality can safely be ignored.

Construed literally, models may describe ideal cases that do not, perhaps cannot, occur in nature. The ideal gas is a model that represents gas molecules as perfectly elastic, dimensionless spheres that exhibit no mutual attraction. There are -- indeed there could be -- no such molecules. But the model captures the interdependence of temperature, pressure, and volume that is crucial to understanding the behavior of actual gases. Explanations that adduced the ideal gas would be epistemically unacceptable if abject fidelity to truth were required. Since helium molecules are not dimensionless, mutually indifferent, elastic spheres, an account that represents them as such is false. But, at least if the explanation concerns the behavior of helium in circumstances where divergence from the ideal gas law is negligible (roughly, where temperature is high

and pressure is low) scientists are apt to find it unexceptionable. For in such circumstances, the effects of friction, attraction, and molecular size do not matter. Models of economic growth represent the profit rate as constant. In fact, it is not. Non-economic factors such as epidemics, corruption, and political unrest interfere. But by bracketing such complications, the economic models capture features that are common to a host of seemingly disparate situations. Even though the full blooded situations seem very different from one another, the model presents a common core and enables economists to (partially) explain seemingly disparate behaviors in terms of that core. Thus representations that are and are known to be inaccurate afford insight into the phenomena they purport to concern.

Thought experiments are imaginative representations designed to reveal what would happen if certain conditions were met. They are not actual, and often not even possible, experiments. Nonetheless, they afford an understanding of the phenomena they pertain to. By considering the experience of a person riding on an elevator with and without the presence of a gravitational field, Einstein shows the equivalence of gravitational and inertial mass. By considering how a light body tethered to a heavy body would fall, Galileo both discredits the Aristotelian theory motion and discovers that the rate at which objects in a vacuum fall is independent of their weight. In other cases, thought experiments flesh out theories by revealing what would happen in the limit. By considering how electrical currents would behave in metals cooled to absolute zero, a computer simulation yields insights into superconductivity. The effectiveness of a thought experiment is not undermined by the fact that the imaginary conditions that set the stage never obtain.

Standardly, philosophers assume that scientific theories aim at truth, and are deficient if they are not true. Even good theories confront anomalies. But anomalies are indications that theories are defective. So the existence of anomalies does not in itself discredit the standard view. Although idealizations, simplified models, and thought experiments neither are nor purport to be true, they are not defective. To account for the cognitive contributions of science, epistemology must accommodate their contributions. Such devices, I believe, function as fictions. So to make my case, I need to explain first how fictions advance understanding and then why it is reasonable to consider these devices fictions.

It is not unusual to emerge from an encounter with a work of fiction feeling that one has learned something. But fictions do not purport to be true. So the learning, whatever it is, cannot plausibly be construed as the acquisition of reliable information. Since fiction is indifferent to literal truth, falsity is no defect in it. A fiction need not be 'realistic'. It can transcend the limits of the possible. It can portray characters with unusual combinations of traits and situations that present unusual challenges and opportunities. It can contrive telling mismatches between characters and their situations. It can uproot characters from one environment and implant them in another. Having done such things, it plays out the consequences. If thought experiments, models and idealizations are fictions, they do the same sorts of things. Like other fictions, they are exempt from the truth requirement. So the fact that the ideal gas law is true of nothing in the world is not a mark against it. The fact that no one ever has ridden and no one ever will ride in an elevator without a gravitational field does not discredit Einstein's thought experiment. If they are fictions, such devices are not supposed to be true. But they are not completely idle speculations either. The consequences they play out are supposed to advance understanding of the actual. The question is: If a fictional representation is not true, how can it shed light on the way the world actually is?

I suggest that it does so by exemplifying features that diverge (at most) negligibly from the phenomena it concerns.[12] To take a pedestrian sample, a commercial paint sample is a chip of a precise color. Surprisingly, it is a fiction. The color patch on the card is not a patch of paint, but of ink or dye of the same color as the paint it represents. The fiction – that it is a patch of paint – affords epistemic access to a fact – the color of paint the patch represents. Not all the paint that counts as matching it is exactly same shade. Any color within a certain range counts as a match. The paint sample thus affords access to that narrow range of colors – colors that diverge at most negligibly from the color on the card. The ideal gas law is expressed in a formula relating temperature, pressure, and volume. The model gas is a fiction in which the formula is exactly satisfied. Real gases do not exactly satisfy the formula. Still the model affords epistemic access to the real gases that fall within a certain range of the ideal gas in the relations of temperature, pressure and volume that they display. Both exemplars afford epistemic access to features that they do not possess, but that diverge negligibly from features that they do possess. Obviously, whether a

divergence is negligible depends on a host of contextual factors. A divergence that is negligible in one context may be nonnegligible in another. Since we know how to accommodate the contextual factors, we are in a position to interpret the exemplars correctly.

A fiction exemplifies certain features, thereby affording epistemic access to them. It enables us to discern and distinguish those features, study different aspects of them, consider their causes and consequences. It is apt to be purposely contrived to bring to the fore factors that are ordinarily imperspicuous. By highlighting features in a setting contrived to render them salient, it equips us with resources for recognizing them and their ilk elsewhere. Othello exemplifies a cluster of virtues and flaws that makes him vulnerable to Iago's machinations. That cluster of traits is perhaps not unusual. But the resulting vulnerability is far from obvious. To make it manifest, Shakespeare shows how Othello's character shatters under the pressure Iago exerts. The play thus exemplifies the vulnerability of a cluster of traits by devising a situation where they break down. It considers what would happen in an extreme case, to point up a vulnerability that obtains in ordinary cases. In effect, it tests the cluster of traits to destruction. Just as the medical experiment is carefully contrived to exemplify the carcinogenicity of *S* by subjecting the mice to massive doses of *S*, the play is carefully contrived to exemplify the vulnerability inherent in a cluster of seemingly admirable traits by subjecting Othello to massive evil.

Of course there are differences. A play like *Othello* is a rich, textured work that admits of a vast number of divergent interpretations. The experiment is designed so that its interpretation is univocal. This is a crucial difference between art and science, but not, I think, a difference between fiction and fact. It is the density and repleteness of the literary symbols, not their fictivness, that makes the crucial difference. Thought experiments combine the freedom of fiction with the austere requirements of science. Like other scientific symbols, their interpretation should be univocal, determinate, and readily ascertained. It should be clear what background assumptions are operative and how they bear on the thought experiment's design and interpretation.

Einstein contrives a thought experiment to investigate what a person riding on a light wave would see. It teases out less than obvious implications of the finitude of the speed of light. It prescinds from such

inconveniences as the fact that a person is too big to ride on a light wave, the fact that anyone travelling at light speed would acquire infinite mass, and the fact that such a person would be unable to see since her retina would be smaller than a photon, and so on. Since such physiological impediments are irrelevant to the thought experiment, they play no role. In effect the thought experiment instructs us to pretend that someone could ride on a light wave without ill effect and to consider what he would observe. Suspension of disbelief is required to adopt the requisite imaginative stance, but what aspects of our situation we should retain and what aspects we should abandon are clear.

A thought experiment affords insight into phenomena only if the driving assumptions about what can be fruitfully set aside are correct. Otherwise, it misleads. But this is so for all experiments. Experiments using a purified sample yield insights into their natural counterparts only if we haven't filtered out significant factors. Studying the properties of a random sample yields insight into the material sampled only if the randomly taken sample is in fact suitably representative. If we randomly select an unrepresentative sample, we will project the wrong features onto the domain. All scientific reasoning takes place against background assumptions. That is the source of both its power and its vulnerability.

To construe a model as a fiction is to treat it as a symbolic construct that exemplifies features it shares with the phenomena it models but diverges from those phenomena in other, unexemplified, respects. A tinker-toy model of a protein exemplifies structural relations it shares with the protein. It does not exemplify its color, size or material. So its failure to replicate the color, size, and material of the protein it models is not a defect. Indeed, it is an asset. Being larger, color-coded, and durable, it is able to make the features it exemplifies manifest so that they can be discerned more easily than they are when we observe proteins directly.

The explanation of the cognitive contribution of fictions in science is that in recognizable and significant respects their divergence from the phenomena they bear on is negligible. I suggest that the same thing accounts for the cognitive contributions of otherwise good theories that contain anomalies. We say that they are right 'up to a point'. That point, I suggest, is where the divergence becomes nonnegligible. Just as an ensemble of gas molecules nearly satisfies the ideal gas law, the motion of a slowly moving

nearby object nearly satisfies Newton's laws. In both cases, the laws provide an orientation for investigating where, how, why, and with what consequences divergences occur. 'Negligible' is an elastic term. Sometimes we are, and should be, prepared to overlook a lot. In the early stages of theory development, very rough approximations and very incomplete models afford a modest understanding of the domain. With the advancement of science we raise our standards, refine our models, and often require a better fit with the facts. That is one way we improve our understanding of what is going on. A closer fit does not always afford a better understanding. Sometimes a stark, streamlined model that cuts through irrelevant complications is more revealing. When a point mass at the center of gravity is an effective way to conceptualize and compute the effects of gravity, a more realistic representation that specifies the actual dimensions of the planets would not obviously be preferable. The fact that in certain respects it is as if the planets were point masses is an interesting and important fact about gravitational attraction. In effect, what I am suggesting is that a theory that is known to be inadequate is consigned to the realm of fiction. It is treated as if it were an idealization. But fictions in science are cognitively significant, so to construe even our best theories as fictions is not to devalue them.

A worry remains: If the acceptability of scientific theories does not turn on their truth, the distinction between science and pseudoscience threatens to vanish. If not on the basis of truth, on what grounds are we to consider astronomy cognitively reputable and astrology bunk? The answer harks back to the previously cited passage from Quine. Although the sentences of science face the tribunal of experience only as a corporate body, they do face the tribunal of experience. Theories as a whole are answerable to empirical evidence and are discredited if they are not borne out by the evidence. Theories containing idealizations, approximations, simplified models, and thought experiments do not directly mirror reality. But because they have testable implications they are empirically defeasible. That is, there are determinate, epistemically accessible situations which, if found to obtain, would discredit the theories. If we discovered, as we could, that friction plays a major role in collisions between gas molecules, that discovery would discredit the ideal gas law and the theories that incorporate it. Pseudoscientific accounts are indefeasible. No evidence could discredit them. They cannot claim to reveal the way the world is, since they would, by their own lights, hold

regardless of how the world turns out to be. This is a critical difference and shows that scientific theories that incorporate fictive devices are nonetheless empirical.

I have urged that science is riddled with symbols that neither do nor purport to directly mirror the phenomena they concern. Purified, contrived lab specimens, extreme experimental situations, simplified models, and highly counterfactual thought experiments contribute to a scientific understanding of the way the world is. I suggested that science's reliance on such devices shows that veritism is inadequate to the epistemology of science. But, one might argue, such devices play only a causal role. They enable scientists to discover the way things are. And perhaps it is significant that non-truths can do that. Nevertheless, epistemology is not primarily concerned with the causes of our beliefs, so the use of such devices does not discredit veritism. The crucial question is whether the conclusions that emerge from the deployment of these devices are true. If so, veritism is vindicated, for the role played by the untruths is causal but not constitutive of scientific cognition.

This strikes me as wrong. The devices do not just cause an understanding of the phenomena they concern, they embody that understanding. Their design and deployment is enmeshed with an understanding of the phenomena they bear on and the proper ways to investigate it. Without that understanding the laboratory experiments, models, thought experiments and samples would not only be unmotivated, they would be unintelligible. We would have no idea what to make of them. Without some constraints on the imaginative exercise, we would have no idea what to imagine when invited to imagine what a person riding on a light wave would see. Moreover, we do not just use the devices as vehicles to generate conclusions, we think of the domain in terms of them. We represent the contents of lakes as water with impurities, the interaction of gas molecules as comporting with the ideal gas law, the orbits of the planets as perturbed ellipses. Because we do so, we are in a position to draw inferences that both test and extend our understanding.

There is a further worry: The only constraint on acceptability I have mentioned is that a theory must answer to the evidence. But a theory that included 'All planets except Mercury have elliptical orbits' would do that. Among the theories that answer to the same body of evidence, some are better than others. What

makes the difference? Unfortunately, the question cannot be settled by appeal to obvious, a priori criteria. Apart from consistency, there are none. With the advancement of understanding, we revise our views about what makes a theory good, and thus our criteria of acceptability. Elsewhere I have argued that epistemic acceptability is a matter of reflective equilibrium: The components of an acceptable theory – statements of fact, fictions, categories, methods, etc. -- must be reasonable in light of one another, and the theory as a whole must be at least as reasonable as any available alternative in light of our relevant antecedent commitments.[13] This is not the place to review that argument. My point here is that because such an epistemology does not privilege literal, factual truths, it can accommodate the complex symbolization that mature science exhibits.

To understand a theory is to properly interpret its symbols. This requires distinguishing factual from fictional sentences, accommodating tacit presuppositions, accurately interpreting the scope and selectivity of exemplars and so forth. To understand a domain in terms of a theory is to be in a position to recognize, reason about, anticipate, explain, and act on what occurs in the domain on the basis of the resources the theory supplies. Understanding thus is a matter of degree. A slight understanding equips us to recognize gross features, to give rough explanations, to reason in general terms, to form crude expectations. With the advancement of understanding our recognition, reasoning, representations and explanations become better focused and more refined.

Harvard University

[1] Herbert Spencer, *Eduction*, London: Williams and Norgate, 1940, p. 77.
[2] Alvin Goldman, *Knowledge in a Social World,* Oxford: Clarendon, 1999.
[3] For simplicity of presentation, I call the unit of science 'a theory'. On this account, models are theories or are parts of theories rather than being independent of them.
[4] W.V. Quine, 'Two Dogmas of Empiricism,' *From a Logical Point of View*, New York: Harper, 1961, p. 41.
[5] This maneuver is modeled on the 'corrected doxastic system' in Lehrer's epistemology. See Keith Lehrer, *Knowledge.* Oxford: Oxford University Press, 1974.
[6] Nancy Cartwright, *How the Laws of Physics Lie*, Oxford: Clarendon, 1983, p. 37. She cites Peter J. Bickel, Eugene A Hammel and J. William O'Connell, 'Sex Bias in Graduate Admissions: Data from Berkeley,' in William B. Fairley and Fredrick Mosteller, *Statistics and Public Policy* (Reading, Mass: Addison-Wesley, 1977).
[7] *Oxford Dictionary of Science*, Oxford: Oxford University Press, 1999, p. 141.
[8] Daniel Dennett, 'Real Patterns,' *Journal of Philosophy*, 88 (1991) p. 28.
[9] Ibid.
[10] A. Gibbard and H. R. Varian, 'Economic Models,' *Journal of Philosophy* 75 (1978) 664-677.
[11] Nancy Nersessian, 'In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling', *PSA 1992,* 2 (1993) 291-301.
[12] See my *Considered Judgment,* Princeton: Princeton University Press, 1996, pp. 180-204 and 'True Enough', *Philosophical Issues* 14 (2004) forthcoming.
[13] *Considered Judgment*, pp. 101-143.