TRUE ENOUGH

Catherine Z. Elgin

True Enough

True Enough

Catherine Z. Elgin

For Jim, Sam, and Gareth.

And in memory of Jonathan Adler.
Thou wert my guide, philosopher, and friend.
—Alexander Pope

Contents

Acknowledgments

In writing *True Enough* I amassed a multitude of debts. The first and deepest is to Jonathan Adler. Since our days in graduate school, he was my dearest friend, sharpest critic, most supportive ally. Although the book is littered with acknowledgments for his suggestions, objections, elaborations, and corrections, they do not begin to do justice to his contributions to my work. His influence is pervasive. The book would have been better if he had lived longer.

Many other friends, colleagues, and students also helped to improve the work, refining my arguments and saving me from a host of embarrassing gaffes. For their insights and advice I thank (in no particular order) Kenneth Walden, Remei Capdevila, Samuel Elgin, Mary Kate McGowan, Kareem Khalifa, Nancy Nersessian, Elizabeth Camp, Stephen Grimm, John Greco, Roman Frigg, James Nguyen, Bas van Fraassen, Georg Brun, James van Cleve, Sabine Ammon, Zachary Barber, Insa Lawler, Duncan Pritchard, Heather Battaly, Judith Jarvis Thomson, A. R. Mudd, Michael Hannon, John Kulvicki, Wayne Riggs, Stephen Yablo, Joseph Cruz, Selim Berker, Robert Schwartz, Amélie Rorty, Jamie Tappenden, Jonathan Mattheson, Alvin Goldman, Sanford Goldberg, Geordie McComb, Michael Stuart, Mauricio Suárez, and a gaggle of anonymous referees. I am sure there are others whom I have forgotten to list. For that I apologize. I am grateful to you all.

I road tested material a variety of venues. I am grateful for the comments and criticisms I received. Among the conferences where I presented drafts of different chapters are: the 2003 Rutgers Epistemology Conference, where Ernest Sosa's challenge got the whole thing started; the 2008 Pacific Division meeting of the American Philosophical Association, where worries raised by Jonathan Kvanvig, Michael Strevens, and Duncan Pritchard spurred me

Earlier versions of sections of the book have appeared in journals and anthologies: Chapter 2 builds on "True Enough," *Philosophical Issues* 14

1   Reconfiguring the Realm

Philosophy valorizes truth. There may be practical or prudential or political reasons to accept a known falsehood. But there can, it is held, never be epistemically good reasons to do so. Nor can there be good reasons to accept modes of justification that are known not to be truth conducive. Such is the prevailing consensus. Although it seems reasonable, this stance has a fatal flaw. It cannot account for the epistemic standing of science: for science unabashedly relies on models, idealizations, and thought experiments that are known not to be true. Modern science is one of humanity's greatest cognitive achievements. To think that this achievement is a fluke would be mad. So epistemology has the task of accounting for science's success. A truth-centered, or veritistic, epistemology must treat models, idealizations, and thought experiments as mere heuristics, or forecast their disappearance with the advancement of scientific understanding. Neither approach is plausible. We should not cavalierly assume that the inaccuracy of models and idealizations constitutes an inadequacy; quite the opposite. I suggest that their divergence from truth or representational accuracy fosters their epistemic functioning. When effective, models and idealizations are, I contend, felicitous falsehoods. They are more than heuristics. They are ineliminable and epistemically valuable components of the understanding science supplies.

These are bold claims. In what follows I develop a holistic epistemology that accommodates the cognitive contributions of science. It acknowledges that tenable theories must be tethered to the phenomena they concern, but denies that truth is the sole acceptable tether. It maintains that felicitous falsehoods relate to their targets via exemplification, the relation of an example to the features it is an example of. It urges that exemplification

provides at least as strong and stable a link to the phenomena as truth does. Although models, idealizations, and thought experiments are inaccurate—sometimes wildly inaccurate—they exemplify features they share with their targets and thereby afford epistemic access to aspects of their targets that are otherwise overshadowed or underemphasized. They are true enough. They advance understanding of the phenomena they bear on. To accommodate science, I advocate shifting epistemology's focus from the knowledge of individual facts to the understanding of a broader range of phenomena. In science and other systematic inquiries, relatively comprehensive constellations of epistemic commitments stand or fall together.

Once truth is dethroned, truth-conduciveness falls as well. Many epistemologists hold that arguments and methods of inquiry are justified by being reliably truth conducive. If so, in deploying them and eschewing arguments and methods that are not truth conducive, we increase our prospects of epistemic success. Having sidelined truth, I can make no such claim. This raises an important question about epistemic normativity. If truth-conduciveness does not vindicate a method or form of argument, what does? I develop a deontological account of epistemic norms. I argue that acceptability turns on epistemic responsibility rather than on reliability. I provide criteria for objectivity that are satisfied by representations that respect deontological epistemic norms. Worries about subjectivity fade, but there is, even in principle, no guarantee that conclusions that are backed by objective reasons are true. What a tenable account does guarantee is that conclusions are reasonable and defensible in the epistemic circumstances.

The position that emerges not only secures the epistemic standing of science, it accords the same status to art. For the symbolic devices that underwrite the acceptability of felicitous falsehoods in science are also deployed in the arts. Thus, I urge, the arts embody and advance understanding if the sciences do. Throughout the book, discussions of the arts and the sciences intertwine. My position vindicates the oft expressed but seldom justified claim that, although works of art are not true or even truth apt, we learn something important from our encounters with art.

In places, I frame my position by criticizing ways others address the same issues. I make no attempt to provide a comprehensive survey of the available alternatives. I discuss other approaches only when and only to the extent that the juxtaposition illuminates the points I seek to make. Nor do my criticisms purport to be refutations. Like Quine (1961), I believe that we

can hold fast to any contention come what may, so long as we are willing to make compensatory adjustments elsewhere in our system of thought. One can, for example, continue to insist that scientific theories are unacceptable unless true, or that to understand is to be able to explain, or that science is cognitive but art is not. But, I will urge, the cost of such tenacity is surprisingly high. It is high enough to make it reasonable to entertain my alternative.

The shape of the book is as follows: I begin by arguing that if epistemology is to accommodate science, it must weaken its commitments to truth and truth-conduciveness. After examining a variety of widely accepted scientific models and showing that standard veritistic tweaks, such as construing them as approximations, fail, I introduce the idea of a *felicitous falsehood*—an inaccurate representation whose inaccuracy does not undermine its epistemic function. I do not contend that we should *believe* felicitous falsehoods. Rather, building on L. Jonathan Cohen (1992), I maintain that we should *accept* them. This is a matter of being willing and able to use them as a basis for inference or action when our overarching interests are cognitive.

To accommodate the fruits of science and other systematic inquiry, I urge, epistemology should shift its center of gravity from knowledge to understanding—in particular, to what has been called *objectual understanding*, the understanding of a topic or subject matter (Kvanvig, 2003). Throughout the book, that is what I shall mean when I use the word 'understanding' without qualification. Understanding is not mirroring. It is neither a matter of believing a host of separately justified individual propositions nor a matter of believing a network of interconnected propositions that mirror the facts. To understand a topic involves knowing how to wield one's commitments to further one's epistemic ends. It involves being able to draw inferences, raise questions, frame potentially fruitful inquiries, and so forth. Since some of the commitments are likely to be felicitous falsehoods and others to be methodological or normative commitments that are not truth apt, understanding is not factive. It is not a type of knowledge; it does not consist exclusively or primarily in believing or accepting truths.

Because the commitments that comprise an understanding are mutually supportive, understanding must be construed holistically. I show how the mutual support of independently doubtful items can increase their credibility. Drawing on my previous work, I argue that an understanding of a topic

consists in accepting a system of commitments in reflective equilibrium (Elgin, 1996). A network of commitments is in reflective equilibrium when each of its elements is reasonable in light of the others, and the network as a whole is as reasonable as any available alternative in light of our relevant previous commitments. Even if some components would be doubtful in isolation, collectively they constitute an interwoven tapestry of commitments that we can on reflection endorse.

What differentiates a felicitous falsehood from an infelicitous one? If truth and truth-conduciveness are sidelined, what replaces them? Rather than being epistemic reliabilists, I urge, we should be epistemic responsibilists. Acceptable epistemic norms are norms that would emerge from the deliberations of suitably idealized epistemic agents. Extrapolating from Kant (1981), I urge that such agents are legislating members of a realm of epistemic ends. The realm of epistemic ends is an idealization of real-life communities of inquiry that eliminates factors that distort actual deliberations. I identify and justify the constitution of the realm, the factors legislators can draw on, and the conditions on their deliberations. I argue that legislators must be, in a political sense, free and equal. They must be free to advance any hypothesis they consider viable and equally entitled to be heard. This does not mean, of course, that they must be equally intelligent or that their views are to be treated as equally cogent or equally worthy of acceptance. It only means that views are not to be rejected without a fair hearing. The political requirements on a realm of epistemic ends fix the framework within which the distinction between good and bad reasons, methods, strategies, and conclusions can be responsibly drawn.

I go on to argue that intellectual integrity requires that certain values—those that figure in or contribute to trustworthiness—are simultaneously moral and epistemic. Here again I focus on science. Since nature is enormously complicated, epistemic agents need to work together to achieve an understanding of it. Since even the best methods are not failsafe, the resources for identifying and correcting received errors must be readily available. Scientific misconduct undermines the enterprise by producing results that are untrustworthy. Because the value of trustworthiness is integral to science—because a scientist needs to trust the findings she builds on—I argue that in teaching what good science is and how good science is done, science education typically imbues novice scientists with certain moral/epistemic virtues. Other disciplines do the same for their practitioners.

As I've construed it, epistemic responsibilism may seem too weak to secure objectivity. Can a discipline claim to be objective if it recognizes that there is an ineliminable human element in whatever vindicates its claims? I argue that the sort of objectivity that epistemology should seek is not that delivered by the 'view from nowhere' (Nagel, 1986) or the 'absolute conception of reality' (Williams, 1978). It is an objectivity that protects against idiosyncrasy, bias, and chance. Drawing on van Fraassen (2008), I argue that it is often perspectival. I discuss two modes of objectivity. One is impersonal. We define magnitudes and develop ways of measuring them that omit the personal element entirely. Or anyway, they eliminate the personal or interpersonal element once the work of defining magnitudes and developing measurement procedures is done. The other is impartial. Although someone must be doing the assessing, it makes no difference who that (suitably trained) someone is. The results of both are trustworthy and would be endorsed by legislating members of a realm of epistemic ends.

Having argued for a holistic account of understanding and explicated the norms that understanding is subject to, I return to the question of what vindicates felicitous falsehoods. I urge that a falsehood or inaccurate non-propositional representation is felicitous only if it exemplifies features it shares with the phenomena it bears on. By making those features salient, such a representation enables us to appreciate their significance for the phenomena. I explicate exemplification and show how it figures in ethics, mathematics, and dance.

I go on to argue that laboratory experiments, thought experiments, and works of fiction also advance understanding by means of exemplification. Because laboratory experiments involve pure samples tested in isolated, pristine environments, they distance themselves from the natural phenomena they bear on. They eliminate or control for irrelevant features in order to highlight relevant ones. Thought experiments involve further distancing. They contrive imaginative scenarios where situations that are not realizable in fact can be entertained and their implications investigated. This enables them to bring to the fore factors that are typically overshadowed. Literary fictions, I urge, are elaborate, extended thought experiments. In all three cases, delicacy of discrimination is needed to determine which features are exemplified and how they project back onto the facts.

Effective models exemplify features they share with their targets. They streamline, eliminating irrelevant factors so that features of interest stand

out. This enables them to reveal occluded aspects of their targets and to dis-
close how seemingly disparate phenomena are alike. Evolutionary game
theory, for example, shows that, from the perspective of natural selection,
predator—prey relations and reciprocal altruism are the same sort of thing.
Following Hughes (1996), I urge that a model represents its target as having
the feature the model exemplifies. This, I argue, is analogous to what occurs
in portraiture.

If legislating members of realms of epistemic ends set the constraints
that bind them, how rigid are those constraints? What prospect is there for
interdisciplinary understanding? I suggest that the restrictions set by dif-
ferent disciplines set strong presumptions but not rigid requirements. This
allows for epistemic cross-fertilization. We can develop epistemic hybrids
that bridge disciplinary divides.

Epistemology is haunted by the worry is that despite our best efforts
we still might be wrong. Nothing in my position immunizes against this
worry. The most I can say is that if a consideration is multiply tethered, it
is relatively secure. That is far from a guarantee. The permanent possibility
of error leads many to embrace fallibilism. This is not so straightforward as
it might appear. The fallibilist is burdened by a version of Moore's paradox
(Moore, 1991). 'I know that p, but I might be wrong that p' has the air
of giving one's word in one breath and taking it back in the next. But if
we repudiate fallibilism, and insist that we should dogmatically endorse
a true conclusion that reaches the threshold for acceptability (wherever
that threshold is set), we confront Kripke's (2011a) dogmatism paradox: If I
know that p, then p is true. In that case, any evidence that $\sim p$ is misleading. I
ought therefore to disregard any evidence that $\sim p$, since if I take it seriously,
I am apt to lower my credence and lose my knowledge. That being so, when
I know, I ought to be dogmatic. I ought to turn a deaf ear to evidence that
tells against p. But to do so seems epistemically irresponsible.

Taken together, the two paradoxes strike me as serious problems for a
theory of knowledge. They do not undermine the possibility of knowledge.
Neither individually nor jointly do they point toward skepticism. But they
undermine the value of knowledge. Concessive knowledge, which tolerates
Moore's paradox, and dogmatic knowledge, which deafens the knower to
objections, both undermine the right to be sure. An epistemic agent would
rightly be reluctant to accept testimony or to take a big risk on the basis
of claims to knowledge that were accompanied by Moore's paradoxical or

Kripke's paradoxical riders. But, I argue, fallibilism with respect to understanding does not face these difficulties. It is not a doctrine, but a stance— indeed, I urge, an epistemically valuable stance. Recognizing where and how we might be wrong discloses something significant about the fabric of our understanding. Possibilities of error, then, are not merely—and perhaps not mainly—vulnerabilities. They are avenues of insight into our epistemic situation.

2 Distancing from Truth

Veritism

In epistemology, the conviction that truth is vital is virtually axiomatic (see David, 2005). Although practical or prudential considerations might favor accepting an acknowledged falsehood, most epistemologists hold that it is never epistemically legitimate to do so. Call any position that takes truth to be necessary for epistemic acceptability *veritism* (see Goldman, 1999). It seems prima facie reasonable. But, I will argue, veritism is unacceptable. For, if we accept it, we cannot do justice to the epistemic achievements of science. Truth ought not be our paramount epistemic objective.

I begin this chapter with a sketch of veritism. My discussion is deliberately schematic because the weakness I focus on is endemic to truth-centered epistemology, being shared by internalist, externalist, virtue-theoretic, and knowledge-first positions. Too strong an emphasis on truth prevents epistemology from accommodating the cognitive contributions of science. But veritistic commitments run deep. Abandoning them requires radical revisions. In particular, I urge that belief, assertion, and knowledge should be sidelined in favor of acceptance, profession, and understanding. I show how this reconfigures the epistemic terrain, making available resources that veritistic positions lack. I go on to sketch a variety of devices that figure in science and philosophy, devices whose epistemic value comes to the fore if we distance epistemology from an inviolable commitment to truth.

According to veritism, truth-conduciveness is the appropriate standard of assessment for epistemic policies, practices, and their products. This holds for all epistemic achievements—knowledge, understanding, wisdom, know-how. The epistemic value of methods for obtaining and evaluating

information, and the criteria used in such evaluations are purely instrumental. They are justified only because and only to the extent that they are truth-conducive. Replicability of scientific results, for example, is epistemically valuable only if replicable results are more likely to be true than irreplicable ones. Should it turn out that replicable results are no more likely to be true than otherwise well supported irreplicable results, replicability would have no epistemic value. As Berker notes, the shape of this argument is familiar. Veritism, he argues, is a form of consequentialism. It differs from more familiar consequentialist positions in that the consequence it wants to promote is true belief or a preponderance of true belief over false belief, not a preponderance of, as it might be, pleasure over pain or utility over disutility (Berker, 2013). If veritists are right, then at least insofar as our ends are cognitive, we should accept only what we consider true, take pains to ensure that the claims we accept are in fact true, and promptly repudiate any previously accepted claims upon learning that they are false.

Following William James (1951), many veritists hold that our overriding cognitive objective is to believe as many truths as possible and disbelieve as many falsehoods as possible (e.g., Alston, 1988, 258; Goldman, 1999, 5; Lehrer, 1986, 6; BonJour, 1985, 9). This is the objective with respect to each and every individual contention. It is, they say, never legitimate to accept an acknowledged falsehood, even if so doing would in the long run result in the acquisition of more or a higher proportion of true beliefs than the repudiation of that falsehood would (Firth, 1981).

As stated, Jamesian veritism is remarkably tolerant. It does not discriminate among true (or false) beliefs. Insofar as our goals are purely cognitive, it says, we should amass as many truths and as few falsehoods as possible. This is easily done. Take any number. Add two. Do it again. And again. And again. With each iteration you get another truth. Your risk of accepting a falsehood along the way is vanishingly small. Should you prefer to restrict yourself to the empirical realm, take any trivial truth, such as 'Cats don't grow on trees'. Add a disjunct. (Its truth-value doesn't matter.) Add another disjunct. And another. You now have numerous true beliefs. This is far easier and far less risky than, for example, mounting a carefully controlled, scrupulously executed, theoretically grounded experiment. Nevertheless, running the experiment is epistemically more valuable. It won't do to insist that the experiment is likely to be more fruitful, at least if the only available way to measure fruitfulness is to count the number of true conclusions or the

proportion of true over false conclusions that can be drawn. Every proposition has infinitely many consequences—indeed, infinitely many obvious consequences—so the fruits of trivial inferences and those of the well-run experiment are on a par. And the automaticity of the procedure for generating trivial bits of knowledge is efficient enough to dwarf the number of truths one could discover by devising and running experiments.

As many epistemologists recognize, this suggests that the Jamesian goal needs to be fine-tuned. Rather than considering all true beliefs on a par, perhaps epistemology should distinguish between significant and trivial truths, or between truths obtained by significant and trivial means. Then the Jamesian veritist could insist that our goal is to believe as many significant truths as possible and to disbelieve as many significant falsehoods as possible, or perhaps to believe as many significant truths as possible and to disbelieve as many falsehoods as possible, whether or not they are significant (since it seems epistemically objectionable to believe falsehoods even if they are trivial). Alternatively, he could hold fast to the veritist goal as a purely cognitive end, but recognize that as finite beings with limited resources, we need, for practical reasons, to prioritize some truths over others.

Not all veritists are Jamesians. Pritchard focuses on a position he calls 'Epistemic Value T-Monism', the view that "true belief is the sole fundamental epistemic good" (2010, 14). For the T-Monist, the Jamesian concern with proportionality drops away or becomes of merely instrumental value. We should, on the T-Monist view, worry about false beliefs only if harboring them somehow interferes with our amassing true beliefs. Sosa (2007) credits only aptly formed true beliefs—those that emerge from the proper use of our epistemic competences.

Treanor goes further. He disputes the egalitarianism inherent in the Jamesian position, arguing that it rests on "the implicit assumption that any two truths contribute the same amount of truth to the balance sheet" (2013, 599). This, he claims, is incorrect. Rather than being concerned with how many truths someone believes, we should be concerned with how much truth he believes. And the latter is not a function of the former. Trivial truths, Treanor argues, contribute little to a person's overall body of true belief; significant truths contribute a great deal. Undermining the idea that we need to (or even that we can) count and aggregate true beliefs sidelines the triviality worry. Still, Treanor remains a veritist. Like his Jamesian opponents, he thinks that truth is epistemically central.

So does Hetherington (2011). He maintains that we should take the locu-
tion 'knowing better' seriously, and recognize that knowledge admits of
degrees. If an epistemic agent's belief that p is true, she knows that p in a
minimal way; if she has some justification for p, she knows better that p;
and if she can defeat skeptical challenges to p, she knows perfectly well that
p. But even the most minimal sort of knowledge requires truth. Epistemic
gradualism does not solve our problem.

Most veritists take knowledge to be the paradigm of epistemic success,
where knowledge is nonfortuitously justified or reliable true belief. Some
consider knowledge valuable for its own sake. Others think that the valu-
able aspect of knowledge lies in its embodying true belief. Either way, the
reason to favor justified beliefs (those backed by sufficient evidence) and
safe beliefs (those that could not easily have been false) is that they are
liable to be true. Justification and safety are of merely instrumental value.
Since propositions are the contents of beliefs and the bearers of truth-values,
they are what is known.[1] If this is right, the sort of justification of inter-
est to epistemology is in the first instance the justification of individual
propositions. To justify a given proposition is either to infer it from already
justified propositions or to show how belief in it emerges from reliable
belief-forming mechanisms. S is justified in believing p on the basis of q,
and q on the basis of r, and so on. On some accounts the sequence ends in
propositions that need no further justification (Fumerton, 2010); on others,
the sequence is infinite (Klein, 1999).

Holists are apt to consider this picture misleading. They maintain that
epistemic acceptability is, in the first instance, acceptability of a fairly com-
prehensive theory, or system of thought—a constellation of mutually sup-
portive commitments that bear on a topic. Let us call such a constellation
an account.[2] It consists of contentions about a topic and the reasons adduced
to support them, the ways they can be used to support other contentions,
and higher-order commitments that specify why and how evidence sup-
ports them. It also contains normative and methodological commitments
specifying the suitability of categories, the criteria of justificatory adequacy,
and the ways to establish that the criteria have been met. An account's
priority over its components is epistemic, not historical. There is no con-
tention that people come to accept an account before coming to accept the
various items that comprise it. Rather, the claim is that, regardless of the
order in which they are acquired, commitments are epistemically justified

only when they coalesce to constitute a tenable account. An account as a whole is tenable just in case it is as reasonable as any available alternative in light of the current relevant antecedent commitments (Elgin, 1996). The acceptability of individual claims, as well as of methods and standards, is derivative, stemming from their role in a tenable account. Perhaps a linear model of justification works for the belief that the cat is on the mat, but disciplinary understanding is holistic. The claim that electrons have negative charge is justified only to the extent that the scientific account that embeds it is.

As standardly construed, knowledge is granular: it comes in discrete bits. The objects of knowledge are individual facts, expressed in true atomic propositions or stated in true atomic declarative sentences. Judy knows (the fact) that the bus stops at the corner; George knows (the fact) that cows eat grass. Each grain of knowledge is supposed to answer to a separate fact. An epistemic agent acquires knowledge by amassing grains. To be sure, one grain can support another. Much of our knowledge is a product of inference from other things we know. But the unit of knowledge is an individual sentence, belief, or proposition—something that apparently bears a truth value in isolation. The commitment to granularity is manifest in the widespread adherence to Jamesian veritism. Indeed, the Jamesian goal is unintelligible unless we suppose that knowledge comes in discrete, countable bits (Treanor, 2013).

Disciplinary understanding is not an aggregation of separate, independently secured statements of fact; it is an integrated, systematically organized account of a domain. There is no prospect of sentence-by-sentence verification or sentence-by-sentence justification of the contentions that comprise a disciplinary account, for most of them lack separately evaluable consequences. Independent of an account of heat transfer, nothing could count as evidence for or against the claim that a process is adiabatic. Independent of an ethological theory, nothing could count as evidence for or against the claim that grooming behavior manifests reciprocal altruism. Collectively the components of an account are apt to have testable implications; separately they do not. In Quine's words, they "confront the tribunal of sense experience not individually but only as a corporate body" (1961, 41). Nor is the problem merely epistemic. Some scientific statements evidently lack truth-values in isolation. If the individuation of the items they purport to refer to—a species, a retrovirus, or a lepton, for example—is provided by

an account or cluster of accounts, there may be no fact of the matter as to whether such statements are true independent of the epistemic adequacy of the account or cluster that characterizes them.

If it is to accommodate the cognitive contributions of the disciplines, epistemology should be holistic, construing the primary unit susceptible of justification to be a more or less comprehensive account of a topic rather than an individual proposition. Elements of such an account then are justified largely by their contribution to a coherent, relatively comprehensive body of thought. 'Knowledge' is usually taken to pertain to discrete propositions. An epistemic agent knows that *p*. Tenable accounts are said to afford an understanding of their topics. This suggests that epistemology should shift its focus from knowledge to understanding, or at least broaden its focus to comprehend understanding as well as knowledge. By itself, this should not spark panic among epistemologists. But, I maintain, a greater revision is called for. To accommodate science, epistemology must relax its commitment to truth. I will argue that the relation between truth and epistemic acceptability is both more tenuous and more circuitous than is standardly supposed. It is often epistemically responsible to prescind from truth to achieve more global, and more worthy cognitive ends.

At first blush, this looks mad. To retain a commitment to a falsehood merely because it has other epistemically attractive features seems the height of cognitive irresponsibility. Allegations of intellectual dishonesty, wishful thinking, false consciousness, or worse immediately leap to mind. But science routinely transgresses the boundary between truth and falsity. It smooths curves and ignores outliers. It develops and deploys simplified models that diverge, sometimes considerably, from the phenomena they purport to represent. It subjects artificially contrived lab specimens to forces not found in nature. It resorts to thought experiments that defy natural laws. Even the best scientific accounts are not true. Not only are they plagued with anomalies and outstanding problems, but where they are successful, they rely on lawlike statements, models, and idealizations that are known to diverge from the truth. Veritism might reasonably be undaunted by the existence of anomalies and outstanding problems, since they are readily construed as defects. The more serious problem comes with the laws, models, and idealizations that are acknowledged not to be true but that are nonetheless critical to, indeed at least partially constitutive of, the understanding that science delivers.

Consider the problem that models pose. The ideal gas law, $pV = nRT$, represents the relation between pressure, volume, and temperature in a gas comprised of dimensionless, spherical molecules that exhibit no mutual attraction. There is no such gas; indeed, if our fundamental theories are even nearly right, there could be no such gas. All material objects have spatial dimensions. Being subject to gravity, all attract one another. No molecules are spherical. Nonetheless, the ideal gas model is integral to thermodynamics. To be sure, there are more refined models, such as the van der Waals equation and the virial equation, that incorporate some of the features of actual gases that the ideal gas model omits; but they too are idealizations, not accurate representations of any actual gas. Yet these models, which are true of nothing real, figure in a genuine understanding of how actual gases behave.

Thought experiments should also vex the veritist. They are not actual experiments and often not even possible experiments. Einstein demonstrated the equivalence of gravitational mass and inertial mass by inviting scientists to compare the experience of a person riding in an elevator in the absence of a gravitational field and her experience while at rest in the presence of a gravitational field. Lacking external cues, she could not tell which situation she was in. To actually run the experiment we would need to place an unconscious subject in a windowless enclosure, send her to a region of outer space distant from any significant source of gravity, restore her to consciousness, and query her about her experiences. This is morally, practically, and physically unfeasible. It is evidently also unnecessary. Physicists take the thought experiment by itself to establish the equivalence.

Far from being defects, models, idealizations, and thought experiments figure ineliminably in successful science. If truth is mandatory, much of our best science turns out to be epistemologically unacceptable and perhaps intellectually dishonest. Our predicament is this: We can retain the truth requirement and construe science either as cognitively defective or as noncognitive, or we can reject, revise, or relax the truth requirement and remain cognitivists about and devotees of science.

I take it that science provides an understanding of the natural order. By this I do not mean merely that an ideal science *would* provide such an understanding or that at the end of inquiry science *will* provide one, but that much actual science has done so and continues to do so. I take it, then, that much actual science is cognitively reputable—indeed, estimable.

So an adequate epistemology should explain what makes good science cognitively good. Too strict a commitment to truth stands in the way. Nor is science the only casualty. In other disciplines such as philosophy, history, political science, and economics, as well as in everyday discourse, we often convey information and advance understanding by means of sentences and other representations that are not literally true. An adequate epistemology should account for these as well.

A tenable account is a tapestry of interconnected commitments that collectively constitute an understanding of a domain. My thesis is that some representations that figure ineliminably in tenable accounts make no pretense of being true, but are not defective on that account. Indeed, I will argue, their deviations from truth are epistemically valuable—often more valuable than the unvarnished truth about the phenomena would be. If I am right, tenable accounts and the understandings they embed have a more intricate symbolic structure than epistemologists standardly suppose. Nevertheless, I do not think that we should jettison concern for truth completely. The question is what role a truth commitment should play in a holism that recognizes a multiplicity of sometimes conflicting epistemological desiderata.

Stepping Back

Consensus has it that epistemic acceptability requires something like nonfortuitously justified and/or reliable, true belief. The justification, reliability, and belief requirements involve thresholds. The connection between the belief's being true and its being justified or reliable should not be due to (the wrong kind of) luck (see Pritchard, 2010). But truth is supposed to be an absolute matter. Either a belief content is true or it is not. I suggest, however, that the so-called truth requirement on epistemic acceptability involves a threshold too. I am not saying that truth itself is a threshold concept. Perhaps such a construal of truth would facilitate treatments of vagueness, but that is not my concern. For my purposes, classical bivalence is acceptable. Either a sentence, belief, or proposition is true or it is false. My point is that epistemic acceptability turns not on whether it is true, but on whether it is *true enough*—that is, on whether it is close enough to the truth. "True enough" obviously has a threshold.

This raises a host of questions. But before addressing them, let me issue a disclaimer. I do not deny that (unqualified) truth is an intelligible concept or a realizable ideal. We readily understand instances of the (T) schema:

'Snow is white' is true ≡ snow is white

'Power corrupts' is true ≡ power corrupts

'Neutrinos have mass' is true ≡ neutrinos have mass

and so on. So long as it evades the paradoxes, a disquotational theory of truth suffices to show that the criterion expressed in Convention (T) can be satisfied. One might, of course, want more from a theory of truth than satisfaction of Convention (T); but to make the case that the concept of truth is unobjectionable, any minimalist theory that does not engender paradox suffices. Moreover, we can often tell whether a sentence is true. We are well aware not only that

'Snow is white' is true ≡ snow is white,

but also that

'Newly fallen snow is white' is true.

The intelligibility and realizability of truth, of course, show nothing about which sentences are true or which truths we can discover. Nevertheless, as far as I can see, nothing about the concept of truth discredits veritism. Since truth is a self-consistent, intelligible concept, it is open to epistemology to insist that only truths are epistemically acceptable. Since truth is a realizable objective, such a stance does not lead inexorably to skepticism. I do not deny that veritism is an available epistemic stance. But I think it is an unduly limiting one. It prevents epistemology from accounting for the full range of cognitively estimable achievements.

If epistemic acceptance is construed as belief and epistemic acceptability as knowledge, the truth requirement seems reasonable (see Wedgwood, 2002). For cognizers like ourselves, there seems to be no epistemically significant

gap between believing that *p* and believing that *p* is true. Ordinarily, upon learning that his belief that *p* is false, an epistemic agent ceases to believe that *p*. Moreover, he considers it epistemically obligatory to do so. One ought to believe only what is true. If he cannot manage to divest himself of the belief, he considers his failure a defect in his epistemic character. Perhaps a creature without a conception of truth can harbor beliefs. A cat, lacking the capacity for semantic ascent, might believe that there is a mouse in the wainscoting without believing that 'There is a mouse in the wainscoting' is true.[3] In that case, the connection between believing that *p* and believing that *p* is true is not exceptionless. But whatever holds for cats, it does not seem feasible for any creature that has a conception of truth to believe that *p* without believing that *p* is true. If epistemic acceptance is a matter of belief, acceptance is closely linked to truth (Adler, 2002). So is assertion. Although asserting that *p* is not the same as asserting that *p* is true, it seems plain that one ought not to assert that *p* if one is prepared to deny that *p* is true or to suspend judgment about whether *p* is true; nor ought one assert that *p* is true if one is prepared to deny that *p* or to suspend judgment about whether *p*. Assertion and belief, then, seem committed to truth. Knowledge shares that commitment (Williamson, 2000). Whether or not knowledge is equivalent to justified or reliably generated true belief, once someone discovers the falsity of something she took herself to know, she withdraws her claim to knowledge. She says, 'I thought I knew it, but I was wrong', not 'I knew it, but I was wrong'.

Being skeptical about analyticity, I do not contend that a truth commitment is part of the meaning of 'belief', 'assertion', and 'knowledge'. Nor, for my purposes, does it matter whether belief is more primitive than knowledge or vice versa. Whatever the details, the truth commitment tightly intertwines with our views about belief, assertion, and knowledge. It seems best, then, to retain that connection and revise epistemology by making compensatory adjustments elsewhere. Once those adjustments are made, knowledge and belief turn out to be less central to epistemology than we standardly think.

I do not then claim that it is epistemically acceptable to *believe* what is false or that it is linguistically acceptable to *assert* what is false. Rather, I suggest that epistemic acceptance is not restricted to, and does not always involve, belief. L. Jonathan Cohen distinguishes between acceptance and belief. With some modifications, I deploy his distinction. According to

Cohen, "belief that *p* is a disposition, when one is attending to issues raised or items referred to by the proposition that *p*, normally to feel that it is true that *p* and false that *not-p*, whether or not one is willing to act, speak, or reason accordingly. . . . To accept that *p* is to have or adopt a policy of deeming, positing, or postulating that *p*" (1992, 4). The term 'belief' is ubiquitous in epistemology. To avoid (or at least minimize) confusion, I will use the term 'conviction' for what Cohen calls 'belief'. To be *convinced* that *p* is to be disposed, when attending to issues raised or items referred to by *p*, normally to feel that it is true that *p* and false that ~*p*. To *accept* that *p* involves being willing to take *p* as a premise, as a basis for action or, I add, as an epistemic norm or a rule of inference, when one's ends are cognitive. This includes being willing to give others to accept that *p* via testimony. Acceptance is not a disposition to represent, but a disposition to act. I impose two restrictions: first, I am concerned only with cases where our ends are cognitive. So I set aside cases where acceptance is purely prudential, practical, self-destructive, or whimsical. Second, I restrict acceptance to assertoric—that is, non-reductio—inferences, and to actions that are not intentionally self-defeating (these being the practical counterparts of reductio inferences). I grant that reductios are cognitive and that in making a reductio argument we temporarily accept a premise in order to demonstrate that it is unacceptable. But reductios are not the sorts of inferences that concern me, since acceptance-for-reductio is too short-lived to serve my purposes.

Something more is required. An epistemic agent who accepts a consideration is not just *willing* to deploy it when her ends are cognitive; she is also *able* to do so. Someone who, on the basis of testimony, believed that modus ponens is a valid rule of inference, but had no idea how or when to apply it, would not be in a position to accept modus ponens. In my usage, then, to accept that *p* is to be willing and able to take *p* as an assertoric premise, epistemic norm or rule of inference in one's reasoning or as a basis for action when one's ends are cognitive. To reject that *p* is to accept that not-*p*. To withhold that *p* is to fail to accept that *p* and fail to accept that not-*p*. *Acceptance* is action oriented in a way that conviction per se is not. In what follows I will continue to use the term 'belief' as it is standardly used, to cover cases where conviction and acceptance align.

Although there is considerable overlap, acceptance and conviction can diverge. One can be convinced of propositions that in a given context one does not accept. All snakes in Australia are venomous. So an Australian

might find that she can't help but feel that every snake she sees is venomous. She is convinced. Nevertheless, she may train herself to accept that the garter snakes she encounters in Maine are harmless. She does not call the animal control officers each time she sees a garter snake slither across the campground. Conversely, people often accept things of which they are not, or not yet, convinced. They accept propositions for the purposes of argument or as a working hypothesis, and draw out their consequences. Judah Folkman suspected that cancer can be suppressed by inhibiting angiogenesis in tumors. If tumors lack an adequate blood supply, he reasoned, they will be unable to grow. Initially, this was only a suspicion. He accepted it as a working hypothesis, drew inferences, and designed experiments based on it. When the long series of experiments yielded the results he anticipated, he became convinced. But the acceptance was earlier, and independent of the conviction (Folkman, 1996).

The term 'working hypothesis' is multiply apt here. A working hypothesis is a hypothesis that a scientist works with; it is not an idle hunch. It does some work for him, framing and focusing his investigation. And it is accepted only so long as it works—so long, that is, as it is epistemically fruitful. I contend that acceptance rather than conviction is epistemologically central. The Australian camp counselor is saddled with a conviction that she recognizes is unwarranted. It is too ingrained for her to easily give up. This is unfortunate. But if she manages to bracket it and accept that many North American snakes are harmless, she is not epistemically defective.

The shift from belief to acceptance requires an additional shift, for belief is closely tied to assertion. Since I contend that much of our cognitively responsible verbal outputs concerns acceptance rather than belief, I cannot comfortably hold that that output consists in assertions. I maintain that uttering or inscribing seriously and sincerely for cognitive purposes—call it 'professing'—is not limited to asserting. We regularly reason verbally within the parameters of accounts we are not convinced of. Atheist philosophers profess 'God exists and is not a deceiver' in discussions of Descartes. Nominalists profess the Theory of Forms in discussions of the divided line. Physicists profess '$F = ma$' in discussions of slow-moving middle-sized physical interactions, even though they consider Newton's laws to have been superseded.

Some might maintain that what I have called professing is just a matter of pretending. On such a view, in discussing the Sixth Meditation, I pretend

that I am convinced by Descartes's proofs for the existence of God and reason within the pretense. In discussing the motion of the billiard ball, I pretend that F = ma. But professing is not the same as pretending. Moore's paradox is an assertion of the form '*p* but I do not believe that *p*'. To assert that *p* is to put it forth as something one believes. To assert that one does not believe that *p* is to deny that one believes that *p*. So there is something absurd or self-defeating about asserting a Moore's paradoxical sentence. The paradox cannot be avoided by retreating to pretense. An actor on stage saying, 'Buckingham is loyal to the king' pretends to be asserting that Buckingham is loyal to the king. But if, without breaking character in the middle of his speech, the actor were to say, 'Buckingham is loyal to the king but I do not believe that Buckingham is loyal to the king', his utterance would be as absurd and self-defeating as a standard Moore's paradoxical assertion. Similarly, a child having a make-believe tea party might pretend that her toy dog Mopsy says, 'I want another cookie'. But if she were to pretend that Mopsy says, 'I want another cookie, but I do not believe I want another cookie', her pretend-utterance would be equally absurd and self-defeating (Szabó, 2001).

To pretend to assert that *p* is to act as if one had asserted that *p*. But to profess that *p* is to do something different. It is to make *p* available to function as a premise or rule of inference in a given context for a given cognitive purpose. If one is professing rather than asserting, there is nothing absurd about saying, 'God exists and is not a deceiver but I do not believe that God exists and is not a deceiver'. In fact, it may be a good way of signaling that 'God exists and is not a deceiver' is professed but not asserted in a particular context. As Cohen (1992, 72–73) argues,

where an utterance of 'It is raining' implies acceptance that it is raining but not belief that it is raining, there should be no feeling of oddness or anomaly. And in fact it is not at all difficult to imagine such cases with more complex structures as in

"All right. Your arguments from economic theory are unanswerable and I have to concede your point. Index-linked wage settlements are inflationary. But, although I am bound to accept this, and everything that follows from it, I still don't really *believe* [= I am still not really convinced] that index linked wage settlements are inflationary."

As so often in linguistic analysis, a larger slab of discourse constrains us to suppose a speech-act is different from the one that a smaller one suggests.

The close connection between the conviction that p and the conviction that p is true explains why conviction and belief are in general impervious to the will (Williams, 1973).[4] A conviction is true only if things are the way its content represents them to be. Fred's conviction that it is raining is true only if it is in fact raining. But if his conviction were responsive to his will—if, that is, wishful thinking were an effective mechanism for forming convictions—by willing, wishing, or desiring strongly enough, he could convince himself that it is raining even if it is not, and even if he as ample evidence that it is not. Self-deception is real, so the imperviousness in question is not absolute. But self-deception is difficult to sustain. For it is hard to retain a conviction that flies in the face of one's evidence, while continuing to hold one's other convictions firm (see Adler, 2002). Normally we find ourselves convinced of one thing or another. Sometimes, to be sure, we become convinced through arduous investigations or laborious chains of reasoning. But even in these cases, the fruits of our cognitive labors seem to simply emerge from the labor. They do not present themselves as a matter of choice.

This creates a conundrum for epistemology. On the one hand, we do not want to claim that it is up to an epistemic agent whether to believe a proposition for which she has and recognizes that she has sufficient evidence. On the other hand, epistemology is a normative discipline. It is given to issuing pronouncements about what epistemic agents should believe in various circumstances. If we are passive with respect to belief, then the 'should' seems impertinent. If Fred can't help but believe what he does, it's no use telling him that he should believe something different (or, for that matter, telling him that he should keep believing what he does). What he believes is not up to him.

Here the distinction between conviction and acceptance is helpful. Conviction is passive. As Fred surveys the arid landscape, he can't help but be convinced that it is not raining, however much he might wish it were. But acceptance involves agency. Even if it is impertinent to criticize epistemic agents for convictions they cannot help but harbor, it is not impertinent to criticize them for the premises and rules of inference that they accept. For acceptance is voluntary. And it is reasonable to ask whether, or to what extent, the acceptance of a given premise or rule furthers the agent's cognitive objectives. Fiona can decide to accept the

proposition that Amherst is 90 miles from Boston, and use it as a basis for reasoning about Massachusetts geography. In so doing, she need not think the proposition is true. She need only think it is true enough for her current purposes. Rather than assuming that as a matter of fact Amherst is exactly 90 miles from Boston, she deems the distance close enough to 90 miles that it does no harm to identify the actual distance, whatever it is, with 90 miles. To identify *a* with *b* is not to hold that *a* is identical to *b*. It is to hold that in the given circumstances it does no harm to treat *a* and *b* as identical (Goodman, 1977).

This is not to suggest that any consideration whose acceptance would cause one to realize one's cognitive objectives or increase the probability of doing so is thereby acceptable. But a consideration whose integration strengthens the tenability of an account has, for that reason, a claim to acceptability.

Felicitous Falsehoods

Understanding is often couched in and conveyed by symbols that are not true. Many of them do not even purport to be true. Such symbols are epistemically felicitous falsehoods.[5] Not all scientific models are propositional. Besides equations and verbal models, science is rife with diagrams, such as harmonic oscillator depictions in physics texts, three-dimensional models such as tinker-toy models of proteins, real-life simulations, and computer-based simulations. Nonpropositional models and diagrams are not, strictly speaking, false. But if interpreted as realistic representations of their referents, they are inaccurate in much the same way that false descriptions of an object are inaccurate. All represent their referents as they are not. So despite the fact that it is a bit of a misnomer, for ease of exposition, I label all such models falsehoods; if despite (or even because of) their inaccuracy they afford epistemic access to their objects, they are felicitous falsehoods.[6] I contend that we cannot understand the cognitive contributions of science, philosophy, or even our everyday accounts of things, if we fail to account for the epistemic functions such symbols perform.

Here are some familiar cases:

Curve smoothing: Ordinarily, each data point on a graph is supposed to represent an independently ascertained truth. (The temperature at t_1, the

temperature at t_2, . . .) By interpolating between and extrapolating beyond these truths, we expect to discern the pattern they instantiate. If the curve we draw connects the data points, this is reasonable. But the data rarely fall precisely on the curve adduced to account for them. The curve then reveals a pattern that the data do not instantiate. Strictly, it seems, veritism requires accepting the data only if we are convinced that they are true, and connecting these truths to adduce additional truths. In that case, the line should connect all the data points no matter how convoluted the resulting curve turns out to be. This is not done. To accommodate every point would be to abandon hope of finding order in most data sets, for jagged lines and complicated curves mask underlying patterns and regularities. Nevertheless, it seems cognitively disreputable simply to let hope triumph over experience. Surely we need a better reason to skirt the data and ignore the outliers than the fact that otherwise we will not get the kind of theory we want. Nobody, after all, promised that the phenomena would obligingly accommodate themselves to the kind of account we want.

We often have quite good reasons for thinking that the data ought not, or at least need not, be taken as entirely accurate. Sometimes we recognize that our measurements are relatively crude compared with the level of precision we are looking for. Then any curve that is within some δ of the evidence counts as accommodating the evidence. Sometimes we suspect that some sort of interference throws our measurements off. Then in plotting the curve, we compensate for the alleged interference. Sometimes we are confident that the measurements are accurate, but think the phenomena measured are complexes only some of whose aspects concern us. Then in curve smoothing we, as it were, factor out the irrelevant aspects, construing them as noise. Perhaps with a bit of hand-waving, veritism could accept such rationales for smoothing curves. Sometimes, however, we have no explanation for the data's divergence from the smooth curve. But we may correctly judge that what matters is the smooth curve the data indicate, not the jagged curve they actually instantiate. Whatever the explanation, we accept the curve, taking its proximity to the data points as our justification. We ignore the outliers, thinking that there is something-we-know-not-what unacceptable about them. We understand the phenomena as displaying the pattern the curve marks out. We thus dismiss the data's deviation from the smooth curve as negligible.

Ceteris paribus claims: Many lawlike claims in science obtain only ceteris paribus.[7] The familiar law of gravity,

$$F_g = \frac{Gm_1 m_2}{r^2}$$

is not universally true, for other forces are in play. The force between charged bodies, for example, is a resultant of electrical and gravitational forces. Nevertheless, we are not inclined to jettison the law of gravity. The complication that charge introduces just shows that the law obtains only ceteris paribus, and when bodies are charged, ceteris are not paribus. This is not news. 'Ceteris paribus' is Latin for 'other things being equal', but it is not obvious what makes for equality in a case like this. Strevens (2012) maintains that where a ceteris paribus claim is correct, its divergences from truth are not difference makers. This is plausible, but uncongenial to veritism. Sklar (1999, 702) glosses 'ceteris paribus' as 'other things being normal', where 'normal' seems to cash out as 'typical' or 'usual'. Then a 'ceteris paribus law' states what usually happens. In that case, to construe the law of gravity as a ceteris paribus law is to contend that although there are exceptions, bodies usually attract each other in direct proportion to the product of their masses and in inverse proportion to the square of the distance between them. Again, veritism could handle this.

But Sklar's construal does not always work. Some laws do not even usually hold. The law of gravity is one, since in addition to gravity, other forces, such as charge, always affect falling bodies. Snell's law, $n_1 \sin i = n_2 \sin r$, which expresses the relation between the angle of incidence and the angle of refraction of a light ray passing from one medium to another, is a more vivid case.[8] As standardly stated, the law is perfectly general, ranging over every case of refraction. It is not true of every case, though; it obtains only where both media are optically isotropic. The law then is a ceteris paribus law. But it is not even usually true, since most media are optically anisotropic (Cartwright, 1983, 46–47). One might wonder why physicists don't simply restrict the scope of the law: 'For any two optically isotropic media, $n_1 \sin i = n_2 \sin r$'. The reason is Gricean: expressly restricting the scope of the law implicates that it affords no insight into cases where the restriction does not obtain (Grice, 1989). Snell's law is more helpful. Even though the law is usually false, it is often not far from the truth. Most media are anisotropic,

but many of them—and many of the ones physicists are especially inter-
ested in—are nearly isotropic. The law supplies good approximations for
nearly isotropic cases. So although explanations and calculations that rely
on Snell's law do not yield truths, they are often not off by much. The cog-
nitive utility of the law involves two factors. One is that the law is approxi-
mately true for nearly isotropic media; the other is that those are the media
of interest. The acceptability of the law turns, evidently, not on its being
true, or being usually true, or even being usually approximately true. It is
approximately true in the cases that physicists care about.

The law's generality is valuable for another reason as well. Sometimes it
is useful to first represent a light ray as conforming to Snell's law, and later
introduce 'corrections' to accommodate anisotropic media. If we were inter-
ested only in the path of a particular light ray, such a circuitous approach
would be unattractive. But if we are interested in optical refraction in gen-
eral, it might make sense to start with a prototypical case, and then show
how anisotropy perturbs. By portraying anisotropic cases as perturbations,
we point up affinities that direct comparisons would not reveal. The issue,
then, is what sort of understanding we want. Showing how a variety of
cases diverge from the prototypical case affords valuable insights into the
phenomenon we are interested in. And what makes the case prototypical
is not that it usually obtains, but that it cleanly exemplifies the features we
deem important.

Stylized facts are close kin of ceteris paribus claims. They are "broad gener-
alizations, true in essence, though perhaps not in detail" (Bannock, Baxter,
& Davis, 1998, 396–397). They play a major role in economics, constitut-
ing explananda that economic models are required to explain. Models of
economic growth, for example, are supposed to explain the (stylized) fact
that the profit rate is constant.[9] The unvarnished fact, of course, is that
profit rates are not constant. All sorts of noneconomic factors—such as war,
pestilence, drought, and political chicanery—interfere. Manifestly, stylized
facts are not (what philosophers would call) facts, for the simple reason
that they do not obtain. It might seem, then, that economics takes itself to
be required to explain why known falsehoods are true. (Voodoo econom-
ics, indeed!) This cannot be correct. Rather, economics is committed to the
view that the claims it recognizes as stylized facts are in the right neighbor-
hood, and that their being in the right neighborhood is something eco-
nomic models need to account for. A critical question is what being in the

right neighborhood amounts to. The models may show them to be good approximations in all cases, or where deviations from the economically ideal are slight, or where economic factors dominate noneconomic ones. Or the models might afford some other account of their often being nearly right. The models may differ over what is actually true, or as to where, to what degree, and why the stylized facts are as good as they are. But to fail to acknowledge the stylized facts would be to lose valuable economic information (for example, the fact that if we control for the effects of noneconomic interference such as wars, epidemics, and the president for life absconding with the national treasury, the profit rate is constant). Stylized facts figure in other social sciences as well. I suspect that under a less alarming description, they occur in the natural sciences too. The standard characterization of the pendulum, for example, strikes me as a stylized fact of physics. The motion of the pendulum that physics is supposed to explain is a motion that no actual pendulum exhibits. What such cases point to is this: The fact that a strictly false description is in the right neighborhood is sometimes integral to our understanding of a domain.

Idealizations: An ideal pendulum—the one discussed in physics books, not the one in manuals for clockmakers—is a rigid body consisting of a small mass oscillating through a small angle at the end of a string. It provides a model for simple harmonic motion. The idealization neglects friction and gravity, and it treats the mass of the string as 0. Explanations that represent the pendulum as a simple harmonic oscillator would be epistemically unacceptable if abject fidelity to truth were required. Since every swinging bob is subject to friction and gravity, and every string has some mass, the assumptions built into the model are false. If veritism is correct, its falsity makes it epistemically unacceptable. But scientists find it unexceptionable.

A fortiori arguments from limiting cases: Some accounts focus on a single, carefully chosen case and argue that what holds in that case holds in general. If so, it does no harm to represent the phenomena as having the features that characterize the exemplary case. Astronomy sometimes represents planets as point masses. Manifestly, they are not. But because the distance between planets is vastly greater than their size, their spatial dimensions can safely be neglected. Given the size and distribution of planets in the solar system, what holds for properly characterized point masses also holds for the planets. Another familiar example comes from Rawls (1971). *A Theory of Justice* represents people in the original position as mutually disinterested.

Rawls is under no illusion that this representation is accurate. He recognizes that people are bound to one another by ties of affection of varying degrees of strength, length, and resiliency. But, he believes, if political agents have reason to cooperate even under conditions of mutual disinterest, they will have all the more reason to cooperate when ties of affection are present. If what holds for the one case holds for the others, then it does no harm to represent people as mutually disinterested. That people are mutually disinterested is far from the truth. Conceivably, every person cares about the fates of some other people. But if Rawls is right, the characterization's being far from the truth does not impede its function in his argument.

The foregoing examples show that in some cognitive contexts we accept claims that we do not consider true. But we do not indiscriminately accept falsehoods either. The question then is, what makes a claim acceptable? Evidently, to accept a claim is not to take it to be true, but to take it that the claim's divergence from truth, if any, is negligible. As Strevens (2008) puts it, the divergence from truth is not a difference maker. The divergence need not be small, but whatever its magnitude, it can be safely neglected. We accept a contention, I suggest, when we consider it true enough. The success of our cognitive endeavors indicates that we are often right to do so. If so, a contention is acceptable when its divergence from truth is negligible. In that case it is true enough.

As we have seen, contentions that are true enough diverge from truth in different ways. Some diverge in irrelevant respects; others, in relevant respects, but only a little. Yet others diverge (maybe a lot), but only rarely. Some diverge in cases their users do not care about; others, at a finer grain than their users care about. Some diverge radically, but in ways that are informative.

How might the veritist handle such divergences? One possibility is simply to stick to his guns. He might insist that only truths are epistemically acceptable. So if a scientific account ineliminably deploys models or idealizations that are not true, that account is unacceptable. If he is feeling charitable, the veritist might grant that such devices could play a heuristic role in the presentation or application of the theory. But he would insist that for the account to be epistemically acceptable, they must be excisable with no loss of cognitive content or epistemic justification. Such a position is likely to implausibly discredit much of our best science.[10]

How much is not clear. As we have seen, veritism can apparently accommodate some of the representations I have labeled felicitous falsehoods by construing them as approximations. It might, for example, accommodate Snell's law, by restricting the domain to nearly isotropic media, and endorsing not

(1) $n_1 \sin i = n_2 \sin r$

but

(2)   approximately $n_1 \sin i = n_2 \sin r$.

One difficulty with this approach is that scientists seem to use (1) rather than (2) in their reasoning. Perhaps the veritist can insist that the 'approximately' operator, although tacit, is always operative.

But not all felicitous falsehoods are even approximately true. Rawls's representation of deliberators in the original position as mutually indifferent is far from the truth. So is the Hardy—Weinberg model, which represents the redistribution of alleles in the absence of evolutionary pressures. It tracks alleles in an infinite population of organisms that mate randomly and are not subject to mutation or migration. That the population is infinite counteracts the effects of genetic drift; that mating is random ensures that neither physical proximity nor genes that give rise to attractive phenotypes have a preferential advantage; that neither mutation nor migration takes place ensures that no novelties are introduced into the gene pool. The model is no approximation. Populations are not nearly infinite (whatever that might mean). Mating is not nearly random. However indiscriminate actual mating behavior is, physical proximity is required. In the long run, mating only with nearby partners promotes genetic drift. Natural selection and genetic drift are ubiquitous. Migration is widespread. Mutation and random fluctuations are, in real life, unavoidable. Still, to understand the effects of evolution, it is useful to consider what would happen in its absence. By devising and deploying an epistemically felicitous falsehood, biologists find out. Veritists evidently have to simply deny that accounts that ineliminably deploy devices like the Hardy—Weinberg model or the mutually indifferent deliberators behind the veil of ignorance are epistemically acceptable.

Some approximations are accepted simply because they are the best we can currently do. They are temporary expedients, which we hope and perhaps expect to eventually replace with truths. Such approximations can be

construed, in Sellars's (1968) terms, as promissory notes that remain to be discharged. The closer we get to the truth, the more of the debt is paid. They are, and are known to be, unsatisfactory. But they can be successfully improved upon to get them closer to the truth.

Not all approximations have this character. Some are preferable to the truths they approximate. It is possible to derive a second-order partial differential equation that exactly describes fluid flow in a boundary layer. Such an equation describes, for example, how air flows directly over an airplane wing. The equation, being nonlinear, does not admit of an analytic solution. We can formulate the equation, but no one knows how to solve it. This is highly inconvenient. To incorporate the truth into the theory would bring a line of inquiry to a halt, saying in effect: 'Here's the equation; it is impossible to solve'. Fluid dynamicists prefer a first-order partial differential equation that approximates the truth, but admits of an analytical solution (Morrison, 1999, 56–60). The solvable equation advances understanding by providing a close enough approximation yielding numerical values that can serve as evidence for or constraints on future inquiry. The approximation then is more fruitful than the truth. There is no hope that future inquiry will remedy the situation if, as may be the case, the second-order equation cannot be solved numerically, while the first-order equation can.[11] That reality forces such a choice upon us may be disappointing, but under the circumstances it does not seem intellectually disreputable to accept and prefer the tractable first-order equation. A veritist might argue that acceptance of the first-order approximation is only practical. It is preferable merely because it is more useful. This may be so, but the practice is a cognitive one. Its goal is to understand turbulent flow. In cases like this, the practical and the theoretical inextricably intertwine. The practical value of the approximation is that it advances understanding of a domain. A felicitous falsehood thus is not always accepted only in default of the truth. Nor is its acceptance always second best. It may make cognitive contributions that the unvarnished truth cannot match. These contributions seem out of veritism's reach.

The veritist's final alternative is to say that the use of epistemic success terms like 'knowledge' and 'understanding' in cases that involve felicitous falsehoods is a courtesy usage. Just as we might say, perhaps a little condescendingly, that a young child has some understanding of photosynthesis if she thinks that sunlight is the flower's food, the veritist might say, a little

condescendingly, that Kepler had an understanding of planetary motion, even though his contention that the Earth's orbit is elliptical is strictly false. Similarly, the veritist might say that a currently promising theory that has its share of successes but some outstanding anomalies constitutes an understanding of its subject matter. In so saying, he is in effect taking out a lien on the future. He credits the current theory because he expects future science to correct its errors and fill its gaps. Arguably, he could say the same thing about a theory that makes currently ineliminable use of felicitous falsehoods. He would then be betting that science will eventually discover the resources to eliminate them without replacing them with other felicitous falsehoods. Fluid dynamics, for example, might find a way to either solve the second-order differential equation that currently cannot be solved, or discover a way to exactly characterize and calculate turbulent flow with a solvable equation that states the exact truth. The difficulty with this stance is that although scientists expect that current models and idealizations will be refined with the advancement of inquiry in their field, they do not expect that science will outgrow the use of such devices. Rather than being seen as temporary crutches, modeling and idealizing are considered valuable ways to embody scientific understanding. So even if the Sellarsian lien on the future is a reasonable perspective to adopt about flawed but promising factual claims, it does not seem a plausible way to understand the epistemic utility of models and idealizations. If contemporary science is our guide, they are here to stay.

Current science is riddled with epistemically felicitous falsehoods. Current scientists are disinclined to repudiate them. This leads me to doubt that they will disappear with the advancement of inquiry. The veritist disagrees. He thinks such devices will ultimately be eliminated or consigned to merely heuristic uses. Crystal balls are notoriously cloudy, so prognosticating about the future of science is a risky business. Instead, in what follows I will develop an epistemological position that shows how felicitous falsehoods function in understanding, what they contribute, and why they are epistemically valuable. What follows will not show that such devices are ineliminable. But it will suggest that their elimination is neither necessary nor obviously desirable.

This chapter has raised an important difficulty for epistemology and pointed toward a resolution. That resolution requires holism, nonfactivism, and a reconception of the basis of epistemic normativity. In the next several

chapters I will show how to satisfy all of these requirements. Only then will we be in a position to appreciate the way they equip us to account for the epistemic functions of felicitous falsehoods.

Although I have focused on examples drawn from natural science, I do not contend that all understanding reduces to scientific understanding, or that the use of epistemically felicitous falsehoods is peculiar to scientific understanding. They are, for example, also found in philosophy, history, and the social sciences. As will emerge, I believe that science is epistemically much closer to the arts than is standardly believed. My point in initially focusing on natural science is strategic. Because science is manifestly epistemically estimable, and is self-conscious about its methods and standards, it is often a useful place to focus. The requirement that epistemology accommodate science is nonnegotiable. If science incorporates felicitous falsehoods into its accounts, epistemology needs to explain what they contribute. Nevertheless, I will begin by considering understanding in general, temporarily leaving issues pertaining to felicitous falsehoods in abeyance.

3   From Knowledge to Understanding

As the examples sketched in the previous chapter show, the epistemic excellence of mature science is not easily construed as knowledge. To accommodate not just science, but disciplinary epistemic success in general, we should focus on understanding. I begin to explicate understanding and sketch how it differs from knowledge. Following Kvanvig (2003), I distinguish between propositional understanding, which is captured in an individual proposition, and objectual understanding, which is captured in an account. The cognitive competence involved in understanding is generally characterized as grasping. Propositional understanding involves grasping a fact; objectual understanding consists of grasping a range of phenomena. This seems right. But it is not clear what grasping is. I suggest that to grasp a proposition or an account is at least in part to know how to wield it to further one's epistemic ends. To make this out requires a lengthy discussion of know-how. That discussion is far from exhaustive. It is structured to highlight the aspects of know-how that figure in objectual understanding. Disciplinary understanding, I argue, is best construed as objectual. In the first instance we understand a range of phenomena via an overall account; only derivatively, drawing on the resources of that account, do we understand that or why something is the case. I argue that scientific (objectual) understanding is nonfactive. Although it is not indifferent to the facts it concerns, an account that accommodates those facts need not consist exclusively or predominantly of truths.

The Limits of Knowledge

Bill thinks that apatosauruses and brontosauruses are distinct. He has good reason to think so. But he cannot know that they are distinct unless in fact

they are. We do not consider false beliefs to be knowledge, no matter how well founded they may be. Once Bill discovers that, contrary to what he believed, the apatosaurus is the same sort of dinosaur as the brontosaurus, he rescinds his claim ever to have known that they were distinct. He may say, 'I thought I knew that they were distinct, but I was wrong'. He may even say, 'I had it on excellent authority that they were distinct, but I and my authorities were wrong'. He will not say, 'I knew they were distinct, but I was wrong'. The term 'knowledge' is factive. An epistemic agent does not know that p unless 'p' is true.

Inasmuch as each bit of knowledge is keyed to a particular fact, knowledge is granular. Even so, it is (close to) impossible to know one thing at a time. Insofar as knowledge requires justification, in order to know that p, an agent's belief that p must be backed by considerations that bear on whether p is the case. It seems that these considerations must themselves be known; otherwise even if p is (externally) justified, the agent would not be justified in believing that p. She would have no reason to trust her conviction. But if she must know the justificatory chain leading to p, an agent will have to know a lot in order to know (much of) anything.[1] There is then a tension between the apparent granularity of knowledge and the nongranularity of justification. One way to alleviate the tension might be to consign the justifiers to 'background knowledge' and ask whether, together with the empirical evidence, the 'background knowledge' affords sufficient grounds to underwrite a particular claim. This might work for simple inductive cases, or the sort of amassing discrete bits of evidence that one sees in detective stories. But problems arise when an account figures in the justification for a claim. If the background ethological account together with the empirical evidence justifies the conclusion that this bit of grooming behavior manifests reciprocal altruism and if that conclusion is true, it might seem that we are in a position to know it. Unfortunately, things are not so straightforward. Such a connection reveals whether an account supports a claim, but the assumption that the 'background knowledge' is genuine knowledge cannot be sustained. The justification for many of the claims making up a scientific account is reciprocal; they support one another. How they do so will be discussed in the next chapter. Here the point is simply that systematic accounts are not granular in the way that mainstream epistemology takes knowledge to be.

Another, potentially more promising, strategy is to take holism at its word and relax granularity. The atomic sentences that comprise an account cannot be separately justified. Even if the focus of attention is a particular claim, evidence always bears on an account as a whole. Evidence for the claim that a given process is adiabatic is evidence for an entire account of heat transfer, which is tested along with the claim. Evidence that the grooming behavior displays reciprocal altruism is evidence for an entire account of primate behavior, which is tested along with the claim. In principle, this is epistemologically unproblematic. The contention that knowledge is propositional says nothing about the length of the propositions that constitute knowledge. We might accommodate holism by treating an account as a conjunction of its component atomic propositions and saying that the evidence bears on the truth or falsity of that long conjunction. If the conjunction is true, is believed, is justified or reliably produced, and is epistemically lucky, 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 epistemic standing of science, for its requirements are rarely if ever met. In particular, the truth requirement is seldom satisfied. As we have seen, many epistemically estimable accounts contain models and idealizations that do not even purport to be true. For now, however, this complication will be set aside. The current difficulty is that even the best scientific accounts confront anomalies. They entail consequences that conflict with the evidence. A conjunction is false if any of its conjuncts is; so if a scientific account and the evidence that bears on it are a conjunction, an anomaly—being a falsifying instance—tells decisively against the account that it bears on. Since a theory that confronts an anomaly is false, it does not amount to knowledge.

Perhaps we can evade this predicament. The characterization of an account as a conjunction might seem to offer hope of isolating anomalies and screening off their effects.[2] 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. In any of these circumstances, (1) 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 reliably generated. Since neither (a) nor the evidence for (a) lends any support to (b), (b)'s acceptability is not undermined by the repudiation of (a). On standard accounts of knowledge, assuming that the relation between the truth of (b) and the evidence for (b) is not vitiated by epistemic misfortune, we are in a position to know that (b). If the components of a scientific account were related to one another as loosely as (a) and (b) are related in (1), we could simply excise the anomalous sentences and be left with a justified, reliable truth—something that could be known.

The problem is that the components of an account lack the requisite independence. An account is a tightly interwoven tapestry of mutually supportive commitments. Simply excising anomalous sentences would leave a moth-eaten fabric that would not hang together. Toward the end of the nineteenth century, physicists devised a variety of increasingly drastic expedients to accommodate the perihelion of Mercury. But even at their most desperate, they did not suggest simply inserting an exception into the theory. Although 'All planets except Mercury have regular orbits' is evidently 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 can be a good investigative tactic, but simply dismissing them as exceptions that need not be accommodated 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 perihelion of Mercury, be symptomatic of a subtle but significant misunderstanding of the phenomena. Science would lose potentially valuable information if it dismissed anomalies as exceptions that it could simply disregard. There is then no hope of excising anomalous sentences without undermining the epistemic support for the rest of the account. The account rather than the individual sentence or proposition is the unit we need to focus on.

These points are familiar and uncontroversial, but their epistemological consequences are worth noting. An account can be construed as a conjunction of the sentences that appear in it. But science does not consist of knowledge expressed by such conjunctions. For the conjunction of the

sentences that constitute an epistemically estimable scientific account is almost surely false. The unavailability of sentence-by-sentence verification or justification discredits the idea that science separately delivers knowledge of each component atomic sentence. The hopelessness of selectively deleting the falsehoods in, and the false implications of, an account undermines the plausibility of claiming that scientific knowledge is what remains when an account's falsehoods have been expunged. Knowledge requires truth. And there seems to be no feasible way to get the scientific accounts we admire to come out true. So knowledge is not the cognitive condition that good science standardly engenders.

Refocusing Epistemology's Lens

To accommodate science, I suggest, epistemology should shift its focus from knowledge to understanding. In speaking of understanding, we should distinguish between the objects of understanding and the vehicles of understanding (see Greco, 2014). The vehicles of understanding are whatever embody the understanding—perhaps sentences, propositions, accounts, nonpropositional representations, or dispositions. An object of understanding is whatever the understanding is about. Objects of understanding may be concrete or abstract. An epistemic agent can understand linear algebra as well as particle physics. She can understand why $\sqrt{2}$ is irrational as well as why the noble gases are inert. Nor, I maintain, need the objects of understanding be articulable. A commuter can understand the New York subway system and an avid fan can understand the Celtics' defensive strategy without being able to express that understanding in words or endorse a description of the phenomena expressed in the language that she speaks.

A critical question is whether 'understanding', like 'knowledge', is factive. If it is, then the shift from knowledge to understanding may avail us little. Pretty plainly, understanding somehow answers to facts. The question is how it does so. If 'understanding' is factive, all or most of the propositional commitments that comprise a genuine understanding are true. And the epistemic status of nonpropositional elements is instrumental and dependent on their role in underwriting the epistemic standing of those truths. Many epistemologists consider understanding factive (see Kvanvig, 2003; Grimm, 2006). But, I will argue, a factive conception is too restrictive. It does not reflect our practices in ascribing understanding and it forces us

to deny that much contemporary science embodies an understanding of the phenomena it bears on. This is too high a price to pay. A more generous, flexible conception of understanding accommodates the deliverances of science, reflects our practices, and shows a sufficient, but not slavish, sensitivity to the facts it bears on.

We cannot settle the question by inspecting ordinary usage, as we did with 'knowledge'; for the term 'understanding' is used in a variety of ways, none of which seems to dominate the rest. Some are irrelevant to epistemology; others pull in different directions. 'I understand' can hedge an assertion or attenuate its force. 'I understand that you are angry with me' may be a mild overture that gives you space to politely demur. 'Not angry, but disappointed', you might reply. This is a moderating use. Or I might say, 'I understand that the committee has the authority to decide' when I am not sure that it has the authority, but have reason to think so. Here 'I understand' indicates a backing away from a full-fledged claim to epistemic entitlement. This is hedging. These are not the sorts of usages that concern us here. Our concern is with cases in which understanding is a sort of epistemic success. I shall use the term 'understanding' as a success term for having an epistemically suitable grasp of or take on a topic. Just what constitutes a suitable grasp or take remains to be seen. Here the point is that in such cases the understander has a claim to epistemic entitlement.[3] The questions, then, are what is the bearer of that entitlement and what is the claim to it? I contend that a nonfactive explication of 'understanding' yields a concept that better suits epistemology's purposes than a factive one does.

I recognize the existence of factive uses of the term 'understanding', just as I recognize the existence of hedging and moderating uses. Nor do I deny that epistemology could incorporate a factive conception of understanding. My contention is that a factive conception cannot do justice to the cognitive contributions of science and that a more flexible conception can. But understanding is not restricted to science, so I begin with an historical example:

In 490 BCE the Athenian army confronted the invading Persians on the plain of Marathon. The Persians vastly outnumbered the Greeks. Nevertheless, the Athenian commander, Miltiades, decided to attack. Contrary to standard military practice, he arrayed his forces in a line with a weak center and strong flanks. The Persians

readily repulsed the center of the line. When the Persians pursued the retreating Greeks, the flanks moved in and surrounded the Persian army, attacking them from all sides. They thereby won the Battle of Marathon. (paraphrased from Herodotus, 1985)

The first order of business is to identify the primary object of understanding. Two obvious candidates present themselves: individual facts and more comprehensive domains. The former is propositional understanding; the latter objectual understanding (Kvanvig, 2003). One might doubt that the two are distinct. Perhaps the body of information that constitutes an objectual understanding of a topic is a single long conjunctive proposition. Then objectual understanding reduces to propositional understanding. Alternatively, perhaps the understanding of an individual matter of fact qualifies as understanding only because it is backed by a complex justificatory network. Then the agent's grasp of the network is implicated in her understanding of *p*.

The basic unit of propositional understanding is the individual sentence or proposition. One possibility is that its form is:

x understands that *p* is so.

For example,

Jack understands that Athens defeated Persia in the Battle of Marathon.

If this is correct, the difference between knowledge and understanding seems slight. To understand that Athens won the battle, Jack would presumably require a nonfortuitously justified true belief to that effect. Since he learned it from Herodotus, and Herodotus is a good source of such information, most epistemologists would probably grant that he knows it. If this is all it takes to understand that Athens won the battle, then understanding is factive, because knowledge is.

We might be disinclined to think that Jack understands that Athens won if all he can do to support his belief is point to his source. If he did not read Herodotus's full account of the battle or did not comprehend or register the information it provides, he has no idea why Athens won. He merely knows that Athens won. It may be preferable, then, to take understanding to be a

matter of knowing *why*, rather than merely knowing *that*. In that case, the fundamental schema is:

x understands why *p* is so.

Many hold that to understand why is to be in a position to explain (see Strevens, 2008; Grimm, 2006). Jill understands why Athens was victorious because she can explain the victory. On this sort of account, to understand a fact is different from merely knowing that fact. But it is still evidently a form of knowledge—knowledge why *p* is so. According to Lipton, "Understanding is not some sort of super-knowledge, but simply more knowledge: knowledge of causes" (2004, 30). This is too narrow. Knowledge of causes might account for understanding that the moon causes the tides, or that Athens won the battle of Marathon, but it cannot account for understanding that $\sqrt{2}$ is irrational. Nonetheless, it may be that even in mathematics understanding is knowing why. This suggests that advocates of a knowing why account would do better to explicate propositional understanding in terms of knowledge of dependencies, and let knowledge of causes be a special case (see Greco, 2014; Strevens, 2010; Grimm, 2014). This is more plausible. Understanding is not just knowledge, but a particular sort of knowledge. Since knowledge of dependencies, like all knowledge, is factive, understanding is too.

Perhaps the demand for *knowledge* of dependencies is too strong. Understanding might be factive even if it is not knowledge. Factivity requires only that the proposition or account that embodies an understanding be true, not that it be known. In view of the plethora of threats to knowledge, factivity may be the best we can hope for. The worry comes from Gettier cases. Where there is a misalignment between truth-makers and justifiers, an agent can have a justified true belief that does not amount to knowledge. Following Pritchard (2010), let us say that a belief is environmentally unlucky if it is formed or sustained in an epistemically hostile environment. Environmental misfortune defeats knowledge since a belief formed in such an environment could too easily have been false. The question is whether such misfortune undermines understanding.

Pritchard says 'no'. Elaborating on an example developed by Brogaard (2005), he sketches the following scenario: Suppose Fred arrives home to find his house in flames. He asks someone who appears to be a fire marshal

why the house is on fire. She reports that the fire was caused by faulty wiring. Given her expertise, we would ordinarily conclude that Fred knows on the basis of her testimony that faulty wiring caused the fire. But suppose that there are several people milling around, all dressed as fire marshals. En route to a costume party, they stopped to look at the blaze. If it is just by luck that Fred asked the actual fire marshal rather than one of the costumed partygoers, he does not know. He could too easily have chosen an unreliable informant. Nevertheless, Pritchard maintains, Fred understands why the house is on fire, since he has accurate information that coheres with his relevant background beliefs. "The agent . . . has all the true beliefs required for understanding why his house burned down, and also acquired this understanding in the right fashion. It is thus hard to see why the mere presence of epistemic luck should deprive the agent of understanding" (2010, 78–79). I do not find this case compelling. But if it is, or if a more compelling case could be devised, then understanding is not knowledge, for one can have propositional understanding where the anti-luck condition on knowledge is violated.

Some (Kvanvig, 2003; Grimm, 2014; Hills, 2015) argue that understanding is keyed to explanation. To propositionally understand a fact is to grasp an explanation of that fact. In mathematics the explanation is not causal; in empirical disciplines it is. If an agent can provide a suitable explanation of a fact, we are inclined to credit her with understanding that fact. The challenge comes where the agent cannot provide or comprehend an explanation. Does this completely disqualify her from understanding a given fact?

Strevens (2008, 2010) takes the basic unit of understanding to be the proposition. On his picture, Sara understands why *p*, only if she can explain *p* or grasp an explanation of *p*, where the explanation is a derivation from more fundamental truths that ground *p*. If *p* is a matter of empirical fact, the explanation is causal. It includes all and only causal factors that make a difference to whether *p* obtains. If *p* is a fundamental truth (such as a fundamental law of nature), then the explanation why that figures in understanding it is an explanation of what derivative or higher-order facts *p* is a cause of. According to Strevens, then, all understanding of the material world is causal. To understand the Athenian victory is to be able to explain or comprehend an explanation of how it arose from more basic truths. Other sorts of understanding—mathematical, ethical, aesthetic—are still grounded in

explanations that reflect real, metaphysical dependency relations, but the explanations are not causal (Strevens, 2010).

There are a number of worries about this picture. Not all understanding involves explanation or grounding of any sort. A savvy commuter who understands the New York City subway system may have no explanation of the regularities she observes. Her understanding is exhibited in her facility in negotiating the subway system, in accommodating her travel to its vicissitudes. She knows, for example, that the number 2 train is generally faster than the number 3, but not when it rains. She knows where to stand on the 14th Street platform to maximize her prospects of getting a seat. She readily accommodates herself to delays and disruptions, having a feel for when it makes sense to wait things out and when it is better to switch to a different line, and so forth. Such understanding is, or is akin to, know-how. I will have more to say about that below.

Nor is it obvious that all scientific understanding is causal. Ecology and population biology, for example, study patterns of population growth and diminution. Animal populations are known to regulate their own size. But the mechanisms by which populations self-regulate are not known. Still, it seems, even if they cannot fully explain how it is done, ecological models of population size embody an understanding of the dynamics of population regulation (Sarkar, 2014).

Mathematics poses an even greater difficulty. According to Strevens, "If mathematical explanation is possible at all, there must be a dependence relation whereby some mathematical facts depend in some sense on other mathematical facts. It is in virtue of this dependence relation that some mathematical facts are fundamental and others should be understood as holding in virtue of the fundamental facts"; he goes on to say, "A proof in mathematics may or may not be explanatory. In order to be explanatory, the entailment structure of the proof must mirror the relevant mathematical dependencies" (2010, 14–15). Whatever the merits of this as a view of mathematical *explanation*, mathematical *understanding* seems not so tightly tied to metaphysical grounds. Any axiomatized branch of mathematics can be reaxiomatized. Under different axiomatizations, different patterns emerge. All contribute to mathematical understanding in that each reveals networks of mathematical relations that others obscure. It seems tendentious to insist that someone who grasped a valid proof of a theorem had no understanding of the subject because the theorem in question

derived from an axiomatization that did not mirror the right dependence relations.

Questions arise even with respect to the understanding of ordinary matters of fact. Must the elements of the grounding explanation be independently understood? Strevens (2008) recognizes that the ultimate laws of nature cannot be explained in terms of anything more basic. In understanding them, we understand how they underwrite less basic claims. This is not my worry. Rather, I am concerned about the epistemic relation between explanandum and explanans. If the explanandum must be epistemically prior to the explanans, it is going to be difficult to understand the Athenian victory, since the elements of the explanation—the details about the deployment of forces and the effectiveness of the strategy—seem hardly more secure than what they are adduced to explain.[4] Since ultimately, most of the information that figures in the explanation comes from Herodotus, if the explanation must be known, although Jill can explain the victory, she seems to be not much better off epistemically than Jack.

If the elements of the explanans need not be known, there is something potentially misleading about claiming that to understand why p is to be able to explain it or comprehend an explanation of it. Such a contention suggests that the explanans is known independently of, and is epistemically prior to, the explanandum. What we actually see is something more holistic, with the understanding embodying not only the fact to be explained but also the elements of the explanation and whatever supports them. Herodotus's whole account of the battle then stands or falls together. In that case, however, it appears that the basic schema for understanding is

x understands φ

where φ is a topic, discipline, or subject matter. This is objectual. Jill, in the first instance, understands the Battle of Marathon, and thereby derivatively understands and can explain or comprehend an explanation of particular matters of fact such as the course of the battle, the strategies and tactics of the two armies, and the history of relations between Athens and Persia. The epistemological standing of 'Athens defeated Persia in the battle of Marathon' then implicates a more comprehensive understanding of the history of Greece or of Persia or of warfare. I do not deny that the sort of understanding why that figures in explanation is real. I simply doubt that it is

either as comprehensive or as fundamental as its adherents think. I believe that the understanding of a particular matter of fact derives from the understanding of a suitably unified, integrated, tenable body of information that bears on that fact. Objectual understanding, I take it, is the core conception of understanding. It affords a basis for distinguishing between understanding a topic and knowing particular truths about it.

Objectual Understanding

At a first approximation, an understanding is an epistemic commitment to a comprehensive, systematically linked body of information that is grounded in fact, is duly responsive to reasons or evidence, and enables nontrivial inference, argument, and perhaps action regarding the topic the information pertains to. This rough approximation is far from a full explication. But we can augment it with accepted verdicts about what we take to be clear cases. Astronomy affords an understanding of the motions of celestial bodies and their effects; astrology does not. Chemistry affords an understanding of the constitution of matter; alchemy does not. Biology affords an understanding of the origin of species; 'intelligent design' does not. An adequate epistemology should, at least for the most part, respect such verdicts.

   Although one mark of understanding on this conception is an ability to explain or comprehend an explanation, no such ability is required. Gareth is incapable of explaining the apparent retrograde motion of the outer planets. Nor can he comprehend a verbal description of the phenomenon. But he can demonstrate what happens using an orrery (Lipton, 2009). If his demonstration is effective, and if he can display enough cognitive dexterity in accommodating different scenarios to show that his demonstration is not a fluke, it is hard to deny that he understands the phenomenon. Pat understands the New York subway system in that she can use it to get around the city efficiently, being aware of such matters as which lines are fastest at different times of day, how foul weather affects different subway lines, which transfer points are likely to be uncrowded, and so forth. She may have no idea how to explain such matters, no appreciation of a transportation engineer's explanation of these matters, and no reason to think they need explaining. Her understanding is exhibited in the adroitness of her subway-riding behavior and in the advice she gives to other riders, not in any capacity to explain. A physician understands the course of a disease,

being aware of the sequence of signs and symptoms, the duration of the contagious phase, the potential complications, and the ranges of responsiveness to treatment, even if neither she nor anyone else can explain why the disease presents in the way it does.

Whether this sort of understanding is factive is the question I need to address. Understanding on my view is a (perhaps tacit) endorsement of a fairly comprehensive, interconnected constellation of cognitive commitments. The understanding encapsulated in individual propositions derives from an understanding of larger bodies of information that include those propositions. In understanding the Athenian victory in the Battle of Marathon, Jill grasps how the proposition stating the fact that Athens won fits into, contributes to, is justified by reference to, and figures in the justification of a more comprehensive understanding that embeds it.

Obviously, not just any comprehensive, mutually supportive set of cognitive commitments will do. A coherent body of manifestly unfounded contentions does not constitute an understanding of the phenomena they purportedly bear on. Even if it is coherent, astrology affords no understanding of the cosmic order. The issue that divides factivists and nonfactivists is not whether understanding must answer to the facts, but how it must do so. Following Plato (1997), let us call the required connection between a comprehensive, coherent account and the facts it bears on an understanding's tether. Even if astrology offers a comprehensive, internally coherent account of the cosmos, it yields no understanding because it lacks a suitable tether.

To be sure, we sometimes say things like 'Joe understands astrology', or 'Paul understands mythology', or 'Michael understands rationalism', meaning only that the epistemic agent knows his way around the field. He knows how its contentions hang together, and is adept at reasoning within the framework that they constitute. In such cases, the epistemic agent's commitments are tethered to the doctrine or account, not to the phenomena that doctrine or account purports to be about. One can understand an account regardless of its fidelity or infidelity to the phenomena it concerns. Such an understanding takes an account—astrology, rationalism, or whatever—to be its object rather than the phenomena the account purportedly pertains to. There is nothing epistemically special about this sort of understanding. It is of a piece with understanding a natural phenomenon or a historical episode or a transportation system. Still, it is worth noting

that it is possible for an account to be the object of understanding even if, owing to its untenability, it affords no understanding of its purported subject matter.

Understanding the Athenian victory involves more than knowing the various truths that belong to a suitably tethered comprehensive, coherent account of the matter. The understander must also grasp how the various truths relate to each other and to other elements of the account. Ceteris paribus, she should also be both willing and able (and perhaps be aware that she is willing and able) to use that information—to profess it, to reason with it, to apply it, perhaps to use it as a source of working hypotheses about related matters when her ends are cognitive.[5] Someone who knows geometry, for example, knows all the axioms, all the major theorems and their derivations. You can acquire this knowledge by rote. But someone who understands geometry can reason geometrically about new problems, apply geometrical insights in different areas, assess the limits of geometrical reasoning for the task at hand, and so forth. Understanding a historical event like the Athenian victory is not exactly like understanding geometry, since the applications and extensions are more tentative, the range to which insights can reasonably be applied is more restricted, the evidence for a successful application is empirical (and may be hard to come by), and so on. But in both cases understanding involves an adeptness in using the information one has, not merely an appreciation that things are so. Evidently, in addition to apprehending connections, an understander needs the ability to use the information at his disposal.

Know-How

Understanding is widely held to be a matter of grasping. The difficulty is to spell out in nonmetaphorical terms what grasping is. I suggest that an important element of grasping is knowing how to exploit the information or insight one's understanding provides. Someone who understands a proposition knows how to wield it to further her cognitive (and perhaps practical) ends. Someone who understands a topic knows how to use the epistemic resources her take on that topic affords. To make this out I need to take a detour to consider the nature of know-how.

Ryle (1949/2009) contends that knowing how is entirely a matter of habits and dispositions. To know how to ride a bicycle, he believes, is just to be

disposed to behave (or to be in the habit of behaving) in certain ways while astride a bicycle. Obviously knowing how to ride a bicycle does not consist in a disposition to engage in a single specific behavior. Knowing how to ride a bicycle is not like a propensity to dissolve in water or to shatter when struck. Ryle construes knowing how to do something as a multitrack disposition, "consisting of more or less dissimilar exercises" (1949/2009, 56). Knowing how, he believes, is a propensity to perform any of a variety of systematically linked but distinct acts in a range of diverse but not wholly unanticipated circumstances. Since circumstances vary, sensitivity to circumstances gives rise to a cluster of available responses.

A standard criticism is that Ryle's theory is excessively behaviorist. He explicates virtually every mental predicate as a disposition to overtly behave somehow or other (Carr, 1979). But we need not accept Ryle's entire theory of mind to accept (or adapt) his explication of know-how. Nor need we hold that all the habits and dispositions involved in knowing how to do something are habits and dispositions to overt behavior. A dispositional account can recognize that knowing how often involves dispositions not only to behave, but also to think, notice, infer, and/or feel; to ignore, marginalize, emphasize, and/or find salient. To construe knowing how dispositionally is to characterize it in terms of propensities, or readinesses, or reluctances to do various things in various circumstances. It may involve propensities to think certain things, to represent things mentally in certain ways, to feel certain emotions, and to refrain from thinking, representing, or feeling others. No commitment to behaviorism is required for a dispositional account of know-how. Indeed, the knowing how that figures in epistemic understanding is largely if not completely mental. It is a disposition to reason well about a given topic in a range of relevant circumstances.

Whether or not Ryle is correct to *identify* knowing how to do something with a multitrack disposition, know-how clearly *involves* multitrack habits and dispositions. A person does not know how to ride a bicycle unless she is disposed to peddle, steer, and maintain her balance on a bike. And in different circumstances (when the surface is slippery, in traffic, on steep hills, over rough terrain, etc.) different fine-grained behaviors are required to peddle, steer, keep one's balance, and so forth. A person does not know how to recognize ad hominem arguments unless she is disposed to spot the fallacy in a variety of rhetorical guises.

Knowing how to φ does not require reliably φ-ing. In basketball, even the best outside shooters make fewer than half of the three-point shots they attempt.[6] That they even come close to 50 percent is a tribute to their know-how. A recently paralyzed pianist still knows how to play the *Moonlight Sonata* even though she can no longer display her know-how. An emergency medical technician knows how to perform CPR even though she is never called upon to do so. These caveats are all compatible with a Rylean position.

But knowing how is not a mere multitrack disposition. Knowing how is an achievement. It involves a capacity to do something well, or rightly, or correctly. An adequate explication should do justice to this normative character. Some habits are bad; some are neutral; some are good. Some dispositions are benign; some are beneficial; some lead us astray. If someone who takes himself to be serious about tennis is complacent about habitually stepping on the base line while serving, he evidently does not know how to serve. If he is unconcerned about or even unaware of his habit of wiping his brow before serving, it makes no difference. That habit is irrelevant to whether he knows how to serve. If he is indifferent to his propensity to hit the ball into the net, he does not know how to serve. On the other hand, if he reliably serves into the diagonally opposite service box, rarely double faults, occasionally aces, and recognizes the value of doing so, he knows how to serve. What is missing from Ryle's account is the normative element. But what sort of normativity is at issue here?

Abel (2012) characterizes knowing how as rule-following: actions take place within practices, and the rules of the practices supply the norms that govern the actions those practices embed. Whereas bad habits and dispositions are propensities to flout relevant rules, and neutral habits and dispositions are uninfluenced by the rules, knowing how consists in a propensity to follow the rules. If Jim knows how to play chess, his chess-playing behavior typically follows the rules of chess. If Jane knows how to play tennis, her tennis-playing behavior typically follows the rules of tennis. This seems almost trivial. But it raises a number of questions.

One is whether all knowing how is a matter of following the rules of a practice. Games like tennis and chess are plainly rule-governed. So are practices like standing in line and paying one's taxes. Participants have a pretty good idea what the rules are and what it is to follow them. But not all knowing how is keyed to the norms of a practice. Consider knowing how

to swim (by which I mean no more than to propel oneself through deep water so as not to drown). Dogs can swim. They do not follow the rules of a practice. Why should we think that our basic ability to propel ourselves through water is different from a dog's? If a child learned to swim by mimicking the motions of her dog, would we say that she did not know how to swim? Or consider Abel's example of tying a necktie. Suppose someone regularly engages in a variety of deviant contortions that reliably result in a perfect Windsor knot. Should we say that he does not know how to tie the necktie simply because he fails to follow the rules for necktie tying that are canonical in his society? Or should we say, 'If it works, it works'? Some know-how appears straightforwardly consequentialist. The standard for displaying it seems to be no more than reliably producing the desired result. In such cases the end justifies the means. Beyond the rules that govern all actions within a given practice (such as 'Don't kill anyone while doing it'), the rationale for the behavior that constitutes a straightforwardly consequentialist action seems to be justified by the ends it seeks to promote, not by rules of the practices (if any) it belongs to. It is not obvious, then, that knowing how to perform straightforwardly consequentialist actions is properly explicated as rule-following.

Still, much know-how seems to be a matter of following the rules of a practice. And it is plausible that the know-how required for epistemic acceptance is practice based. So let us consider such cases.

What is it to follow a rule? A plausible answer is that to follow a rule is to intentionally regulate one's behavior by reference to the rule. The novice chess player mutters to herself, 'The bishop moves diagonally', then moves her bishop along a diagonal. As a general account, this will not do. First, it apparently sets off an infinite regress. Since language is itself a rule-governed practice, the novice would have to invoke rules for interpreting her muttering, and rules for interpreting those rules, and so forth. We cannot do this indefinitely. Nor can we plausibly maintain that there is a self-interpreting level of discourse where the regress ends. Second, it is no accident that my example concerns what a *novice* chess player might do. An experienced player—someone who knows how to play chess—does no such thing. In Wittgenstein's (1953, §219) terms, she acts blindly. Although she may deliberate about strategy and tactics, she has so internalized the rules governing ways the various pieces can move that they have become second nature to her. For her, *to be a chess bishop* is to be able to move only along a

diagonal. Once she recognizes a chess piece as a bishop, how it can move is settled. No inference need be drawn, no rules consulted. The capacity to act blindly while being subject to norms requires explanation.

If we act blindly in exercising know-how, what role do the norms play? We evidently do not consult them or intentionally regulate our behavior by reference to them. Once we know how to do something, doing it in appropriate circumstances is second nature to us. It might seem then that knowing how is simply a matter of automatically, unthinkingly behaving in accord with the norms of a practice.

This will not do. Knowing how is sensitive to why we automatically, unthinkingly behave as we do. Consider the following case: Except in New York City, drivers in the United States are permitted to turn right at a red light unless a sign saying 'No Turn on Red' is posted at the intersection. In New York City, right turns on red are never permitted. Drivers from out of town tend to be unaware that New York is an exception to the general rule. Suppose Meg, a denizen of a small town in Kansas, is driving in New York. Unsurprisingly, she finds the experience harrowing. She stops at every red light, not because she is aware of or sensitive to the law, but because she considers New York drivers and pedestrians reckless and wildly unpredictable. She deems it safer to proceed only when the light is green. She acts in accord with the law, but not on account of it. The regularity in her behavior is not an instance of following the traffic law. Although it may be an instance of knowing how to drive safely, it is not an instance of knowing how to obey New York traffic laws. Acting in accord with the law is not the same as acting on account of the law (see Kant, 1981). And only acting on account of the law qualifies as knowing how.

Virtue

Again we face the tension: On the one hand, knowing how to follow the rules of a practice seems to require us to be cognizant of those rules; on the other, we act blindly. How is it possible to do both? Here it pays to turn to Aristotle. The virtuous person, Aristotle says, does the right thing, in the right way, at the right time, for the right reason; and she does so from a firm and stable character (Aristotle, 1985, 1105a30). But she does not, and need not, deliberate about what to do. She need not even be conscious or expressly aware of why she does what she does. For being virtuous,

she has internalized the rules. Doing the right thing in the right way, at the right time, for the right reason has become second nature to her. Aristotle (1985, 1103b) likens virtues to crafts. Anything that can be done can be done well or badly. And to do something well—not accidentally, but as a result of a stable, acquired disposition—is to do it with a trait that is, or is at least analogous to, a virtue. A good harp player knows how to play the harp and normally displays that know-how when playing the harp; a good builder knows how to build well and normally displays that know-how when building.

I suggest that, being sensitive to norms, agents who follow rules blindly exhibit something akin to Aristotelian virtues. Although Aristotle restricts the term 'virtue' to characteristics that make the actions of certain agents morally or intellectually good, I shall use it in a broader sense. Virtues are what make the actions of certain agents (those who standardly do the right things in the right way at the right time for the right reasons) good of their kind. The virtues integral to a practice are various, and some are matters of degree. The propensity to follow the rules of chess at all is a real but minimal virtue in a chess player. The propensity to devise and execute complicated strategies effectively is a greater chess-playing virtue, for it makes one a better chess player.

What are rules for the novice transmute into virtues when they are internalized so that they automatically, unthinkingly guide practice. The attractive element in Ryle's account of multitrack dispositions is that it accommodates sensitivity to circumstances. What Abel and Aristotle add is that the sensitivity in question is not just to the physical, material, or sociological circumstances, but also to the normative circumstances. Human behavior is circumscribed by norms. Being duly sensitive to circumstances involves being sensitive to the norms of the practices one takes part in. Such sensitivity is a part of knowing how to participate in those practices, for the norms govern what may be done, what must be done, and what must not be done within the practice. If this sensitivity has become second nature, we need not deliberate, and may not be able to articulate the norms that constrain and guide us. Still, our behavior is responsive to those norms.

Dispositions have a counterfactual dimension. To ascribe a disposition is to indicate something not only about what does happen, but also about what would happen had circumstances been different. The glass that never is struck and never breaks nonetheless has the disposition of brittleness

if it would break if it were struck. Similarly, I suggest, for dispositions that involve norms. Someone who has internalized the rules of the road automatically and unthinkingly follows them. She does the right thing. But given that her behavior is automatic, what makes it the case that she does it for the right reason? The answer depends on what counterfactuals are true of her. Meg is not only ignorant of the law pertaining to right turns in New York, she is also completely indifferent to it. She would not turn right on red in New York no matter what the law allowed. Although Mark, like Meg, stops at every red light as he drives in the city, he would often turn right on red if the law allowed. His driving behavior is constrained by the law in a way that hers is not. His propensity to modulate driving to the local laws is evidence that he knows how to drive in New York. His disposition is, as hers is not, sensitive to the normative structure of the New York City traffic laws. Displaying that sensitivity constitutes acting for the right reason.

Internalizing the norms of a practice does not just engender a disposition to behave, but a normative disposition—a disposition to hold oneself accountable. Someone who has internalized the norms of a practice considers herself subject to criticism if she violates those norms. She may flinch, or blush, or correct herself, or glance furtively around to see if anyone noticed. She may resolve to do better next time. She may also consider herself entitled to disapprove of, criticize, or correct other participants in the practice who violate its norms, and perhaps to praise or admire those who observe the norms.

The norms that concern us are epistemic norms, norms that bear on thinking or reasoning or arguing well. A fluent speaker of a language typically follows its grammatical rules automatically. She may be unable to articulate the rules she follows. She may even harbor doubts about the correctness of the rules a knowledgeable grammarian proposes. When asked what is wrong with a particular construction, she might have nothing more helpful to say than that it sounds funny. This is surely an instance of following the rules blindly. But a fluent speaker is not a flawless speaker. Occasionally she says something odd. Some odd utterances are simply unexpected. In a discussion of reverberations, Maude says, 'A duck's quack does not echo'. Although her claim is surprising, there is nothing untoward about her utterance. Not considering it problematic, she is not embarrassed about having uttered it and is not inclined to correct it. Nor are other speakers of the language apt to take her to task. Other utterances are factually

incorrect. These may go uncorrected because the speaker is unaware of the error. If she realizes her error, she is likely to rescind or correct her claim. But to become aware, she needs either to acquire new information or to be reminded about information she already has. Then she admits (at least to herself), 'I was wrong to say that Peoria is in Indiana; it is in Illinois'. Her correction has a different semantic content from her original claim. Yet other utterances are grammatically flawed. Here, the error may be obvious to the speaker as soon as the words leave her mouth. Perhaps she says, 'The data demonstrates that the ice caps are melting'. She immediately recognizes that 'data' is a plural noun and requires a plural verb. The correction she makes is to simply change the verb form. The substantive semantic content of the claim remains the same. If she fails to recognize her error, other speakers—even those who know nothing about what the data are or show—can correct her mistake. The recognition of a need for correction shows that the agent takes herself to be answerable to a norm. The kind of correction she makes indicates what norm she takes herself to be accountable to.

An adept reasoner does much the same thing. She constructs sound arguments, draws on relevant evidence, holds herself responsible for flaws in her reasoning. The corrections she makes or accepts, as well as the principles she adduces, are indications of the norms she takes herself to be answerable to.

What may be articulable as the rules and conventions that, from the outside, constitute an agent's rule-following behavior, function as quasi-Aristotelian virtues in the agent's own sense of what he is doing. Once he knows how to play chess, tie a necktie, construct an argument, or speak grammatically, he does the right things at the right times for the right reasons, and does so from a steady disposition. The regularities that characterize his behavior have become second nature. And the right reasons are internalized norms.

Abel maintains that knowing how is more fundamental than knowing that. If so, the model provided by the novice chess player is misleading. She internalized antecedently articulated rules. And she was expressly aware of the rules before she internalizes them. This sometimes happens. But, Abel maintains, often no articulated rules are available. A speaker learns her native language by being brought up in a community where it is spoken. She models her utterances on those of other speakers, and subjects her

linguistic behavior to correction from them. She counts as fluent when, in the opinion of her compatriots, she speaks like a native. That is all it takes. The process can occur without explicit instruction in the rules of grammar. Indeed, it can occur in a linguistic community whose grammar has never been codified. Unlike chess, where the constitutive rules are prior to the practice, grammatical rules precipitate out of the practice. What makes a construction grammatical is that fluent speakers treat it as such. They understand it and do not think it needs correction. Rather than saying that the language learner internalizes the rules, it would be more accurate to say that the linguist externalizes the linguistic virtues of competent speakers.

This sheds light on how to resolve Wittgenstein's worry about rule-following (Wittgenstein, 1953; Kripke, 1982). Like Abel, Ginsborg (2010) recognizes that not all norms are reducible to rules. To evade the skeptical puzzle Kripke identifies as arising from the absence of rules, she argues that some norms are primitive. When asked what comes next in a series, or what the sum of 68 and 57 is, she maintains, there is an answer that we primitively ought to give. That answer is right: other answers are wrong. This is so even though it is possible to articulate a rule that a 'deviant' answer would conform to. Ginsborg characterizes primitive normativity negatively. It is "normativity which does not depend on conformity to an antecedently recognized rule" (2010, 233). She evidently assumes that either normativity is grounded in rules or it is *sui generis*. If there are virtues that are not the internalization of antecedently recognized rules, they are, on her view, primitively normative. We ought to act on them, but there is just no saying why.

This seems wrong. Practices are public; a performer learns to conform to their norms, and is subject to criticism and correction if he fails to conform. Knowing how is, as Abel says, a product of "triangular relations of subject, other subjects or agents, and the world" (2012, 248). A four-year-old child is given the first numbers in the sequence '1, 2, 3' and asked what comes next. She might answer '4, 5, 6'; or she might answer '5, 8, 13'. If she does the latter, her answer is mathematically impeccable; the numbers 5, 8, 13 are the next ones in the Fibonacci series. Still, we want to say, her answer is incorrect. If correctness turns entirely on the rules of mathematics, we have no grounds for criticizing her. But if correctness turns on attuning herself not just to the rules of mathematics but to the practice she is participating in, things are different. Given the practice in effect in her preschool classroom,

she has reason to give and prefer the standard answer. Assuming that the class is not studying arcane mathematical sequences, the immediately subsequent integers are standard. Because Abel recognizes the importance of other agents, he has the resources to assess practices, and assess particular behaviors within practices from a perspective that Ginsborg, who looks only at the agent, her past intentions, and her current mindset, cannot. Abel can say, as Ginsborg cannot, that what makes an action correct is that it is required (or permitted) by a practice, and that practice promotes something that the members of a given community regard as worth achieving. He can look outward to the public good that the norms achieve, rather than exclusively inward to what the agent thinks she is trying to do.

We are brought up in practices, and often learn from examples. So, unlike the chess novice, we may never have learned the "rules" of practices we participate in. Indeed, such practices may have few articulable rules. We were simply socialized to emulate adept practitioners. This is why native speakers can fail to recognize, much less be able to state, the grammatical rules of their language. Outsiders—perhaps linguists or anthropologists—may formalize the normative regularities that they find in the practice. But practices proceed without expressly formulated rules. The fact that practitioners not only typically conform to certain regularities but teach others to conform, and correct or disparage the behavior of nonconformists, makes it manifest that norms are operative. Normatively informed behavior, then, is more extensive than explicitly rule-governed behavior.

The novice tennis player hits hundreds of balls in learning how to serve. The novice cellist saws away for untold hours in learning how to bow. Abel insists that learning by doing is typically needed to acquire know-how. Aristotle agrees. One becomes just by doing just acts. But Aristotle emphasizes that "the sources and means that develop each virtue also ruin it" (1985,1103b9). What the novice does repeatedly can be repeatedly done well or repeatedly done badly, or intermittently done well and done badly. Only if it is done well will repetition lead to the development of a virtue. Rote repetition will not do. The novice's actions must be monitored (by himself or others), and encouraged or corrected as appropriate.

That being so, the student needs standards by which to judge. But if the actions are ones for which there is no adequate articulable rule, where do they find the standards? They appeal to exemplars—telling instances—where the action is manifestly well done. With or without the aid of a

mentor, the novice emulates the behavior of those who already know how to perform the actions he seeks to master. We will discuss exactly what such emulation involves in chapter 9. For our purposes here, it is enough to recognize that it is an avenue to acquiring know-how.

The epistemic norms of knowing how are quasi-Aristotelian virtues; they are goods realized in action and may be uncharacterizable apart from the practices they belong to or the ends they promote. When this is so, it is impossible to state exactly what knowing how involves. But this does not make knowing how epistemically inaccessible, or learning how mysterious. A student can learn how to perform the action by emulating exemplary performances of it. Once his behavior accords with his exemplar, he knows how to do the action in question. Knowing how, then, is not inscrutable so long as we have the resources to identify and interpret exemplary instances.

Although knowing how is subject to epistemic norms, it is fundamentally different from knowing that. It is not an attitude toward the truth of a sentence or proposition. That Fred knows how to tie a necktie is either true or false. But his knowing how is not itself truth evaluable. Nor is knowing how a matter of belief. Molière's M. Jourdain knew how to speak in prose, even though he did not believe that he knew how to do any such thing. Henceforth, for clarity, when I use the term 'knowledge' and related terms without qualification, I will be speaking of knowledge that.

I said earlier that an understanding of a topic involves know-how. We now can see that it involves knowing how to wield the commitments that bear on the topic—how to draw the inferences and perform the actions that the understanding licenses. That know-how is, in Rylean terms, a multitrack disposition—an ability and propensity to make certain inferences and eschew others, to perform certain actions and refrain from others, to engage in and endorse certain forms of higher-order evaluation and criticism and avoid and repudiate others. The network of commitments functions normatively. It constitutes the permissions and prohibitions that constrain and facilitate the epistemic agent's relevant thoughts and actions (see Hetherington, 2011). In Sellars's (1963, 169) terms, the network structures a space of reasons. If the agent has internalized the commitments, she may deploy them blindly. In that case they function as quasi-Aristotelian intellectual virtues.

Earlier I characterized epistemic acceptance of φ as a willingness and ability to use φ as a basis of inference or action when one's ends are cognitive.

That ability is a matter of know-how. When an epistemic agent accepts a proposition, rule, norm or account, she knows how to use it to further her relevant cognitive ends.

Factivity

Knowledge is factive in that one does not know that *p* unless '*p*' is true. So if propositional understanding is knowledge of dependencies, it is factive. We would not say that Lavoisier understood that phlogiston is the volatile element in combustion, since it isn't. But objectual understanding concerns topics rather than individual propositions. So what it means to claim that objectual understanding is factive is a bit harder to make out. Perhaps objectual understanding is factive if it is impossible to understand a topic—say, the history of Athenian warfare—unless one is committed to some identifiable, suitably comprehensive proposition that is true. That proposition might be the long conjunction of all the atomic propositions belonging to the coherent body of information that constitutes the understanding. (This parallels the interpretation of coherence theories of knowledge as requiring the truth of the conjunction of the atomic propositions in a coherent system of beliefs.) On such an account, understanding would be a sort of knowledge, namely the knowledge captured in long, subject-matter-connected conjunctions.

This proposal faces several problems, beyond its difficulty in accommodating science. First, it does not respect the requirement that the understander grasp the relations among the atomic propositions—that the understander appreciate how they bear on one another. Although the body of information understood must be coherent, if the understander need only know the conjunction, there is no requirement that she grasp the coherence. Second, it does not accommodate the insight that the student who understands geometry can do more with it than the student who knows the axioms, the main theorems, and their derivations by rote. Third, it is at odds with the recognition that not all the propositions that comprise a genuine understanding of a topic need be true. We would be inclined to say that a historian understood the Athenian victory even if he harbored a few relatively minor false beliefs about the matter.

Although Kvanvig (2003) considers understanding factive, he concedes these points. He does not believe that understanding a topic consists in

believing a long conjunction. Nor does he insist that every proposition in the comprehensive body of information that constitutes an epistemic agent's understanding be true. Rather, he maintains, an agent cannot understand a topic unless most of the propositions and all of the central propositions that constitute her coherent take on that topic are true. He allows that a few peripheral falsehoods can degrade an agent's understanding of a topic without destroying it. That understanding is factive in this sense is the thesis I want to dispute.

Unlike knowledge, understanding admits of degrees. A freshman has some understanding of the Athenian victory, while her teaching fellow has a greater understanding and her professor of military history has an even greater understanding. Epistemology should explain what such differences in degree consist in. A factive account can easily recognize three dimensions along which understanding can vary: breadth, depth, and significance. The professor might have a broader understanding of the Athenian victory, being able to embed his coherent body of largely true beliefs into a more comprehensive account of Greek warfare. He might also have a deeper understanding. In that case, his web of belief is more tightly woven; it contains more propositions, and/or more nontrivial inferential connections among propositions. But according to a factivist epistemology, both the student and the professor understand the Athenian victory insofar as they grasp coherent bodies of predominantly true propositions, and believe the propositions that belong to those bodies. The student and the professor might weigh the facts differently. Even if each believes a given truth and incorporates it into a coherent account of the matter, the professor might consider it highly significant, while the student considers it just another truth about the battle. If the truth really is significant—if, for example, it is central to explaining how the Athenians' novel battle formation contributed to the victory—then the professor's better understanding consists in his appreciating the significance of the truth, not merely in his recognizing that it is a truth. Strevens (2008, 2016) is a factivist who allows for what I call felicitous falsehoods; he holds that not all (or even most) of the elements of an objectual understanding need be true. But, he insists, those that are difference makers must be true. So the professor's broader understanding consists in his incorporating more difference makers into his account; his deeper understanding consists in there being tighter connections among the difference makers. And what is important is not the

weight the professor assigns to every element of the account, but the weight he assigns to the difference makers. Factive accounts, then, can accommodate some differences in degrees of understanding.

However, there is another dimension along which we can measure greater and lesser understanding which factivists cannot take on board. It involves conceding that some accounts, even though they are not true, nonetheless display a measure of understanding. The growth of understanding often involves a trajectory from contentions that, although strictly false, are in the same general neighborhood as contentions that are closer to the truth. The sequence may terminate in truths. But even the earlier steps in the sequence should fall within the ambit of epistemology. For they are, to an extent—often to a considerable extent—cognitively valuable.

An eight-year-old's understanding of human evolution might include as a central strand the proposition that human beings descended from apes. A more sophisticated understanding has it that human beings and the other great apes descended from a common hominid ancestor who was not, strictly speaking, an ape. The child's opinion displays some grasp of evolution. It is clearly cognitively better than the belief that humans evolved from butterflies. But it is not strictly true. Since it is central to the child's take on human evolution, factivists like Kvanvig must conclude that her take on human evolution does not qualify as understanding. Unless we can plausibly abstract from the child's claim about apes to some more generic 'ape-like creatures', which may be difficult if she does not distinguish between apes and ape-like creatures, Strevens cannot take her to harbor an understanding either. In that case, epistemology need give no account of what makes the child's grasp of evolution cognitively good, or cognitively better than a view of her classmate who holds that humans evolved from butterflies. This may not appear to be a major objection, since the child's opinion is fairly naïve. Perhaps it is reasonable to conclude that she does not (really) understand. But the pattern exhibited in this case is endemic to scientific education. We typically begin with rough characterizations that properly orient us toward the phenomena, and then refine the characterizations as our understanding of the science advances. Think of the trajectory from naïve folk physics through Newtonian mechanics to relativity and quantum mechanics.

Kvanvig (2003) believes that when we construe such a take on a subject as understanding, we use the term 'understanding' in an honorific sense,

just as we use the term 'knowledge' in an honorific sense when we speak of 'the current state of scientific knowledge', while conceding that some of what belongs to the current state of scientific knowledge is false. Such honorific usages of epistemic terms, he maintains, are extended usages that fall outside the scope of epistemology. Only in an extended sense, then, does the child have any understanding of evolution.

Perhaps it would be feasible to dismiss such uses of 'understanding' as merely honorific if they applied only to young children or neophyte students of a subject. I think otherwise, for I believe that epistemology should have something to say about what makes the views of the child who thinks that humans evolved from apes better than the views of a child who thinks that humans evolved from butterflies. But the main problem with the contention that understanding is factive is that the trajectory displayed as she moves from the naïve view of human evolution up to the view held by the evolutionary biologist is the same trajectory that science displays in the sequence of accounts it develops. It is intellectually arrogant to insist that it is only in an extended sense that a contemporary biologist understands morphological evolution, given that the precise role of cis-regulatory sequences (noncoding sequences of DNA) is currently unclear. That there are always unclarities at the cutting edge of inquiry does not diminish the epistemic status of the understanding that brought us to the current cutting edge.

A central tenet of Copernicus's theory is the contention that the Earth travels around the sun in a circular orbit. Kepler improved on Copernicus by contending that the Earth's orbit is not circular but elliptical. Newton realized that because the planets exert gravitational force on each other, the Earth's orbit is not quite elliptical. Having abandoned the commitment to absolute space, current astronomers can no longer say that the Earth travels around the sun simpliciter, but must talk about how the Earth and the sun move relative to each other. Despite the fact that Copernicus's central claim was strictly false, the account it belongs to constitutes a major advance in understanding over the Ptolemaic account it supplanted. Kepler's and Newton's accounts are further advances in understanding, and the current account is yet a further advance. The advances are clearly cognitive advances. With each step in the sequence, we understand the motion of the planets better than we did before. But no one claims that science has as yet arrived at the ultimate truth about the motion of the planets. Should we say

that the use of the term 'understanding' that applies to such cases should be of no interest to epistemology?

One might follow Kvanvig and contend that even here the use of 'understanding' is honorific. We apply the term in these cases only because we think that the strides scientists have taken are on the way to the truth—the comprehensive, general account of celestial motion that gets it right. In effect, current science borrows its epistemic status from its descendants. Sellars (1963) argues that in a mature science, later theories should show why their predecessors were right to the extent that they were. So later theories are supposed to at least partially vindicate their predecessors. Where this does not happen, we are apt to conclude that the earlier scientists did not understand the phenomena that their theories purported to explain. We do not, for example, consider phlogiston theorists to have had any understanding of combustion.

Suppose we concede this point. Let us admit that in saying that the various astronomical accounts embody an understanding, we take out a lien on the future of science. Still, I would urge, the cognitive achievements embodied in such accounts should be a central concern for epistemology. Even if we do not yet have (and may never get) the truth, we have made real cognitive progress. We understand the motions of celestial bodies better than our predecessors did. Epistemology should explain what makes current understanding better. If we say that the uses in question are honorific, epistemology should explain why certain attitudes toward certain topics are worthy of honor.

Another aspect of science is even more troublesome for the factive view—namely, science's penchant for idealization. Science streamlines, simplifies, and distorts. It devises and deploys comparatively austere models that diverge from the phenomena it seeks to explain. The ideal gas law accounts for the behavior of gases by describing the behavior of a model gas composed of dimensionless, spherical molecules that are not subject to friction and exhibit no intermolecular attraction. The Hardy—Weinberg formula describes the distribution of genes in an infinite population whose members mate randomly. There are no such things. Indeed, as far as we can now tell, there could be no such things. But the ideal gas law figures in the understanding provided by thermodynamics and the Hardy—Weinberg formula figures in the understanding provided by population genetics. Scientists purport to understand the phenomena in terms of these laws. The

contention that such laws degrade but do not destroy understanding is simply not credible.

Idealization is not taken by scientists to be an unfortunate expedient, but rather to be a powerful tool. Although they expect today's idealizations to be replaced, they harbor no expectation that in the fullness of time idealizations will be eliminated from scientific theories. So the 'promissory noteishness' that we saw in talk about the progress in our understanding of celestial motion has no role here. Elimination of idealizations is not a desideratum. Nor is consigning them to the periphery of a theory. It is simply not the case that the bodies of information that constitute scientific understanding are, or that their ultimate successors can be expected to be, composed of truths, with any residual falsehoods occurring only at the periphery. The ideal gas law lies at the core of thermodynamics, and some such model is likely to lie at the core of any successor to current theories.

I concede that many of the propositions that fall within the scope of 'the current state of scientific knowledge' are not strictly *knowledge* because they are not true. If we are being scrupulous, we should probably not speak of the current state of scientific *knowledge* unless we are convinced that the propositions in question are true. But the contention that 'understanding' is factive does not have the same strong support from ordinary language. When 'understanding' applies to large, often somewhat inchoate bodies of information, it takes a direct object that is not a proposition. Jill understands the Athenian victory, Joe understands the motions of the planets, Pat understands the New York subway system. We typically acknowledge that people can have a measure of understanding even if the contentions making up the bodies of information they endorse diverge somewhat from the truth. So our ordinary use of 'understanding' as applied to bodies of information does not supply a strong argument for a factive analysis. There is, however, a recognition that 'understanding' is a cognitive success term. If I am going to reject the factive analysis, I need some way to identify or characterize that cognitive success.

4 The Fabric of Understanding

In the previous chapter, I argued that disciplinary understanding—indeed, any sort of systematic understanding—is objectual. Here I argue that it follows that accounts must be assessed holistically. Rather than asking whether each component sentence expresses a fact, we should ask whether the account as a whole is in reflective equilibrium. If so, I will urge, it is acceptable in the epistemic circumstances. After explicating reflective equilibrium I show how adopting it as a standard of acceptability enables us to appreciate the way individually weak considerations can weave together to make a strong case. I then consider objections. Although my theory is not a pure coherence theory, coherence plays a major role. I discuss familiar worries about coherence theories and explain why they do not undermine my position.

Reflective Equilibrium

The move to holism raises questions about justification. After several hundred years of theorizing, it is fairly easy to see why empirical evidence for a hypothesis is truth conducive. Relevant empirical evidence increases the probability that the hypothesis is true. If an account is construed as a long conjunction whose conjuncts are statements of its commitments, it might seem that the same holds of it. The total evidence increases the probability that the long conjunction is true, thereby justifying the conjunction. The difficulty is that under such a construal, even the best accounts turn out to be unjustified. If p and q are evidentially independent, and each has a probability of less than 1, the probability of the conjunction p & q is less than the probability of each conjunct. The longer the conjunction, the lower the probability. Comprehensiveness is considered an epistemic virtue, but

if justification consists in probability raising through empirical evidence, it is not. Holism ought not then contend that the justification of the wholes derives from the justification of the parts. Or at least it ought not contend that the probabilistic justification of the parts increases the probability that the whole is justified as well.

The challenge, then, is to explain how systematic interconnections give rise to justification—how the fact that considerations dovetail affords reason to accept them. In previous works I argued that a system of cognitive commitments in reflective equilibrium is epistemically tenable (Elgin, 1996, 2011). This is a procedural account of epistemic tenability. An account is tenable just in case it is, or is rationally reconstructable as, a result of a process of adjudication that brings a collection of initially tenable commitments into reflective equilibrium. It is as reasonable as any available alternative in light of those initially tenable commitments, and the commitments that comprise the account are in general more reasonable in light of one another than they would be alone. I do not claim that reflective equilibrium shows an account to be true or even to probably be true. I claim only that a system of epistemic commitments in reflective equilibrium is the best we can do in the epistemic circumstances. But that is enough to make it acceptable right now. This needs to be spelled out.

Inquiry, as Quine insists, always begins *in medias res*. We start with opinions, values, methods, and standards that we consider relevant and that we are inclined to credit. Although we recognize that they are less than wholly satisfactory as they stand, they comprise our current best take on the matter under investigation. These commitments can include hunches, rules of thumb, superstitions, and old wives' tales as well as confirmed generalizations, proven theorems, firmly established principles, and solid evidence. The only requirement at the outset is that we have some inclination to accept them. Because we do, they are initially tenable. Clearly that inclination does not endow them with a high level of tenability. To be initially tenable is not to be presumptively acceptable, or prima facie acceptable, or ceteris paribus acceptable. It is merely to have some very slight measure of epistemic standing. If a consideration is initially tenable, we need a reason to give it up. But since initial tenability is weak and precarious, even a weak reason may suffice to abandon it. So construing a consideration as initially tenable is not saying much in its favor. Reasons to give up initially tenable commitments are legion.

That being so, why should we think that the mere fact that we have some inclination to accept a consideration gives it any level of epistemic standing? This much can be said in its favor. Since beliefs form the basis for action, the success of our actions affords some evidence that the beliefs that underwrite them are true enough. And the persistence of the inclination, however slight that inclination, is evidence that the consideration has not unduly impeded our actions. Granted, brute luck may explain why certain untoward initially tenable commitments have not yet caused our efforts to go awry. Or the untoward commitments may bear on factors that are too distant from day-to-day experience for the success or failure of our actions to make much difference to their continued acceptance. Or we may have made compensating errors. Nevertheless, that they have not caused enough trouble to induce us to abandon them is at least a minuscule mark in their favor. Moreover, we learn from experience. When we learn that premonitions tend not to be borne out, we cease to credit them. Although we may continue to experience feelings of foreboding, once we lose any inclination to credit their deliverances, they cease to be initially tenable. The fact that we have not yet had reason to give a commitment up is then a small reason to accept it.

Our initially tenable commitments pertaining to a topic may at the outset constitute a motley collection; they include our current convictions about the topic, the methods for investigating it, the standards to which a cogent account of that topic should be held, and the cognitive ends such an account should serve. Taken together they are apt to leave relevant questions unanswered and relevant problems unsolved. The elements of the collection may be in tension with one another. Our convictions may fail to satisfy the standards we consider them subject to. We may have incompatible or noncotenable convictions and/or mutually unsatisfiable cognitive desiderata. Considerations that seem independently plausible may seem jointly implausible. To achieve an acceptable account, we correct, reject, refine, and extend the commitments we started with. Some, such as 'Fish have gills', may come to be seen as crude approximations. Others, such as 'Objects falling near the Earth fall at a rate of 32 ft/sec^2', turn out to be acceptable only within certain limits—for example, where air resistance is negligible. Yet others turn out, like Frege's Basic Law V, to be hopeless, regardless of their antecedent plausibility.

Nor is our task just to corral and tame our initially tenable commitments. For even duly pruned and connected, a constellation of initially

tenable commitments is apt to be gappy. Achieving reflective equilibrium often requires integrating into our emerging account commitments that we have no independent inclination to believe. The scientific community was originally committed to the existence of positrons not because it had any direct evidence of them, but because it was strongly committed to symmetry principles and to the existence of electrons. If electrons exist and symmetry holds, then there exist positively charged counterparts to electrons—that is, positrons. The relation between the embedding account and the positron hypothesis is one of mutual support. The hypothesis gains tenability from its place in the account; the account increases in tenability as well. Given the other commitments, the asymmetry that would result from denying the electron a positively charged counterpart would be an irregularity that required explanation that physics was in no position to provide. By rejecting, correcting, revising, and augmenting, we bring commitments into accord. Because the elements of the resulting account are reasonable in light of one another, they are in equilibrium; because the account as a whole is as reasonable as any available alternative in light of the relevant antecedent commitments, its equilibrium is reflective. It is an account that the community of inquiry can, on reflection, endorse.

Initially tenable commitments perform two functions. Being, at the outset, our current best take on how matters stand and on how to figure out, measure, and assess how they stand, they provide the raw material out of which to construct tenable accounts. They capture what we start out thinking we understand about the domain. In constructing a better account, we leverage that understanding. Their mutual support makes for equilibrium. By playing a second role, initially tenable commitments ensure that the equilibrium is reflective. They serve as touchstones against which to assess revisions.

Their capacity to play the second role may seem doubtful. I conceded that the initial commitments are to one degree or another inadequate. To use them as a touchstone for judging revisions may therefore appear misguided. How can an account be vindicated by being shown to be reasonable in light of considerations that are acknowledged to be unsatisfactory? Because our initial commitments constitute our previous best guesses about the topic, we use those commitments as a threshold for assessment. The account we arrive at should be recognizable as an improvement upon them. We engage

in inquiry because we suspect that they are inadequate and in need of correction, fortification, elaboration, and/or extension. We revise our previous commitments when we arrive at something we consider better. The requirement that an acceptable account answer to our initial commitments is not a requirement that it incorporate or validate them. But where it does not, it should, at least for the most part, show why they seemed as reasonable as they did when they did.

Epistemic advancement, then, consists in improving on the commitments we currently hold, where improvement itself must be measured by current standards. We are in no position to adopt a God's eye view and say that in absolute terms one account is better than another. But we do not need a God's eye view to characterize cognitive changes as improvements. We can recognize, for example, that Newton's account of planetary motion is an improvement over Kepler's because Kepler's account accommodates only the relation between the Earth and the sun, while Newton's also accommodates the gravitational attraction of one planet on another. We appreciate that having acknowledged the significance of the sun's gravitation, we have no grounds for ignoring other sources of gravity.

The method of reflective equilibrium is dialectical and its results are provisional. We judge our commitments and potential revisions of our commitments against one another to see which combination seems best on balance. But we also judge our criteria for being best on balance by seeing if they yield verdicts that we can on reflection endorse. Reflective equilibrium can be upset by new findings. However well it fits with our other ornithological commitments, the conviction that birds cannot hover is discredited by the discovery that hummingbirds hover. (We might accommodate the discovery by abandoning the conviction that birds cannot hover or by abandoning the conviction that hummingbirds are birds. But something's got to give.) No account or component of an account is irrevocable, but an account in reflective equilibrium is acceptable until problems emerge or improvements are envisioned. The justification for this verdict lies in the fact that the commitments that comprise such an account are collectively at least as reasonable as any available alternatives in the epistemic circumstances.

The requirement that each element be more reasonable in light of the others than it was alone may seem too demanding. After all, one might think, when an observer sees a camel before her very eyes, in good light, in

the middle distance, and uncamouflaged, her belief that she is looking at a camel needs no further support. I agree that her perceptual deliverance alone yields a presumption that she sees a camel. That she is actually looking at a camel is initially tenable. But that deliverance does not stand alone. It gains or loses backing from collateral commitments. If the perceiver harbors an account (even a Moorean commonsensical account) that includes the beliefs that camels are visible and that organisms like her are capable of seeing them, her perceptual belief is more reasonable than it would be if she had no views about the visibility of camels or if she thought that they were, like proverbial pink elephants, figments of the imagination that appear only to the inebriated. Her belief is strengthened if backed by a plausible reason for why there might be a camel in front of her. Is she at the zoo, at a circus, wandering the Sahara? If not—if, for example, her experience is as of a camel on Main Street—she might need to discredit the hypothesis that her apparent camel sighting is a hallucination or a dream. If she has no reason to think that she is taking mind-altering drugs or is under posthypnotic suggestion, the absence of such competing potential explanations contributes to her warrant. But if she has no idea whether camels are visible, whether she has been taking mind-altering drugs, whether she has recently been hypnotized, or whether human beings can see, her belief that she sees a camel is epistemically precarious. Much of the support we depend on in cases where things seem obvious is so deeply tacit and uncontroversial that we may be unaware of it (Adler, 2002). But that tacit background contributes to the tenability of even the most seemingly obvious deliverances.

Some elements of a tenable account have no independent presumption in their favor. They gain all their support from their place in the account they figure in. Had physics abandoned or attenuated its commitment to electrons or to symmetry, early twentieth-century physicists would have had no reason to believe that positrons exist. Similarly, Robinson Crusoe, seeing what he takes to be human footprints in the sand, concludes that he is not alone on the island, basing his conclusion on the conviction that the presence of what look to be footprints is a strong indication of the recent presence of human feet. If he discovered that the configuration that he took to be human footprints could as easily be produced by local sand crabs, his belief would lose its support.

Mutual Support

A critical question is how reflective equilibrium provides justification. Why should we think that mutual accord among our commitments, backed by a tie to our previous commitments affords any indication of how things are? Perhaps this will become clearer if we consider a case.

Yesterday Meg's Latin book was stolen from her school locker. Three students may have witnessed the theft. None of them is very reliable. Anne is given to proving theorems in her head and tends to be oblivious to her surroundings when preoccupied with a tricky proof. To compensate for her habitual distractedness, she draws plausible inferences about mundane events, and often fails to register whether an opinion is due to observation or to such an inference. Being rather vain, Ben frequently does not wear his glasses. Like Anne, he draws plausible inferences about events around him, and tends not to remember having done so. Chauncy is simply a liar. Presumably he knows when he is speaking sincerely, but given the fluency and frequency of his lies, virtually nothing he says is trustworthy. Not surprisingly, the social circles of the three students do not intersect; none would deign to speak to the others. When questioned about the theft, Anne and Ben report what they think they saw, but confess that they are not sure what they actually witnessed and what they inferred. Chauncy insists that his report is accurate, but in view of his record, his claim is suspect.

Individually, none of the reports would count for much. Had only one of the putative witnesses been present, the most we could reasonably conclude would be that the thief might fit the description. But all three reports agree, and agree in alleging that the thief had an unusual appearance: he had spiked green hair. This makes a difference. Even though individually each report is dubious, and the probability of a green-haired textbook thief is low, the fact that the three reports provide the same antecedently improbable description inclines us to believe it. Their accord evidently enhances the epistemic standing of the individual reports (C. I. Lewis, 1946, 346). We have more reason to believe each of them in light of the others than we have to believe them separately. The question is why? How can multiple statements, none of which is tenable on its own, coalesce to yield a tenable conclusion? How can their relation to other less than tenable claims enhance their tenability?

Given the unreliability of the witnesses, we would not be surprised if they were wrong about the thief. But we would not expect them to all be wrong in the same way. Their agreeing needs an explanation. Were they coconspirators, the explanation would be straightforward: They plotted to tell the same tale. But not being on speaking terms with one another, they are unlikely to be in cahoots. If the descriptions they provided fit a relevant stereotype, a penchant for plausibility might explain their accord. But someone with green spiked hair is far from anyone's stereotype of a textbook thief. So despite Anne's and Ben's propensity to draw inferences based on plausibility, their descriptions of the thief seem not to result from such an inference. Evidently the best explanation of the agreement is that the reports are accurate (Lipton, 2004).

It is not just our ability to exclude obvious alternatives that leads us to credit the allegation. A variety of collateral considerations support it. Some bear directly, albeit weakly, on the content of the claim. The principal's secretary dimly recalls seeing an odd-looking stranger lurking in the hallway, but cannot remember what made the stranger look odd. The custodian thinks he recently saw a container of hair dye in the trash, but cannot remember which day. Although the tentativeness of these reports makes them less than wholly creditable, they are suggestive enough to buttress the eyewitness testimony. Other collateral considerations concern the witnesses and their circumstances. Book thefts are observable events, so there is nothing inherently dubious about a claim to have seen someone steal a book. The light and sight lines were such that the witnesses could have seen what they report from where they said they were standing. The witnesses, teenagers themselves, are adept at recognizing furtive adolescent behavior. None was subject to psychological experiments with implanted memories. None was on drugs. And so on. Separately, these factors count for little. Either their credibility is low, their bearing slight, or their evidential support minimal. But they coalesce to make a solid case. This suggests that the epistemic tenability of the several reports and the conclusion they sanction derives from their mutual supportiveness.

Although our focus is on the status of the allegation, it is really the account as a whole that is or is not acceptable. Many of the justificatory relations are reciprocal. The allegation is acceptable only if (at least much of) the rest of the constellation of supporting considerations is. Since the eyewitnesses are individually unreliable, and the contentions of the collateral

witnesses or the bearing of those contentions on the theft is tenuous, the acceptability of the testimony likewise depends on the acceptability of the allegation. The epistemic status of the allegation is inseparable from the status of the rest of the story. Some of the background information may be separately secured, but to a considerable extent, the various components of the story stand or fall together. Epistemological holism contends that the primary unit of acceptability is an account comprised of mutually supportive elements. The acceptability of individual elements largely derives from the acceptability of the account they belong to or follow from. This suggests that epistemic justification is at base a property of a suitably comprehensive, coherent account, when the best explanation of its coherence is that the account is at least roughly correct. The epistemic justification of individual claims derives from their figuring in such an account.

Coherence evidently plays a major role. Although there is no universally accepted criterion of coherence, at least this is required: the components of a coherent account must be mutually consistent, cotenable, and supportive. Logical consistency is not enough. Two claims can be consistent with one another because they are mutually irrelevant. 'Napoleon died in exile' and 'Penguins cannot fly' are consistent with one another, but neither lends support to the other. There is no more reason to accept their conjunction than there is to accept the separate conjuncts. Since both cotenability and supportiveness are matters of degree, coherence is too. So if it can be shown that epistemic justification requires coherence, there remains the question of how coherent an account must be in order for it to be epistemically justified. Before facing that worry, though, other challenges need to be met. At least two worries immediately arise. The first is that coherence is too demanding an epistemic requirement. The second is that it is not demanding enough.

Even where we take ourselves to be on solid ground, contravening considerations are not uncommon. Mrs. Abercrombie, the aging geometry teacher, says that during the relevant period she saw a young man sporting a green hat. A green hat is not green hair, so strictly speaking her report is in tension with the reports of the other putative witnesses. Mr. Polk, the hall monitor, insists that no one was in the corridor at the time of the alleged theft. Mr. Miller, the classics teacher, disputes the allegation on the strong sociological grounds that students do not steal Latin books; they shun them. These reports are clearly relevant to and at odds with the emerging

account of the theft. If we incorporate them into it, we render it incoherent. But they are initially tenable, so we have no legitimate reason to disregard them. The problem is this: The discussion so far suggests that the credibility of the various claims comprising an account depends on how well they hang together. If so, the failure of other, equally relevant information to cohere with them threatens to discredit the account.

Although true, this is not so daunting as it appears. The immediate threat of incoherence comes from assuming that we must take seemingly contravening considerations at face value, incorporating them into the account of the theft as they stand. We need do no such thing. Rather, we assess contravening considerations just as we do the rest of our evidence. Recall that we did not take the eyewitness reports at face value. We initially deemed them suspect because our background information indicated that the informants are unreliable. The credibility of the reports increased because of their agreement with one another and the support provided by collateral information. That agreement gave us reason to think that the general unreliability of the witnesses did not affect the standing of these particular reports. Contravening considerations are subject to similar assessments. Mrs. Abercrombie, being dreadfully near-sighted and woefully out of date, cannot even imagine, much less see, that a green thatch on someone's head is his hair. That being so, her characterization of the suspect as wearing a green hat seems close enough to count as supporting rather than undermining the original allegation. It is, then, true enough. Although Mr. Polk flatly denies what the others say, there are reasons to doubt his veracity. The three putative eyewitnesses purportedly saw each other in the corridor during the interval when Mr. Polk denies that anyone was there, so his contention is dubious on independent grounds. Since he occasionally goes AWOL to smoke a cigarette, it is not implausible that he was absent when the theft occurred. Mr. Miller's argument cannot be so easily discredited. The prior probability of a Latin book theft is slight. But the book is gone. Meg put it in her locker when she arrived at school. It was not there when she returned. None of her classmates admits to having borrowed it. Even if Latin books are not normally attractive targets for teenage thieves, the book's having been stolen may explain its absence better than any available alternative would. Just as other considerations compensate for the improbability of a green-haired thief, other considerations compensate for the improbability of a Latin book's being stolen. In determining the acceptability of a claim,

we assess the considerations that afford evidence pertaining to its tenability. This is not always a simple yes/no matter. We may find that although an evidence statement is unacceptable or unsupportive as it stands, with suitable modifications, it would be. And we may find that the modifications themselves are acceptable. Coherence remains crucial. Sometimes it is achieved directly, sometimes by discrediting or disarming threats.

Although we are initially inclined to take putatively eyewitness testimony at face value, we reject Mr. Polk's claims when we find that they conflict with the testimony of the other witnesses, and we reconstrue Mrs. Abercrombie's claim so that it supports rather than conflicts with the contention that the thief had green hair. Since none of the school's students, faculty, or staff has green hair, we entertain and finally accept the hypothesis that the thief was an outsider, despite our initial conviction that it was an inside job. If our methods prove wanting, or our standards inadequate, they too are subject to modification or rejection. If we seek to understand the theft rather than merely identify the thief, we may appeal to psychological and sociological insights about motivation and the conditions that give rise to it.

The coherence that affords epistemic justification is not just coherence among object-level considerations. We have higher-order commitments about what sorts of object-level considerations are trustworthy, what level of credibility they enjoy, how they ought to mesh, and what to do when commitments clash. These higher-order commitments supply reasons to revise or reject some contentions rather than others when conflicts occur. The coherence that constitutes epistemic justification is something we achieve, not something that simply falls out of the relations in which our object-level deliverances happen to stand to one another.

Although epistemology generally focuses on the beliefs of a single individual, I began with a public case because the otherwise unlikely agreement of independent witnesses clearly shows how the best explanation of the coherence of a given body of claims enhances their acceptability. The case of a single individual can be trickier. People sometimes confabulate. This is a matter not of lying or intentionally seeking to mislead, but of self-deception. A confabulator composes a coherent narrative by unconsciously ignoring, bracketing, or downplaying factors that detract from the story he seeks to construct. It might seem that such self-deception undermines my case for epistemic holism, since an account that achieves coherence by

bracketing, ignoring, or misinterpreting relevant factors has no claim to be epistemically acceptable.

But coherence through confabulation is not so easy to achieve as philosophers imagine. Adler (2002, 67–72) argues that self-deception typically undermines coherence. Confabulators achieve coherence among their first-order beliefs by sacrificing coherence between first- and second-order beliefs. A doting father deceives himself into thinking that his daughter's dismal grades are due to her dimwitted teachers' failure to recognize her quirky brilliance. To sustain this belief he needs to overlook factors that he normally takes to bear on the assessment of student performance—her dreadful study habits, her incomplete homework, her ignorance of the most basic facts about the subjects she is studying. He may 'explain away' these factors to his own satisfaction, thereby sustaining his self-deceptive belief. Nevertheless, he weakens the overall coherence of the relevant neighborhood of his belief system by carving out exceptions for a special case. His thinking violates his otherwise accepted methodological commitments about how to assess such matters. He would not, for example, invoke them to account for his paperboy's poor grades. This is not to say that the self-deceptive agent can recognize his error. He has, after all, managed to deceive himself. But because self-deception weakens coherence, it does not automatically discredit the contention that coherence plays a major role in justification.

Even if mundane self-deception is not a serious worry, delusions might pose a stronger challenge. Suppose a delusional mental patient believes he is Napoleon in exile. He interprets all of his experiences in the locked ward in terms of his delusion. He takes the nurses, doctors, and attendants to be lackeys, courtiers, and palace guards; visitors to be loyal subjects; those who refuse to do his bidding to be traitors; those who tell him he is ill or mistaken in his beliefs to be part of the plot to prevent him from retaking the throne. Unlike the self-deceptive father, he does not carve out special exceptions. He explains all of his experiences in terms of the delusion. Whenever a tension occurs, he rejects or radically reinterprets the deliverance that conflicts with his self-identification with Napoleon. His beliefs may well be more coherent than the self-deceptive father's. But notice how deeply deluded he would have to be to achieve anything like the level of coherence that normal epistemic agents effortlessly achieve. He would have to believe that he and everyone around him was speaking French rather than English.

He would have to believe that oil lamps or candles rather than electric lights were illuminating the area. He would have to believe that the vehicles he sees are horse-drawn carriages rather than cars and trucks. He would have to construe computers, televisions, and elevators as early nineteenth-century artifacts. He would probably have to blind himself to a vast number of familiar items—Velcro fasteners, ballpoint pens, cell phones, and so forth. I do not deny that this is possible. But it is important to see how thorough-going a delusion must be to retain coherence. It is doubtful that he could achieve anything like the level of coherence that ordinary clusters of beliefs typically exhibit.

Obviously, when epistemic agents are self-deceptive or deluded, the coherence of their beliefs is not explained by their being true or true enough. If it is hard to tell whether they are deceiving themselves, it is hard to tell whether coherence confers epistemic standing on their beliefs. But to understand how, why, and when coherence engenders credibility, it is best to put this complication aside. Then we see that the story I have told about the book theft could be told of a single epistemic agent as well. If the agent's account reaches an acceptable threshold of coherence and the best explanation of that coherence is that it is at least true enough, then so long as she has no overriding reason to think otherwise, she is justified. Anne is aware of what she thinks she saw, and what she thinks the other witnesses report. She is privy to the relevant background information about apparent sight lines and the like. Since her various relevant cognitive commitments mesh and the best explanation of their meshing is that they are at least roughly right, according to my epistemological holism, she is justified in accepting them.

The second worry is that coherence is not demanding enough, for it can readily be achieved through epistemically illicit means. A good nineteenth-century novel is highly coherent, but not credible on that account. Even though *Middlemarch* is far more coherent than our regrettably fragmentary and disjointed views about the book theft, the best explanation of its coherence lies in the novelist's craft, not in the truth or approximate truth of the story. The novel's coherence affords virtually no reason to think it is anywhere near literally true. This is surely right. But rather than taking this objection to completely discredit the contention that coherence conduces to epistemic acceptability, I suggest that it indicates something different: coherence conduces to epistemic acceptability only when the best

explanation of the coherence of a constellation of claims is that they are (at least roughly) correct. That is, the claims that purport to be true are at least roughly true and those that purport to be true enough are true enough.

One might argue that even the best nineteenth-century novel does not pose as great a threat as we sometimes suppose. No matter how deeply immersed I am in the story, a single glance up from the page suffices to convince me that I am not in a drawing room in nineteenth-century England. The content of the story, though internally coherent, manifestly fails to mesh with the rest of my experience. This is so, but the question is what to make of it. On the one hand, too restricted a cluster of mutually supportive claims seems inadequate to engender tenability. We cannot render the story tenable simply by ignoring everything else we believe. On the other hand, insisting that all our commitments need to cohere seems unduly demanding. If acceptability requires coherence with everything we accept, or with everything we accept for cognitive purposes (Lehrer, 1986), it is but a short step to skepticism. One wayward belief, however remote from current concerns, could discredit an entire constellation of beliefs. Epistemological theories that ground justification in coherence then face a problem of scope.

Worries about scope, however, seem not to do justice to the difficulty that confronts us here. Faced with a clash between the deliverances of the novel and those of my glance, it is obvious which I should accept.[1] There is no temptation to resolve the tension by dismissing perceptual deliverances or taking *them* to be fiction. They seem to possess an epistemic privilege that prevents considerations of coherence from overriding them. The capacity of perceptual deliverances to override the claims of a tightly knit novel may seem conclusively to demonstrate that epistemological justification cannot consist in coherence.

Perception's Putative Privilege

Although I do not contend that justification consists in coherence alone, the matter deserves further consideration. Until the source of perception's epistemic privilege is clear, it is premature to rule coherence out. Foundationalists argue that knowledge requires that there be some independently credible beliefs. Many foundationalists hold that perceptual deliverances are among the independently credible beliefs because perceptual deliverances derive at least some of their warrant from the circumstances in which

they occur, not their relation to other deliverances. Exactly how credible they are is a matter of dispute (BonJour, 1985, 26–30). But they must, foundationalists contend, have some measure of credibility that does not derive from their accord with other convictions. Reliabilists argue that a deliverance is epistemically acceptable if produced by a reliable mechanism. Since some perceptual mechanisms are reliable, some perceptual deliverances are acceptable. Since the reliability of perceptual mechanisms is independent of the relations of their deliverances to other deliverances, perceptual deliverances are independently credible.

There are at least two separate insights here. The reliabilist argument targets the need for a link to the world. The reason for crediting the deliverances of a casual glance while dismissing those of the novel is, they maintain, that perception provides the link. The way the world is constrains our perceptual deliverances more immediately and directly than it does our other beliefs. Insofar as the contents of knowledge claims concern the way the world is, it makes sense that the constraints the world supplies should override other considerations. The foundationalist position underscores the idea that some deliverances—in particular, those of perception—seem to display some measure of credibility independently of their connections to other beliefs.

What the objections show is that if perception is to provide the sort of check on theorizing that we think it should, egalitarianism with respect to object-level deliverances will not do. An egalitarian account would hold that each deliverance has an equal claim on our epistemic allegiance. On an epistemic principle akin to 'One Person, One Vote', it would maintain that there is no basis for privileging some commitments over others. If a perceptual deliverance fails to cohere with an otherwise coherent account, the perceptual deliverance ought to be rejected, since the claims of the many outweigh the claims of the one. But no matter how comprehensive and integrated an empirical account is, no matter how many other beliefs the account manages to incorporate, observations should have the capacity to discredit it. They have that capacity only if the epistemic claims of perceptual deliverances at least sometimes outweigh those of an antecedently accepted account. Still, it does not follow that perceptual deliverances must be immune to revision or rejection on the basis of considerations of coherence. Nor does it follow that the epistemic privilege granted to perceptual deliverances is independent of considerations of coherence.

If we think about our situation when we glance away from the novel, we recognize that we draw on more than the sentences comprising the novel and our current perceptual deliverances. We tacitly rely on a fairly extensive and epistemologically informed understanding of novels and of perception. We know enough about underlying mechanisms and track records to have reason to credit some perceptual deliverances. We know enough about literature to realize that novels typically are not literally true. That constitutes sufficient reason for even casual perceptual deliverances to override the claims of the novel.

Juxtaposing the novel with perception might seem to make the problem too easy, though. Regardless of what we think about perception, if we recognize that a novel is a work of fiction, we have reason to discount any direct claims it may seem to make on our epistemic allegiance.[2] The serious challenge comes from a coherent factual account that conflicts with perceptual deliverances. If holism holds that such an account always overrides perceptual deliverances, it seems plainly unacceptable. However tightly woven an empirical account may be, we would be epistemically irresponsible to ignore recalcitrant evidence. Foundationalists take this latter point to be decisive: if observation can show an account to be unjustified, then coherence cannot be the locus of justification.

This would be so if observation worked in isolation. For then, owing to its epistemic privilege, a single perceptual deliverance would have the capacity to discredit an entire system of thought. But that is a myth. Only observations we have reason to trust have the power to unseat theories. It is not an observation in isolation, but an observation backed by reasons, that actually discredits the account.

The holist response to the challenge presented by observation is this: a priori, perceptual deliverances have no special weight. They are just initially tenable commitments jockeying for inclusion in epistemically acceptable accounts. But as we attend to the fates of our various initially tenable commitments, we learn that the incorporation of some, but not others, yields accounts that are borne out by further experience and that retain their place in a coherent system over time. This gives us grounds for discrimination. We realize that the deliverances we take to be perceptual are more likely to be confirmed than spontaneous deliverances that just leap to mind. So we assign greater weight to perceptual deliverances than to hunches and premonitions. Eventually, hunches and premonitions cease to qualify as

initially tenable. Moreover, we learn that not all perceptual deliverances are on a par. Those that are credible tend to come in mutually reinforcing streams, so isolated perceptual deliverances count for relatively little. We begin to draw distinctions among perceptual deliverances. We discover that peripheral vision is less trustworthy than foveal vision. So we have reason to discount what we see out of the corner of the eye. This is not to say that we dismiss the deliverances of peripheral vision out of hand, only that we demand more in the way of corroboration. A passing glance that reveals what looks to be a goldfinch at the bird feeder requires additional support if there is good reason to think that goldfinches have flown south for the winter. Some of us discover that we are color-blind or tone deaf or myopic. That is, we learn that our perceptions as of colors, pitches, or the dimensions of distant objects are not to be trusted. And so on. We come to assign different weights to perceptual deliverances depending on how well they accord with other things we take ourselves to have reason to credit—other appearances of the same object, the reports of other observers, the implications of our best accounts of the perceptible properties of items of the kind in question, and so forth.

The issue is not simply how well a given content meshes with other things we believe, but how well a given content from a given source in given circumstances does. The weight we attach to perceptual deliverances derives from our understanding of the world and our access to it. Perhaps initially this is just a matter of track records. Some perceptual deliverances seem to integrate better into acceptable accounts than spontaneous thoughts that just leap to mind. Later, as we develop physiological and psychological accounts that explain our perceptual mechanisms, we gain additional reasons to credit some perceptual deliverances. The epistemic privilege that some perceptual deliverances enjoy then derives from an understanding of ourselves as perceiving organisms. That is, the reason for assigning those deliverances significant epistemic weight derives from the coherent account of perception that backs the assignment. Contrary to the foundationalist's contention, the justification for privileging perception derives from the relation of perceptual judgments to the rest of our account of ourselves as cognitive agents interacting with a mind-independent world.

The reliabilist account seems to fare slightly better. What justifies assigning visual inputs significant epistemic weight seems to be that vision is a reliable perceptual mechanism. What justifies dismissing forebodings is

that premonition is not. This is not quite right, though. It is not the brute reliability or unreliability of a source that supplies the justification, but a recognition of that reliability or unreliability. Reliabilists disagree. They hold that the sheer fact that a process is reliable renders its deliverances at least prima facie justified (see Goldman, 1986). They would say that if my forebodings were reliable, they would be justified. Were I to believe them, their deliverances would be knowledge. Maybe so. But the knowledge would avail me little. If I had no reason to trust my forebodings, they would not be acceptable. It would be irresponsible for me to use them as a basis for inference or action when my ends are cognitive. From my perspective, they would appear no different from lucky guesses. Even if my forebodings are accurate and their source is reliable, so long as we have no reason to trust them, they rightly bear little epistemic weight.

This argument explains both why some perceptual deliverances have the capacity to unsettle an account, and why those deliverances are not intrinsically privileged. They owe their epistemic status to their place in our evolving understanding of the world and our modes of access to it. This has two welcome consequences. The first is that the privilege they enjoy is revocable. When Bill learns that he is color-blind, he needs to revise his views about which of his visual deliverances are acceptable. The second is that nonperceptual deliverances can in principle be equally weighty. This is an advantage in accounting for the epistemic status of scientific evidence and of testimony.

A look at modern science shows that it is not just (and perhaps not even mainly) bare perceptual deliverances that have the capacity to discredit a scientific account. The outputs of measuring devices do too. In an effort to retain a tie to classical empiricism, some philosophers of science argue that measuring devices are simply extensions of our senses. Just as eyeglasses enable the myopic to see what otherwise they could not, telescopes and microscopes enable everyone to see what otherwise we could not. So if seeing something in suitable circumstances has sufficient weight to undermine a coherent cluster of commitments, seeing something through a telescope or microscope should be able to do so too.

This idea is not unreasonable so long as we restrict ourselves to devices like optical telescopes and microscopes. But it stretches the bounds of plausibility to contend that radio telescopes, electron microscopes, MRIs, and the like are also mere extenders of the sense of sight. It seems preferable to

forgo the strained analogy and simply characterize such devices as detectors. Then an understanding of what they detect, how they detect, and why they should be trusted supplies reason to accord their outputs considerable epistemic weight. Even without the strained analogy, the argument for crediting outputs of scientific instruments thus parallels the argument for crediting perceptual deliverances. For although they are not perceptual mechanisms or even quasi-perceptual mechanisms, the devices are among our modes of access to the world.

Testimony poses a similar problem. We acquire many of our beliefs from the testimony of others, and we consider those beliefs justified. Some philosophers say that the justification for accepting testimony is a priori, albeit defeasible (e.g., Burge, 1993). Ceteris paribus, we are justified in accepting what people tell us. Others say it is inductive. We should believe only those who have shown themselves to be relevantly reliable in the past (E. Fricker, 1994). The former threaten to endorse gullibility, the latter to unduly limit acceptability. Something more sensitive is wanted. Evidently the question is not whether testimony per se is or is not prima facie acceptable. Some testimony is frankly incredible; some requires a good deal of corroboration; some is straightforwardly acceptable. The acceptability of a bit of testimony depends on how well its content coheres with other relevant deliverances, how well the belief that the testifier is competent with respect to her allegation coheres, and how well the belief that she is sincere coheres (Hume, 1977). Because of its mesh with our background beliefs, straightforwardly acceptable testimony scores high on all these measures. Just as different perceptual deliverances are accorded different weights, so are different bits of testimony. Testimony with sufficiently strong backing can discredit a hitherto tenable cluster of beliefs. Even though the deliverances of perception, testimony, and instrumental readings have no special standing a priori, in light of our evolving theories of the world and our modes of access to it, some of them turn out to bear considerable epistemic weight. This satisfies the demand that acceptable beliefs be appropriately constrained by the way the world is. It also reveals that holism has the resources to recognize that initially tenable commitments can differ in weight, some being more credible than others. The claims of the few can in suitable circumstances outweigh the claims of the many.

Adjudication

In adjudicating among competing claims, we prioritize. We do not consider all initial commitments on a par. When they clash, we preserve some at the expense of others, and consider our decisions reasonable. Systematizing is informed by higher-order commitments that assign weights to first-order commitments. Although in principle any commitment can be revised or rejected in the interests of systematicity, commitments differ in cognitive inertia. Some are readily given up; others are abandoned with great reluctance. We readily relinquish the initially tenable commitment that there is a cowbird at the bird feeder in Massachusetts in January. Since it is at best only weakly interwoven with our other commitments, little is lost by abandoning it.

But, as Quine recognizes, even central, tightly woven contentions may justifiably be surrendered. 'Inanimate material objects are identical when and only when their parts are identical' is a highly plausible principle that many metaphysicians consider necessary. If so, it cannot be false. Being a universal claim, it should hold for particles of a viscous fluid. However, if it holds, 'F = ma' does not. The reason is this: The molecules in a viscous fluid move at different rates. In prototypical applications of 'F = ma', forces act on objects like billiard balls that have some sort of material integrity or sharp boundary. But in viscous fluids, the 'forces' on the 'object' are the effects on momentum of molecules moving in and out of that 'object'. So preserving 'F = ma' requires continually redefining what constitutes a particle, letting different molecules comprise it at different times. Although physicists concede that 'F = ma' does not hold at the quantum level or at speeds and distances where relativistic factors play a major role, it is an extremely valuable law for characterizing the behavior of slow-moving middle-sized items in this region of the cosmos. That includes the viscous fluids flowing around here. Fluid mechanics thus characterizes its particles so as to comport with the law. Rather than insisting that all component molecules of a particle be the same from one instant to the next, they let the individual molecules come and go, but keep the average enclosed mass constant (Wilson, 2006, 158–159). 'F = ma' evidently is so central a law of physics that scientists are willing to make drastic revisions in the criteria for the identity of a particle over time in order to preserve it. In this case, the tension is acute. One way or another, a radical revision in antecedently plausible principles is needed.

Either scientists must revise a very reasonable metaphysical commitment about the identity of an object over time, or they must revise a fundamental law of physics.

The issue of scope remains. The totality of a person's epistemic commitments is unlikely to be coherent. Not only are there outliers and inconsistencies among commitments, but an epistemic agent also harbors constellations of commitments that are relatively isolated from one another. Meg's cluster of beliefs about the pituitary gland, the evidence that bears on the acceptability of these beliefs, the trustworthiness of testimony on the subject, and the proper methods for assessing such matters has few and loose connections to her cluster of views about parliamentary procedure, the evidence that bears on these views, the trustworthiness of testimony about the subject, and the proper methods for assessing them. It seems that she could easily be badly wrong about the former without her error having any significant effect on the tenability of her views about the latter. Outliers and inconsistencies among beliefs are in principle relatively unproblematic. According to holists, outliers lack justification. Because they lack suitable connections to other epistemic commitments, although they are initially tenable, the epistemic agent has little reason to credit them. Inconsistencies among commitments conclusively demonstrate that some of them are false. But it is not obvious that mutual indifference of clusters of commitments is objectionable. It would be unreasonable to consider Meg epistemically defective because of the lack of close ties between the two clusters. On the other hand, if the clusters of commitments are too small and too numerous, complacency over their mutual indifference seems problematic. We do not want to license ignoring inconvenient tensions among commitments by consigning them to mutually irrelevant clusters.

The problem neither has nor needs an a priori resolution. Our evolving theories of the world and our access to it provide us with an appreciation of the relations in which our various clusters of commitments should stand to one another and the requirements they should satisfy. Such a laissez-faire attitude might seem to allow for the acceptability of crazy constellations of views. If we leave it to our evolving worldviews to decide what range of considerations acceptable accounts must answer to, we may be forced to endorse isolated islands of claptrap. The worry is more apparent than real. We have metalevel accounts that enable us to assess the reasons, methods, standards, and evidence that our various object-level accounts appeal

to. Some requirements, such as logical consistency, apply globally. Regardless of the distance between them, unless they can be conjoined without contradiction, Meg's political and endocrinological commitments are not all acceptable. Other requirements, like the need to respect judicial precedents or to accord with biochemical findings, are more limited in range. But even these do not enable us to completely isolate clusters of commitments. Even if Meg's views about endocrinology and politics have few points of contact, her views about endocrinology and hematology have many.

Consistency requirements do more than rule out express contradictions. The general requirement that like cases be treated alike demands that if a consideration has weight in one area but not in another, there must be an acceptable reason for the difference. To be tenable, a system of mutually reinforcing claims must either answer to the logical and evidential standards to which other theories are subject or be backed by a tenable account of why those standards do not apply. Some theories have such backing. There are, for example, cogent reasons why mathematics is not subject to empirical testing. So infinitary mathematics is not threatened by the absence of empirical evidence for its findings. In epistemically objectionable cases, no such reasons are available. The claims of 'creation science', although purportedly mutually reinforcing, are epistemically unacceptable because they are indefeasible. According to their adherents, nothing could count as evidence against them, for whatever we find is the way God created things. To say that something cannot be ruled out a priori is not to say that it cannot be definitely and decisively ruled out.

A familiar objection to coherentism is that coherence is so easily achieved that it affords no reason to believe that the contentions in a coherent account are true, hence no justification for them. This assumes that any coherent collection of claims is as good as any other and that whenever incoherence threatens, we can restore coherence by throwing out whatever poses the threat. Whatever its merits as an objection to pure coherentism, it does not apply to positions that ground acceptability in reflective equilibrium. First, not all initially tenable commitments are object-level statements pertaining to the topic at hand. Among those we seek to accommodate are ostensible epistemic values, such as simplicity and relevance; ostensible epistemic standards, such as standards of statistical significance and of evidential support; ostensible epistemic criteria, such as those specifying what counts as validity and as argument; and ostensibly acceptable methods,

such as controlled experiments and random sampling. It is by no means easy to bring all of these into accord. Second, the objection overlooks the special role that initially tenable commitments play. Because they are considerations that present themselves as creditable, they serve as touchstones for theorizing. They supply epistemic ballast, preventing us from restoring coherence by simply rejecting considerations cavalierly.

Not all initially tenable commitments, of course, present themselves as candidates for belief. Putatively felicitous falsehoods present themselves as being presumptively true enough. Putatively acceptable values, methods, standards, and criteria are elements of the constellations within which doxastic commitments function. So the slight presumption in favor of doxastic commitments extends to axiological and methodological commitments and felicitous falsehoods as well. They, too, are initially tenable.

We devise a flexible network of cognitive commitments that, through continual readjustments, achieves an understanding of the topic that is on balance reasonable. None of the commitments is absolutely unrevisable. Different potential revisions have different costs and benefits. To decide among potential revisions requires asking what sort of understanding we seek, what resources we have to draw on, and what limitations we currently face. We have multiple cognitive desiderata—simplicity, fecundity, elegance, predictive power, and so on. Insofar as is feasible, revisions in our initially tenable commitments should yield an account that satisfies them. Moreover, with the advancement of inquiry, methods, standards, values, and goals evolve. There emerge new understandings of how to find or figure things out, what sorts of methods and results stand the test of time, what sorts of representations are apt, what desiderata are worth pursuing and when and how they might conflict. When the account that emerges is grounded in considerations that are collectively as good as any available alternatives, it is an account we can on reflection accept.

Some might dispute this. If our initial commitments were quite wide of the mark, the account that results might still be far from the truth. This is so. Nevertheless, such an account is an epistemic improvement over our starting point and is as good as anything we have reason to accept in those dismal epistemic circumstances. It affords a relatively stable platform for research that may yield yet further improvements.

To see this, consider a skeptical scenario. Imagine that unbeknownst to themselves, a group of scientists live in a Berkeleyan universe. What

they take to be material objects are really immaterial ideas in the mind of God. Initially, they have fairly crude, unsystematic opinions about how things behave and about how to find out how they behave. Over time, they correct, extend, and systematize their approach to the point where they develop the scientific method. They subject their findings to peer review. They insist on controlled, repeatable experiments whose results are statistically significant. The geniuses among them come up with laws that reveal an astonishing order and regularity among the phenomena they observe. The order those laws describe obtains. The 'orbits' of the 'planets', the 'tides', and 'falling bodies' fall under the same equations. Their account is wrong, however, in that it takes the substrate for that order to be material rather than mental.

There is no denying that this is a big mistake. But, I would urge, we should also acknowledge their significant epistemic achievement. Otherwise we would have to construe the success of their science as a fluke. They can, like all scientists, make further progress by extending, correcting, and deepening their account, improving their methods, fine-tuning their criteria, and so on. Maybe someday they will be in a position to recognize that matter, as they conceive it, does not exist; maybe not. But it seems churlish to discount their achievement because it yields insight only into the structure of reality, not into its metaphysical ground. Rather, I contend, we should recognize that because their opinions, methods, criteria, and standards are in reflective equilibrium, they are reasonable in the epistemic circumstances. The system they have constructed is as good epistemologically as could have been achieved given their unfortunate starting points.

Some might concede that the method of reflective equilibrium is the way we go about developing our accounts. It is, after all, a schematic idealization of the self-correcting scientific method. But, they would insist, it is only a method. What makes an account a good account is not that it was generated by the method, but that it satisfies some independently specified standard. Typically the standard is truth. And what makes the method a good method is that it tends to produce theories that satisfy the external standard. The justification for the method of reflective equilibrium, they would maintain, is that it is truth conducive.[3]

There are at least two replies. One is that we have no assurance that the accounts we accept satisfy the standard. If the standard is really independent of the method, we might legitimately wonder whether the accounts

we credit are true. Going back and garnering more evidence would plunge us back into the method of reflective equilibrium, whose bona fides are in question. This is a version of the problem of the criterion (Chisholm, 1973). Moreover, if the standard is truth, the account devised in the Berkeleyan world is devoid of epistemic merit. Perhaps we can concede that the scientists did the best they could, but in the circumstances, there was no way to come up with an epistemically estimable account. From a purely epistemological point of view, they are no better than their compatriots who cavalierly jump to conclusions without even trying to amass or assess evidence.

This is not just a point about some fictional scientists in a made-up world. We have no conclusive evidence that we are not in their situation. That being so, the acceptability of accounts that satisfy the standards we actually use is doubtful. Unless we can explain why being in reflective equilibrium is truth conducive, we seem hostages to epistemic fortune. From our perspective, whether we live in a Berkeleyan world is a matter of epistemic luck. If we can explain it, we can treat being true or truth conducive as an objective that we attempt to accommodate in devising theories in reflective equilibrium. And, if we like, we can assign truth or truth-conduciveness a high priority so that other desiderata would be given up to achieve it. I will argue below that science tends to be much more flexible in its attitude toward truth, but in principle we could take a hard line.

It might seem that the requirement on epistemic acceptability it is too lenient. 'As good as any available alternative' is not 'better than any available alternative', much less 'better than any alternative, available or not'. Why settle for so little? The method I sketched plainly leaves room for ties. When commitments clash, different trade-offs may be equally good on balance. One might sacrifice a measure of sensitivity for an increase in scope; the other might sacrifice scope for sensitivity. It seems dogmatic to insist that one must be better than the other when, by our standards, both do an equally good job and none does a better job at realizing our cognitive desiderata.

Some philosophers think that such ties just reveal our epistemic inadequacy. There is one way the world is, so there is one and only one correct account of the world. If at a certain point in our investigation, we have two or more equally good accounts (and none better), that just shows that our methods, standards, or criteria are too weak. We need to introduce tie breakers (D. Lewis, 1999a). But even if there is just one way the world is,[4]

the contention that there is exactly one best account of the world is unwar-ranted. What makes for a good account depends not just on the subject matter of the account but also on our goals and resources in theorizing. Even if the world's contribution is fixed, different accommodations to our inevitably limited resources may be equally reasonable. No argument is given that would privilege a preference for a more sensitive account over one with greater scope, or conversely. Sensitivity and scope are both desiderata. We can explain why each of them is desirable. But the conviction that there is a unique optimal balance between them is unwarranted. Requiring that we refine our standards until we achieve uniqueness is idle. Without an epistemically accessible indication of what justifies one refinement over another, any decision would be arbitrary.

To insist that an account is unacceptable unless it is better than any alternative, whether available or not, is perilously close to an argument for skepticism. To suspend judgment until we have reason to believe that an account we have devised is better than any imaginable or unimaginable alternative is to suspend judgment permanently; to accept an account on the assumption that it is better than any imaginable or unimaginable alter-native is intellectually arrogant. It intimates that the account satisfies a standard it cannot plausibly be known or reasonably believed to satisfy. To insist that uniqueness is required is to consign our accounts to inadequacy, whatever merits they display. Or it is to contend disingenuously that the accounts we consider acceptable satisfy standards we have no evidence that they satisfy, indeed standards we cannot even articulate. The best policy, then, is to take the standard of acceptability to be 'as good as any available alternative' and to forgo a demand for uniqueness.

Earlier I said that truth has a more peripheral role in understanding than epistemologists standardly think. We are now in a position to see what that role is. A tenable account affords an understanding of its topic. An account is tenable only if it is in reflective equilibrium. This may require that claims in the account that purport to be true actually be true, or anyway not off by much. Although science endorses felicitously false models, it does not consider false observation reports to be felicitous. Fidelity to the evidence seems mandatory. Observation reports are not, of course, the only asser-tions in scientific accounts. Assertions—about laws, boundary conditions, methods, and so forth—are scattered throughout. It might seem, then, that there is a streamlined answer to our question: tenability requires that the

assertions in an account be true and the professions be felicitously false. For assertions purport to be true, and professions purport to be felicitous. Only then, we might think, would an account embody an understanding of the phenomena it concerns.

This does not quite work, for understanding grows over time. One of the ways it grows is by overthrowing previously accepted accounts. Sellars (1968) maintains that when one theory supplants another, it should have the resources to explain why its predecessor did as well as it did. Often the reason the predecessor was so successful is that some of its central assertions, which were put forth as true, turned out to be felicitously false. Kepler's laws, presented as true, neglected the gravitational force on the Earth of everything except the sun. As Newton recognized, Kepler's laws are false. For the gravitational forces exerted by the other planets affect the Earth's orbit. But because the forces exerted by the planets are relatively insignificant compared with that of the sun, Kepler's laws are not off by much. So Newton's theory satisfies Sellars's demand by reinterpreting statements that purport to be true as felicitously false. That a sentence is asserted shows only that it purports to be true. It does not show that its truth is what it contributes to the tenability of account it belongs to. Truth evidently is epistemically important only to the extent that it contributes to tenability.

My position does not privilege any sorts of beliefs or representations a priori. Nothing favors the verbal, the literal, the descriptive, or the true. Acceptability derives from reflective equilibrium. So whether a representation (true or false, verbal or pictorial, literal or figurative, denotative or exemplificational) is acceptable turns on whether it is an element of an account in reflective equilibrium. What commitments are worthy of acceptance is something we learn by developing increasingly comprehensive, coherent accounts of the world and our access to it. The method enables us to start from whatever initially tenable commitments we happen to have. But because it insists that we subject them to rigorous assessment, such a starting point is not question begging. The standards of assessment are themselves the fruits of epistemic activity and can change in response to feedback (Goodman, 1984, 69). Hence, nothing is in principle immune to revision. An account that we can on reflection accept today may be one that we cannot on reflection accept tomorrow. But so long as an account is in reflective equilibrium, it and its components are justified. What results is neither certainty nor skepticism but a fallible, provisional, reasonable

epistemic stance. We judge our commitments and potential revisions of our commitments against one another to see which combination seems best on balance. But we also judge our criteria for being best on balance by seeing if they yield verdicts that we can on reflection endorse. No account or component of an account is irrevocable, but an account in reflective equilibrium is acceptable until problems emerge or improvements are envisioned. The justification for this verdict lies in the fact that the commitments in reflective equilibrium are as reasonable as any alternative in the epistemic circumstances.

5   Epistemic Normativity

The repudiation of veritism opens an important gap.[1] If truth-conduciveness is not the standard by which epistemic policies and practices are to be judged, what is? I suggest that rather than epistemic practice having an exogenous aim, its aims are delineated by and within the practice itself. Epistemic standards are vindicated not by being reliable promoters or indicators of truth or any other external end, but by being products and promoters of responsible epistemic agency. Epistemic norms, then, are norms of responsible epistemic agency. In this chapter I argue that epistemic norms are norms that would emerge from the deliberations of suitably idealized epistemic agents. Drawing on Kant, I urge that such agents are to be construed as legislating members of a realm of epistemic ends. This is an idealized model of real-life communities of inquiry. The idealization eliminates factors that distort deliberations. I argue that legislators must be, in a political sense, free and equal. They must be free to advance any hypothesis they consider viable and equally entitled to be heard. This does not mean construing them as equally smart or construing their views as equally worthy of acceptance. It means only that they are entitled to be heard. The result is a form of epistemic responsibilism. It explains why a system in reflective equilibrium is worthy of acceptance even though there is no guarantee that it is true.

Reflective Endorsement

If epistemic normativity is keyed to epistemic responsibility, what are the criteria for epistemic responsibility? One familiar requirement on responsibility is 'ought' implies 'can': x is responsible for y only if whether y obtains is under x's (direct or indirect) control. If x cannot bring about y, or

x cannot prevent *y*, then *x* is not responsible for the occurrence or nonoc-currence of *y*.[2] Epistemic agents have little if any control over their epistemic environment. They cannot, for example, control whether they live in a Berkeleyan world. So they are not epistemically remiss if the falsity of their beliefs derives from an undetectable mismatch between the evidence and its metaphysical ground. Nor are agents epistemically remiss if, in a more auspicious world, they form a false belief on the basis of evidence that is scrupulously garnered and carefully analyzed but misleading. Truth-conduciveness, then, is not the criterion of epistemic responsibility. What is the alternative? Even in an inauspicious world, agents can exercise epistemic self-discipline (see Weatherson, 2008). They can decide whether on reflection to endorse considerations that present themselves as candidates for acceptance.

According to Kant (1981), someone who acts autonomously makes the laws that bind her, while someone who behaves heteronomously is bound by constraints he neither makes nor endorses. He is not strictly an agent, for he does not act; he merely reacts. He responds to stimuli. Although Kant describes the heteronomous subject as driven by his inclinations or desires, action is a joint product of belief and desire. So a subject is equally heteronomous if he is driven by beliefs that are not under his control. This suggests that it may be fruitful to import Kant's distinction into the epistemic realm.

Inasmuch as we are cognitive beings with sensory and representational capacities, a variety of things simply strike us as being the case. Deliverances of sensation, emotion, apparent memory, and imagination present themselves as candidates for acceptance. It might seem that this makes us, at base, epistemically heteronomous. We can't help but see, hear, and feel what we do. But the issue is not whether we are passive in the reception of inputs; it is what we do with those inputs. The heteronomous individual simply takes deliverances at face value. For him, seeing is believing. Or perhaps seeing is not always believing, but whether he believes what he sees is not up to him. He simply finds himself believing something on the basis of his deliverances, being under the sway of whatever belief forming mechanisms happen to be operative. Since his beliefs are considerations that just strike him—since he can neither defend nor criticize them—they are not under his control.

The beliefs he finds himself with might be reliable. Perhaps humans evolved in such a way that what strikes us as dangerous usually is dangerous.

If so, when the subject thinks an animal is dangerous, it usually is; when he thinks an animal is harmless, it usually is. Maybe this generalizes. To the extent that his dispositions are reliable, he is objectively secure. Nevertheless, he is in a subjectively precarious position. Reliabilism is a form of veritism that holds that a disposition or process is epistemically creditable just in case it is reliably truth conducive and a true belief is knowledge just in case it is a nonfortuitous product of exercising a reliable disposition or of appropriately using a reliable process. Whether the reliabilist requirement is satisfied is independent of whether the subject thinks it is and independent of her reasons (if any) for thinking it is. Evidentialist veritism holds that a subject's true belief qualifies as knowledge just in case it is backed by adequate evidence, where the adequacy condition covers not just the amount of evidence but also its being nonfortuitously linked to the belief's truth-maker. Again it makes no difference whether the agent thinks that the requirement is satisfied. As far as veritism is concerned, heteronomous, justified, nonfortuitously true beliefs, and the processes and dispositions that underwrite them, are epistemically creditable. Still, insofar as the subject's reflective endorsement or repudiation of his beliefs and the processes that gave rise to them have no bearing on their epistemic status, he is a victim of circumstances.

Perhaps his situation is not so dire as I make it out to be. The believer can register his successes and failures, and credit the dispositions and mechanisms that engender success. He might notice that deliverances of foveal vision tend to be borne out by subsequent experience more frequently than deliverances of peripheral vision. So he assigns foveal deliverances greater credence. But the question then arises for the second-order beliefs, processes, and dispositions he credits. Do they just strike him as correct? Can he defend the way he registers, identifies, records, and tallies successes and failures? Even if his methods are in fact reliable, so long as he has nothing more to say than that it strikes him that this is a good way to proceed—to identify, tally, or measure—he remains vulnerable. Second-order heteronomy is as problematic as first-order heteronomy.[3] Moving to yet higher orders is of no help. Regardless of the level, so long as the best that can be said for the principles or processes he ultimately relies on is that they strike him as correct, he is vulnerable. He may be on safe ground, but he has no reason think he is. He is driven to believe by considerations he happens to be struck by. And he is not responsible for what strikes him.

Is the autonomous agent any better off? She too may be struck by things. Her inputs may be exactly the same as those of the heteronomous subject.[4] But her relation to them is different. Forming opinions is something she does, not something that happens to her. What she does in forming opinions is, as it were, filter deliverances (such as putatively perceptual inputs, apparent memories, or simply ideas that cross her mind) through a critical sieve, accepting only those she considers worthy of her reflective endorsement. What blocks the regress to ever higher-order heteronomy is that second-order endorsement is the product of agency. The agent decides to accept, reject, or withhold. Epistemic subjects are epistemic agents; they take their beliefs, practices, and policies to be answerable to certain norms because they think that epistemically acceptable beliefs, practices, and policies ought to be answerable to those norms.

To be sure, an epistemic agent does not entertain deliverances one by one and ask herself, 'Am I buying this?' Rather, she develops and deploys a variety of methods, mechanisms, heuristics, and habits that enable her to credit or discredit wide swaths of inputs efficiently. Apparent sightings of not unexpected middle-sized objects in good light in the center of her visual field are apt to be readily accepted. Given her prior experiences and other epistemic commitments, having no reason to reject them may constitute reason enough to accept them. But if there is something incongruous about a deliverance—if, for example, she seems to see a camel on Main Street in small town America—she does not simply accept it. She may reject it as a hallucination, an illusion, or a misperception; or she may suspend judgment until she investigates whether a camel is part of a local publicity stunt, political rally, or circus parade. Similarly if she is aware of something untoward about her condition as an observer: if she is half awake or on new medication or intently focusing on something else, she may refuse to credit a surprising deliverance. The critical point is that what she accepts is a product of her reflective endorsement. She is willing to stand behind it because it satisfies her standards. Even if her acceptances are largely automatic, she could, if she chose, withhold acceptance of the content of a deliverance, or of a source of deliverances. What makes the epistemic agent responsible for her opinions is that she takes responsibility for them.

Voluntarism

Can she really choose? Doxastic voluntarists, such as William James (1951), maintain that belief is at least sometimes subject to the will. According to James, if someone thinks his life would go better if he believed that a benevolent God watches over him, it is both possible and epistemically permissible for him to do so. Pretty much everyone acknowledges that the will can indirectly influence belief. An agent can garner evidence, put herself in a position to be influenced, even arrange to be hypnotized in hopes that the belief she desires will result. Sometimes such strategies are effective. But voluntarists like James claim that the influence of the will is direct: A person can believe something simply by deciding to do so. James restricts his voluntarism to realms, like the religious, where empirical evidence does not have much purchase. Such a restriction is imperative if the position is to be at all plausible. It would hard to maintain that a driver could simply decide to believe that she was cruising down a country lane in the south of France when all the evidence points to her being caught in a traffic jam on the Cross Bronx Expressway. But, James maintains, where there is no evidence either way, a person can believe that God exists or believe that God does not exist, as she desires.

Williams disagrees. He argues that because belief is world directed, and belief formation is responsive to evidence, the state that would emerge from such a resolution would not be belief.

> If I could acquire a belief at will, I could acquire it whether it was true or not; moreover I would know that I could acquire it whether it was true or not. If in full consciousness I could will to acquire a 'belief' irrespective of its truth, it is unclear that before the event I could seriously think of it as a belief, i.e., as something purporting to represent reality. At the very least, there must be a restriction on what is the case after the event; since I could not then, in full consciousness, regard this as a belief of mine, i.e., as something I take to be true, and also know that I acquired it at will. (1978, 148; see also Adler, 2002)

In order to be responsive to the world, Williams maintains, belief must be resistant to the will.

The issue is trickier than Williams recognizes. A standard objection to doxastic voluntarism is that if belief (or belief formation) were voluntary, an agent could form any belief she likes, just by deciding to do so. The manifest impossibility of believing whatever one likes regardless of the evidence

seems to settle the case. But voluntary actions are not completely unfettered. Walking is voluntary even though no one can at will walk from San Francisco to Honolulu. Voluntary actions are subject to constraints.

Some are what we might call enterprise-specific constraints. An enterprise both enables and limits the range of a participant's voluntary actions. Only in chess is a knight fork possible; and how a player can create a knight fork is constrained by the rules of the game. Nevertheless, moving one's knight so as to create a fork is a voluntary action. The fact that a player cannot perform that action whenever she pleases—while playing football, for example, or while baking brownies—does not undermine the status of the move as a voluntary action. Nor does the fact that a chess player cannot create a fork just by wanting to. The pieces on the board must be distributed in such a way that a player can move her knight into a position that puts two of her opponent's pieces in jeopardy. This is not always possible. But when it is possible, her decision to move her knight to the relevant position is surely a voluntary act.

How does this bear on belief? As we have seen, to take belief formation to be subject to the unfettered will is a mistake. We cannot make ourselves believe that it is snowing (Feldman, 2008), or that we're driving through France, or that a benevolent God exists (Pascal, 2007) simply by deciding to do so. The question is whether the doxastic realm is, like a chess game, a realm in which voluntary actions are subject to enterprise-specific constraints (see Shah, 2002). I suggest that it is. Beliefs are supposed to be responsive to evidence. Still, agents can and often do cut corners, jump to conclusions, reason carelessly, and think sloppily. They are subject to criticism for their errors and omissions, for they could do better if they tried. Or so, evidently, we think. There is apparently a range of freedom available to agents in their evidence gathering and assessing. The availability of that freedom to maneuver affords a basis for praise and blame, because how scrupulous epistemic agents are is under their voluntary control. The freedom to maneuver is delimited by the epistemic enterprise. To accept that p is to be willing to use p as a basis of inference or action *when one's ends are cognitive*. This is not the same as to be willing to use p whenever one likes. The agent is not free to import into the epistemic realm factors, such as wishes or desires, that reside elsewhere. So the problem with James's 'will to believe' is not, as Williams thinks, that belief is wholly involuntary. It is

that James contends that epistemic agents can draw on factors outside the epistemic realm in deciding what to believe.

Nevertheless, a case sketched by Firth (1981) and elaborated by Berker (2013) might seem to raise a serious difficulty for my position. Suppose Jane is a talented chemist with a promising hypothesis. She needs funding to test it and the only available source of funds is a foundation that only supports research by theists. Jane, unfortunately, is an atheist and a terrible actor. She knows that she cannot persuasively fake religious belief. So she has herself hypnotized into believing that Odin exists. (The foundation insists on belief in a god but is ecumenical about which god it is.) She gets the grant, does the research, and makes contributions that advance the understanding of her subject considerably. Despite the good that came of it, one wants to say that her belief in Odin's existence is epistemically unacceptable. It was illicitly imported from an extra-epistemic realm.

It might seem, however, that I can say no such thing. Jane accepted the contention that Odin exists as a basis for action (the action required to get the grant) when her ends were cognitive (doing the research). So 'Odin exists' looks like a felicitous falsehood. If I am committed to this, my position is in trouble. Luckily, I am not. Jane's false and unwarranted belief is instrumentally relevant to her contributions to science. An epistemic consequentialist should recognize that it is truth conducive, hence presumptively acceptable. But the belief does not figure in Jane's contributions to science. Talk of Odin appears nowhere in her scholarly work. Nor is it presupposed by that work. Once she has obtained the funds, her belief in Odin's existence plays no further role. It is not woven into her fabric of epistemic commitments. Felicitous falsehoods, on my theory, are integral to tenable accounts. They are not merely causal antecedents. My position is nonconsequentialist.

An agent's epistemic behavior is, in a variety of obvious ways, subject to her will. She can decide when to stop gathering evidence, when to stop checking inferences, when to stop investigating other perspectives, and so on. Epistemic acceptance thus is properly subject to 'ought implies can'.

Still, it might seem that to some extent belief is heteronomous. Given her evidence for p, an agent cannot help but believe it. Here the distinction between conviction and acceptance pays dividends. Perhaps she cannot help but feel that p is so; p, then, is her conviction. But acceptance is a

product of autonomous agency. An agent decides to accept *p* as a basis for cognitively serious inference and action when she reflectively endorses that *p*. In acceptance, the will has a role to play.

Williams is right that the world impresses itself on epistemic agents. We do not get to simply wish its deliverances away. Nor can we cavalierly refuse to countenance deliverances because we want to. The issue is what we can endorse *upon reflection*—that is, in light of our other epistemic commitments. We can responsibly reject or suspend judgment about a deliverance if it seems sufficiently implausible. This is what Nozick (1981) calls the Optional Stop Rule. It is always open to epistemic agents to refuse to countenance a conclusion that they consider sufficiently implausible. They can deem that the argument for it contains an undetected flaw. In everyday epistemic practice, we thus have a mechanism for responsibly withholding reflective endorsement. But consigning a deliverance to the realm of misperceptions or hallucinations, or an argument to the realm of those with undetected flaws, has costs. Consigners then face the job of explaining what gave rise to the misperception or hallucination, or where the flaw is located. To bring their views about their recent experience into reflective equilibrium, they need to account for the recalcitrant deliverance.

The role of the will in reflective endorsement is not at the level of asking whether the agent wants to accept a particular contention. An agent cannot reflectively endorse a contention simply because she wants things to be as it says they are. Nor can she reflectively endorse it because, despite its manifest untenability, it would be, as it was in Jane's case, epistemically fruitful to do so. Rather, the agent, along with other members of her epistemic community, sets standards of evidence and thresholds for acceptance. They determine how much evidence is required for a contention of a given sort to be worthy of reflective endorsement. The agent is not at liberty to decide as she likes whether she sees a camel on Main Street. But she and the other members of the community have a measure of liberty in deciding how much evidence she needs for the contention that she sees the camel on Main Street to be acceptable. That liberty is constrained by requirements of consistency and coherence with other matters they reflectively endorse.

Epistemic endorsement, then, is a product of agency. Acceptance is under an epistemic agent's voluntary control. But it does not follow that a responsible epistemic agent can accept whatever she pleases. Reflective equilibrium requires that her epistemic commitments be mutually supportive and

that they constitute an account that is at least as reasonable as any available alternative in the epistemic circumstances.

Reasons

Reflection does not occur in a vacuum. It involves sensitivity to epistemic ends and means, capacities and limitations. It is imbued with (often tacit) background assumptions and is responsive to epistemic circumstances. A sighting of what looks to be camel on Main Street typically warrants further investigation; a sighting of what looks to be a mid-priced American car on Main Street ordinarily does not. An argument leading to an astonishing conclusion often provokes further inquiry; an argument leading to an expected one normally does not. An agent's reflective endorsement is a willingness to be bound by an epistemic commitment (an opinion, method, or standard) because, given her background epistemic commitments, she thinks that her epistemic purposes will be served by her being so bound. She considers the items she reflectively endorses trustworthy bases for inference and action when her ends are cognitive.

It might seem that this too leaves the agent vulnerable. She is willing to stand behind the commitments she reflectively endorses. But should she be? The reliable but heteronomous subject is vulnerable because he cannot reflectively endorse his beliefs. It might seem that the autonomous agent, as I have characterized her, has the opposite problem. She can reflectively endorse her opinions, but they might be unreliable. She is perhaps subjectively secure, but she remains objectively at risk. In fact, if her personal standards are sufficiently skewed, her reflective endorsement may be epistemically worthless. Can we evade an untenable subjectivism here?

When an agent's commitments are ones she reflectively endorses, they are at least prima facie defensible. She can adduce what she takes to be reasons for them. But it is not enough that the agent has something she considers a reason. If her claim that chicken soup alleviates colds were challenged, she could not prevail by saying that she read graffiti to that effect. Reasons, as Korsgaard (1996b) argues, are considerations we give to each other. They are not mere expressions of personal conviction, but considerations our interlocutors should countenance. The agent could give as her reason that she had amassed and evaluated the relevant evidence, that she read about the correlation in the *New England Journal of Medicine* or the

New York Times, that she had been told by her physician that this particular old wives' tale is true. She might even say that her grandmother told her. Assuming her grandmother is not an expert on contagious diseases, this would be a relatively weak reason. But if members of her epistemic community were inclined to give the grandmotherly voice of experience some epistemic weight, it would still be a reason.

Among the considerations that are within the agent's epistemic purview are those she can call to mind. She has learned them and has not forgotten them. These are her *available reasons*. If she recognizes their relevance, she can surely appeal to them to defend her claim. But her purview is wider than that: it includes information she can readily draw on—information obtainable in her epistemic milieu, whether she is currently privy to it or not. These are her *accessible reasons*. In some contexts, we take available reasons to set the limits of accountability. Phil has an available reason to believe that Jane is old enough to vote, because he knows that Jane is older than Jim and he knows that Jim can vote. In others, we appeal to accessible reasons. Meg has an accessible reason to believe that the course has a prerequisite, since that is clearly stated in the catalog, and she has access to the catalog. Accessibility and availability are matters of degree. Some reasons are readily accessible; others take considerable digging. Some available reasons are easily called to mind; others need to be dredged up. The resources available to an agent's epistemic community figure in her range of accessible reasons. She can in principle gain access to whatever information or insight is publicly available to the members of her community. She can draw on her intellectual compatriots and, as a member of that community, she has reason to endorse them.

Reasons are considerations that can properly be adduced to support a conclusion. Drawing on Scanlon (1998), I suggest that a consideration c can properly be adduced to support p only if others who are competent to assess the relation between p and c and motivated to know whether p is acceptable could not responsibly reject the claim that c supports p. Reasons thus are public. No purely subjective consideration counts as a reason; for others who are similarly motivated to find out whether p is acceptable could responsibly reject private intimations. However heartfelt, Pat's feeling that the Red Sox will win is not a reason to think that they will win. It is merely a hunch. Moreover, since nothing epistemically inaccessible qualifies as a reason, a reliable method for establishing that p is not a reason to accept or

believe that *p* until investigators establish that the method is, or is likely to be, reliable. Currently, then, there is no reason to accept the deliverances of Marie's ESP, even if it is reliable. And a motley collection of evidence does not supply a reason to accept that *p*, unless its connection to the conclusion is accessible to the agent.

No truth that fails to satisfy accessible standards of relevance counts as a reason. Considerations that enhance the objective probability that *p* are not reasons if, according to current or foreseeable standards of relevance, they have no bearing on whether *p*. Even if left-handed people are more likely to develop arthritis than right-handed people, so long as we are unaware of the correlation and are not remiss in being unaware, Ben's left-handedness affords us no reason to think that he is more at risk for arthritis than his right-handed brother. The correlation is epistemically inert.

Whether a consideration qualifies as a reason depends on the epistemic circumstances. Public standards of evidence and relevance rest on background assumptions about the topic under investigation and what is known or reasonably believed about it. Considerations that at one point in history could not responsibly be rejected may be readily rejected later, when matters are better understood. In 1975, it was widely accepted that stress causes peptic ulcers. A gastroenterologist could not at that time responsibly reject the contention that the high incidence of ulcers among Wall Street traders was due to their stressful jobs. Once the bacterial basis for ulcers was discovered, stress can no longer be adduced as a reason why so many traders have ulcers. A factor that functions as a reason in one cognitive environment may fail to so function in another.

Reasons are keyed to methodology. Among the factors that figure in whether others can responsibly reject the claim that *c* supports *p* are views about methods for establishing whether *p*. Are standardized aptitude tests a good predictor of college success, or can one responsibly reject the contention that a student's acing the SATs indicates that he will do well? Are focus groups a good way of gauging public opinion, or can one responsibly reject the contention that because a focus group found the candidate's message compelling, the public at large will do so too? Answering such questions requires validating the methods that connect *c* and *p*, and insuring that appropriate standards of rigor are satisfied. A major question is: whose opinions matter? To be exceedingly tolerant, to allow just about everyone to have a say, is to court skepticism. For any interesting thesis, there

are bound to be people whose background commitments are such that they could, without violating their own epistemic principles, reject a contention that *c* is a reason to hold that *p*. A blind person might reject the contention that the car's looking blue in daylight is reason to think that it is blue. He has no direct evidence of that. A statistical novice looking at graduate school admissions might conclude that the university discriminates against women. (Cartwright, 1983, 36–38). Being unaware of Simpson's paradox, he does not understand why, if departments do their own admissions, looking department by department rather than surveying the university as a whole yields the appropriate statistics. A religious fundamentalist might reject the contention that the fossil record affords reason to believe that birds evolved from dinosaurs. He denies that anything could afford genuine reason to believe that evolution occurred. To disqualify such reactions requires a principled way to delineate the class of others whose reactions matter.

More is needed than that the standards be shared. All three rejecters might be deploying standards they share with others. The issue is whether those standards are cognitively acceptable standards, standards whose satisfaction fosters the advancement of understanding. To resolve this, we need to look at the standards and the goods they promote.

In all three cases, we dismiss the rejecters' opinions because we know better. This is not intellectual arrogance. We know what relevant resources are accessible and how and why to draw on them. We dismiss the statistical novice's opinion because it rests on a misunderstanding of statistical methods. This indicates that acceptable reasons must be consonant with the proper use of acceptable methods, where the acceptability of the methods turns on their having been validated and endorsed by experts on the subject and the best available ways to justify conclusions about it. But it is not enough if by chance statistics are properly used. The proper use of the methods should be based on an understanding of the methods, their ranges of application, their powers and limitations. This sounds circular, but it is not. Rather, it is a manifestation of holism. For reflective equilibrium, methods, standards, and results must mesh. When one strand of the network of commitments is doubtful, the rest of the network is properly called on for support. We dismiss the blind man's opinion because it is based on a dearth of direct evidence and a failure to make appropriate use of accessible indirect evidence. This indicates that epistemic agents are expected to

be aware of their own limitations and of the resources they can draw on to compensate for them. We dismiss the fundamentalist's stance because it is dogmatic. It is not sensitive to evidence, since he would retain his position whatever the evidence. This indicates that for something to count as a reason for an empirical claim, it should be responsive to evidence.

The fundamentalist might reply that his opinion is responsive to evidence. Scripture, he maintains, provides incontrovertible evidence that evolution did not occur. He disregards the fossil record because nothing it shows could override scripture. His belief then rests on shaky grounds. He relies on a single source, so his belief is unstable. Should that source be discredited, he has nothing to shore up his belief. He relies on a single, disputed interpretation of the deliverances of that source, so it is fragile.

The scientist is in an epistemically stronger position. Because her opinion is multiply tethered, it is solidly grounded. We would have to be massively wrong about genetics, anatomy, and physiology to discredit the direct biological evidence for evolution; wrong about physics to discredit the evidence provided by carbon dating; and wrong about geology to discredit the evidence of the fossil record. The scientist then can draw on a sophisticated theory of empirical evidence, which explains why and how the sorts of factors biology relies on should be considered trustworthy. A second strength to her position is its fallibilism. Science does not consider any methods or results incontrovertible. Even the best of today's methods and results might need to be reconsidered on the basis of future findings. So science builds in stabilizers to accommodate its recognition that with the advancement of understanding come refinements in methods and epistemic standards. The conviction that standards and methods neither can nor need ever be revised leaves the dogmatist in an epistemically precarious position.

It may seem paradoxical that commitments we are prepared to revise are epistemically more solidly grounded than ones its adherents consider unrevisable. But that is so. The epistemic value of science lies not in the conclusiveness of its results, but in the self-correcting character of its methods and the self-refining nature of its standards. Because we consider a method good, we consider its deliverances trustworthy. Because it delivers results that satisfy our epistemic standards, we consider it a good method. Because results that satisfy these standards promote our epistemic ends, we consider these standards the proper ones. Our views about the particular theses we

take ourselves to have reasons for intertwine with the methods we take to count as generating reasons, and the standards we take reasonable opinions to have to satisfy. The interdependence is dynamic. As our understanding of a topic grows, so does our understanding of the best ways to investigate it, to validate the methods for investigating it, to calibrate the instruments used to investigate it, to identify the standards to which views about it should be held, and hence to characterize what counts as evidence in that area (see Hacking, 1983). Solidity does more than just provide insurance. It also deepens understanding. The various strands tie a solid opinion to different commitments, yielding a wider perspective as to where it fits and how it functions in our overall account.

Still, the publicity requirement may raise a worry. Suppose a brilliant mathematician constructs a valid proof that no one else can grasp.[5] Given what I've said, it might seem that he has no reason to accept the proof, since his colleagues cannot back him up. The worry is unfounded. If the proof proceeds from publicly accepted axioms via publicly accepted rules, it satisfies the publicity requirement. No mathematician has reason to reject it, since it satisfies the community's standards for an acceptable proof. Other mathematicians might, and probably should, suspend judgment about its validity until they can follow the proof. But they lack grounds for rejection. Nonetheless, the theorem prover's position is epistemically precarious. If the proof is so complicated that no other mathematician can follow it, he ought not be confident that it is free of subtle errors. This is why mathematicians check one another's proofs.

The dialectical interplay needed to achieve reflective equilibrium may take time. In the case we've been considering the proof is merely a complicated application of antecedently accepted commitments. The only difficulty is how to bring others to see that it satisfies standards that they all already endorse. Innovations can be more radical. Sometimes it is only with hindsight that we can tell that an initially tenable consideration was acceptable. When it was first proffered, its epistemic status was doubtful. The method of reflective equilibrium accords a presumption in favor of what has already been accepted. But that presumption is defeasible. So a putative proof or finding that skirts, or even violates, currently accepted standards, or is generated by methods that are not currently accepted, has an uphill battle.

But a novelty can prevail. If it is powerful or intriguing enough, it may prompt reconsideration of prevailing standards of acceptability. When first introduced, the Monte Carlo method did not satisfy accepted statistical standards. To assimilate it, statisticians had to rethink the nature of statistical sampling. They had to recognize that the pseudo-random numbers generated by von Neumann's algorithm were random enough to be reliably unbiased, and had to reconcile themselves to the ineliminable use of computers for running simulations (Metropolis, 1987). The power of the method, and the recognition that, for the purposes for which it was used, its deviations from statistical orthodoxy were unproblematic, made a convincing case for the revision.[6]

By itself, the requirement that considerations worthy of reflective endorsement be responsive to reasons says nothing about what range of reasons a particular contention or account must be responsive to. That varies with subject matter, available resources, and the relevant community of inquiry's tolerance for epistemic risk. The point here is only to show that there is a viable conception of a reason and a viable schema for responsiveness to reasons that are independent of truth-conduciveness.

The Realm of Epistemic Ends

We have seen that reasons are public and are keyed to community standards. This raises questions about the nature of the relevant community. Here again it pays to turn to Kant. One formulation of the categorical imperative is that an agent should act only on a maxim that he could accept as a legislating member of a realm of ends. These maxims are not just laws that the members of the realm of ends are subject to, they are laws that members *make themselves subject to.* In the moral realm, Kant (1981) maintains, agents are legislators who enact the laws that bind them. I suggest that the same holds in the epistemic realm. What gives second-order norms their epistemic authority is that they express standards, rules, or principles that epistemic agents can on reflection endorse. Extending Kant's idea to epistemology, then, an epistemic agent should accept only considerations that she could advocate and endorse as a legislating member of a realm of epistemic ends. Let us call this the *epistemic imperative.* Although Kant's argument for the categorical imperative is transcendental, my argument for its

epistemic counterpart is pragmatic. It focuses on the practices of inquiry and seeks to delimit the aspects of those practices that figure in their epistemic esteemability. There is no claim that these aspects are metaphysically necessary.

If agents in a realm of epistemic ends make the laws that bind them, they are not merely subject to the requirements that govern their cognitive behavior; they set those requirements. Since epistemic agents would not make commitments that they consider it inappropriate to be bound by, they think it right that they be subject to the laws of the realm of epistemic ends. They reflectively endorse those laws and the commitments that flow from and are mandated by them. They think that their being bound by a particular constellation of norms fosters their epistemic success.

In making and reflectively endorsing commitments, an agent exercises her autonomy. Because she is an agent, she can both start and end a justificatory path.[7] She considers herself justifiably bound by a network of commitments pertaining to methods, rules of inference, standards of acceptability, and the like, because she believes that being so bound will promote her epistemic ends, given her epistemic resources. The ends may, as the veritist maintains, be truths; they may be truths that satisfy further standards (for example, standards of significance or relevance); they may be nontruths (perhaps nonpropositional representations, effective models, informative idealizations, or illuminating fictions) that she takes to exemplify features of the phenomena; they may be more holistic desiderata pertaining to the system of commitments as a whole. The critical point is that agents set their ends. So even if an agent agrees with the veritist that the overarching end is truth, her relation to that end is different from his.

It might seem that reflective endorsement yields only subjective value. The agent sets epistemic ends for herself, makes epistemic commitments for herself, and on reflection is willing to stand behind them. At best this seems to assure a measure of internal consistency, but not much more. But Kant does not say that a maxim is acceptable only if an agent could enact it as the philosopher king of a realm of ends. Satisfying merely personal predilections is not enough. When an agent reflectively endorses an epistemic principle, she considers it reasonable that her cognitively serious actions—asserting, professing, and inferring—accord with that principle. She repudiates the gambler's fallacy because she recognizes that in committing the fallacy she makes herself vulnerable to Dutch books—sets of commitments

that are jointly incoherent in the strong sense that were a person to bet on all of them, he would inevitably lose. She endorses modus ponens because she recognizes that it is truth preserving. She has no reason to think that as an epistemic agent she should be subject to principles that other similarly situated epistemic agents are not subject to. She does not think, for example, that the desirability of avoiding Dutch books is a matter of personal preference, for she recognizes that commitments vulnerable to Dutch books are not cotenable. Because she takes the principles she reflectively endorses to be reasonable and rational in the epistemic circumstances, she thinks that they should be binding on similarly situated epistemic agents. But she recognizes that epistemic agents should be subject only to principles that they consider worthy of their reflective endorsement. By her own lights, then, the only principles that merit her reflective endorsement are those that similarly situated epistemic agents—those who constitute her epistemic community—could also endorse. If she is wrong to think that her compatriots can reflectively endorse *p*, she is, by her own lights, wrong to endorse *p*.

Her relation to other epistemic agents pays dividends. Their commitments serve as a check on hers. Confronted with what looks to be a camel on Main Street—or indeed, a car on Main Street—she can ask, 'Did you see that?' Arriving at a surprising solution for a complicated calculation, she can ask, 'What did you get?' Her interlocutors' agreement serves as corroboration for the conclusion and for the adequacy of the first-order considerations that led to it. Their disagreement should give her pause.

For all I've said so far, the support compatriots provide is consonant with an epistemic individualism where each agent amasses and assesses her own reasons for a conclusion. She thinks that similarly situated agents will share her conclusions, because each has amassed the same evidence on his own and assessed it using the same standards. If her reasons are good reasons for a particular contention, her standards for assessing such things are appropriate, and she has applied them correctly to her reasons, it is no surprise that her similarly scrupulous compatriots agree. According to epistemic individualism, it is each epistemic agent's standing in the same relation to the evidence, not each agent's standing in any particular relation to other agents, that accounts for their agreement. I suggest, however, that epistemic interdependence runs deeper. It both underwrites certain judgments and grounds the norms that govern epistemic practice.

Compatriots' responses can validate or invalidate the methods by which agents form and justify their opinions. For example, by himself, an agent cannot tell that he is color-blind. He can tell whether he is a consistent judge of color. The items that looked red to him yesterday look red to him again today; variations in the apparent colors of enduring objects correlate with variations lighting conditions; one visual deliverance as of color supports another. Even so, some items that consistently appear red to him might be green. People discover that they are color-blind by learning that other people discriminate shades that they cannot tell apart. If an agent cannot appeal to other people's color perceptions to check his own, he has no way to tell whether his socks match. For all he knows, he might be color-blind. The same holds for the other secondary qualities. On his own, the epistemic agent does not know what he is missing.

Such reliance on others is not restricted to the sensory realm. Without the support of an epistemic community, an individual lacks the resources to discover certain sorts of systematic error, such as confirmation bias. If Life-long Crusoe (who was shipwrecked at birth and reared by wolves) is given to confirmation bias, he weighs evidence that supports his convictions more heavily than evidence that undermines them. He may have no way to discover his bias, particularly if its impact is slight. Nevertheless, many of his beliefs are fated to be unjustified. Suppose, however, that he is not prey to confirmation bias or kindred failings. Then he is, we may assume, as reliable as the rest of us, and his first-order evidence supports his conclusions as strongly as our first-order evidence supports ours. It might seem therefore that he is just as capable of epistemic success as we are. But we have a resource that he lacks. We are members of communities that embed and reinforce certain epistemic values. Confirmation bias involves believing or accepting that p on the basis of considerations that one wrongly takes to be sufficient. If, as is likely, the agent endorses her own basis, it is unlikely that her further scrutiny of that particular basis will disclose their inadequacy. But if she appeals to community standards and subjects her belief to (implicit or explicit) peer review, she can discover the error of her ways. She realizes that whatever she may think of her evidence, it will not pass muster with the relevant community of inquiry. To be sure, the fact that some community or other disagrees neither shows nor automatically gives her reason to suspect that she is wrong in the assessment of her evidence. But if she shares the values of that community, and recognizes

that the opinion in question is properly subject to its standards, she is in a position to see that her belief does not measure up. The protection against epistemic failings like confirmation bias derives not from the mere fact that there is a community whose standards bear on her views, but from her taking her views to be rightly subject to the judgment of that community. She endorses the public standards by which her belief is to be judged. There need, of course, be no actual public assessment. Having internalized the standards she can invoke them to assess her own views. This is why the brilliant mathematician can accept his proof.

We assess our beliefs in light of the standards that an epistemic community has designed to (among other things) filter out failings like confirmation bias. So we have reason to think that beliefs that satisfy those standards are not products of such failings. Life-long Crusoe has no such reason. His reasons for confidence in his beliefs are thus epistemically impoverished compared with ours. His epistemic situation is precarious. It is just by luck that he has managed to avoid the pitfalls.[8]

It is no accident that Kant's term 'legislating members' is plural. Legislators work together to enact laws. To be effective, they must convince their colleagues of the acceptability of the legislation they propose. This requires that the basis for their recommendation be publicly articulable and justifiable to other legislators in light of the evolving commitments and goals that they share. The Kantian realm of ends is not, as some translations have it, a kingdom; it is a commonwealth. So is the realm of epistemic ends. The reasons for acceptable epistemic commitments must be specifiable and justifiable to the other members of the epistemic community. Moreover, since the realm of epistemic ends is supposed to be the arena within which epistemic agents live their cognitive lives, those commitments must mesh. Not only must each be individually acceptable; all must be collectively acceptable. This means that there are strong consistency and coherence constraints on what epistemic legislators can endorse.

It does not follow that all members of an epistemic community are expected to agree about everything. Some commitments—for example, those pertaining to consistency and coherence, are binding on everyone. Others set ranges of permissibility. Any commitment within a given range, if suitably enmeshed with other permissible commitments, is worthy of reflective endorsement. There are, for example, multiple acceptable interpretations of *Hamlet*. An epistemic agent need neither doubt her interpretation nor

assume her compatriot is mistaken if their opinions diverge, so long as both fall within the range of acceptable interpretations. Under some acceptable interpretations, *Hamlet* is about indecisiveness; under others it is about revenge. But even literary criticism's tolerance for diversity is not inexhaustible. Under no acceptable interpretation is the play about the fall of the Roman Empire.

Such permissiveness is not restricted to the arts. Philip Kitcher (1990) argues that rather than setting a fixed threshold for epistemic acceptance, science rightly recognizes a range within which either accepting a new theory or retaining one's commitment to the old one is reasonable. This enables the discipline to balance inertia with creativity. Scientists taking opposing positions need not disagree about what the evidence is, or about how much evidence there is. They may disagree about how to interpret the evidence, about what evidence is relevant, and/or about how much evidence is needed for acceptance. This is why, rather than saying that a responsible epistemic agent should accept only considerations her compatriots would accept, we should follow Scanlon (1998) and say that she should accept only considerations her compatriots could not, by their collective lights, responsibly reject. It is open to them to endorse a consideration if it falls within the purview of an epistemic community and, given the commitments of that community, they could not reasonably reject it.

The epistemic imperative directly vindicates many familiar epistemic virtues. Open-mindedness is one. An epistemic agent cannot reflectively endorse a contention unless she is in a position to think that other members of the epistemic community could endorse it as well. And she is in no position to responsibly think that unless she entertains the alternatives they might consider credible. So she must act open-mindedly in developing the contentions she advocates. Her open-mindedness is not boundless: it need not extend to entertaining hypotheses involving interventions by space aliens, since those hypotheses are not ones that her intellectual compatriots would, or by their own lights should, take seriously. Rigor, responsiveness to evidence, and impartiality receive similar treatment: the scope of the demand a virtue places on the agent is delimited by the epistemic community. Each agent stands in a reciprocal relation to other members of the community. Besides venturing only hypotheses that she considers worthy of their reflective endorsement, she must be in a position to responsibly reflectively endorse, reflectively suspend judgment on, or reflectively

repudiate their hypotheses. She needs to be able to think and act impartially, fair-mindedly, and knowledgeably.

Other epistemic virtues are vindicated by what we have come to understand about our epistemic situation. As we increase our understanding of a topic, we understand more about our understanding of that topic. Object-level advances in understanding thus pay metalevel dividends. We devise new methods and design new instruments. We discover hitherto unforeseen limitations and unacknowledged biases that cast doubt on results generated by old methods and instruments. We uncover mistakes and diagnose their sources. Considerations that our predecessors reflectively endorsed no longer merit reflective endorsement. Considerations they would responsibly have rejected we are in a position to endorse. Two hundred years ago, physicists readily endorsed the contention that space is Euclidean and, had it been offered to them, would have adamantly refused to endorse the contention that mass increases with acceleration. Now we know better. Because the history of inquiry shows that advances in understanding often involve correcting the errors of our forebears, intellectual humility is a virtue; intellectual arrogance, a vice. An epistemic agent, seeing that other agents who satisfied the standards of their communities of inquiry turned out to be wrong, should recognize that even though she satisfies the standards of her community she too might be wrong. She cannot claim that her contentions are permanently acceptable, only that they are as reasonable as any available alternative in the current epistemic circumstances. In professing them and proffering them to the epistemic community, that is what she is willing to defend.

The Division of Labor

Who belongs to a realm of epistemic ends? What qualifies someone for membership? Kant might say that the community of epistemic ends consists of all rational agents throughout history. But little is separately justifiable to or reflectively endorsable by all rational agents. What an epistemic agent is in a position to reflectively endorse depends heavily on her cognitive milieu. The contemporary community of particle physicists reflectively endorses much that is counterintuitive or unintelligible to the rest of us. But it would be wrong to dismiss their epistemic achievements because of our ignorance or incompetence.

The resolution to this problem derives from the division of epistemic labor. Earlier I said that compliance with the epistemic imperative does not require vetting each candidate for epistemic acceptance individually. We routinely endorse deliverances from creditable sources—memory, perception, and so on. One such source is testimony. On her own, each epistemic agent is in a position to know very little. To compensate for our individual shortcomings, we rely on the testimony of others. There is a division of cognitive labor both within and across the communities constituting the realm of epistemic ends. The relevant communities need not be scientific. We defer to the trained auto mechanic for an explanation of the ominous rattle in the engine, to the knowledgeable historian for an understanding of the influence of the cotton gin on American slavery, to the experienced parent for insights into how to soothe a teething baby.

Still, there are communities and *communities*. We might be fairly sanguine about saying that satisfying community standards suffices for epistemic responsibility in particle physics or auto mechanics. But what about the community of psychics? Are we forced to say that if they have managed to contrive, reflectively endorse, and by their own lights satisfy standards for contentions about the paranormal, we should credit their claims? We are not. For outsiders are in a position to assess their standards, methods, and claims, using coarse-grained measures that apply across realms.

One is predictive success. Not every discipline makes predictions; but those that do are subject to assessment on the basis of their track records. For a predictive practice to be epistemically creditable, its predictions should be borne out considerably more often than chance, considerably more often than what one would expect if its distinctive causal commitments were dropped, and should compare favorably with the track records of its rivals.[9] The predictions must be definite enough that it is possible to tell whether they are borne out. And it must be determinate whether their being borne out redounds to the credit of the putative predictor. Psychics make predictions. If the standards of their community are satisfied by predictions that are too vague to be tested, are not borne out when tested, or simply replicate with what would otherwise be expected, then satisfying their standards does not make for epistemic credibility. When Professor Trelawny divines that Harry Potter has a mortal enemy, Hermione protests, "Everybody knows

that!" (Rowling, 1999, 106). Coming up with that insight is no reason to credit divination. If we are to trust divination, Hermione rightly intimates, it should disclose things that we would not otherwise have known.

Similar points can be made about explanatory success, systematicity, fit with the commitments of neighboring areas of inquiry, and so forth. The familiar list of theoretical and empirical desiderata can be brought to bear to assess the claims of a putatively cognitive enterprise. Not only must the commitments be internally coherent and consistent, they must cohere with and be consistent with other things we reflectively endorse. If the discipline is empirical, claims need to be backed by evidence. If the practice involves prediction, the predictions must (often enough) be borne out. These are entirely familiar requirements. The commonwealth of epistemic ends is a federation. Local communities make fine-grained commitments that are answerable to coarse-grained commitments that are more widely shared. The failure of a local community's fine-grained commitments to satisfy (or at least approximate) coarse-grained requirements, unless backed by strong reasons to think that the fine-grained commitments need not satisfy them, is a reason to refuse to reflectively endorse their findings.

What justifies treating particle physics as a community in a federation of epistemic ends is that its commitments satisfy widely shared coarse-grained epistemic commitments. Particle physics respects such requirements as that a scientific account be explanatory, simple, perhaps predictive; that it be responsive to evidence that is judiciously and impartially gathered and assessed; that its results be intersubjectively observable and reproducible; that the account cohere, or at least not be incompatible, with accepted constellations in neighboring disciplines; that its findings stand up under peer review; and so on. Just how these requirements are concretized in a given discipline or subdiscipline depends on local factors; and it may be that only the members of a local community can tell whether this has been done. But it is critical that merely meeting local standards is not enough. The constellation of commitments that meets local standards is, as a whole, answerable to more global standards.

This pattern is reminiscent of what we see in law. Local parking ordinances are justified and binding in a community largely because they were duly enacted. But they are subject to broader constraints—those specified by state and federal laws, which in turn are subject to constitutional

constraints. Regardless of what a particular city council enacted, a law specifying different parking fines for members of different races who commit the same infraction would be impermissible.

The Political Character of the Epistemic Community

To perform their function, legislating members of a realm of epistemic ends must be, in a suitable—political—sense, free and equal. Agreement among free and equal inquirers enhances the epistemic standing of a claim; coerced agreement does not. If inquirers are free, they can adopt any perspective they like, and examine the issue from there. If they are equal, all inquirers have an equal opportunity and an equal right to venture hypotheses, to raise objections and counterhypotheses, and provide reasons for them. It is not enough that members be free to venture their opinions. Their interlocutors must also be disposed to listen to them and take them seriously (Longino, 1990).

Members of a properly functioning epistemic community see those who disagree with them not as rivals to be bested in argument, but as resources whose perspectives may provide valuable insights (Dewey, 1916). By appreciating one another's points of view, deliberators enrich their understanding of an issue and the available initially tenable ways of thinking about it. Entertaining a new perspective may convince an agent to adopt it. She may see that her original assessment of the evidence or the options was limited or biased. Alternatively, such deliberation may lead to the adoption of a conclusion that no one initially envisioned. That the evidence admits of multiple interpretations may reveal that none by itself is adequate and point toward a different solution. By accessing the opinions of others and the reasons for those opinions, epistemic agents augment their data base. They gain new information about the situation and ways it might be fruitful to think about it. So rather than seeing those who disagree with us as opponents, we should see them as allies who, by envisioning things differently, extend our epistemic range. Diversity of opinion thus is not an impediment to deliberation but an asset to it.

On this conception, a community of inquiry provides a venue for fruitful disagreement. Not every disagreement is worth taking seriously. There is little to be gained by entertaining the economic views of someone convinced that extraterrestrial forces control the stock market. But when conscientious,

competent peers with a suitable level of expertise about a topic disagree, the situation is different.

When parties disagree, it might seem obvious that at least one of them is mistaken. Either he overlooked some relevant evidence or erred in his use of it. These are typically identifiable and corrigible errors. Once the reasons are spelled out, it may be straightforward to see where the error lies. One party miscalculated or ignored base rates or overlooked data or whatever. Not all disagreements are like this. Sometimes the root of a disagreement is not an error; nor is it something that needs to be corrected. Parties to a disagreement might assign different (reasonable) weights to different bits of evidence. Fred might, for example, take a blood test to show that the patient's hemoglobin count is dangerously low. George, placing greater emphasis on the trajectory, notes that this count is higher than the previous reading. Fred believes that the patient is unlikely to survive. George believes he is on the mend. Both agree about the facts. Their disagreement concerns which fact is more significant. A disagreement about a matter of fact, then, can be rooted in a disagreement about how to weigh the evidence.

Disagreements can also arise when parties consider different factors salient. When an anorexic patient exhibits a dangerous tachycardia after consuming high levels of caffeine, the emergency room physician judges that excessive caffeine consumption is the cause (Ankeny, 2014). He might overlook or downplay her anorexia, concentrating entirely on the fact that she has overdosed on caffeine. The expert in eating disorders judges that anorexia is the cause, since anorexia gives rise to tachycardia in the presence of a suitable stimulant. Each thinks something true. The difference of opinion lies in the link in the causal chain that they focus on. Their choice of focus stems from what each can do to alleviate the problem and prevent its recurrence. Here the difference is within the field of medicine. But similar disagreements can cut across fields. On a rainy night, a car crashes into a tree. The police say that speed caused the crash; the auto engineer says that the unresponsiveness of the steering column was the cause; the highway architect places the blame on the way the road was banked. In such cases, no one is saying anything false or misleading. Given their different interests and areas of expertise, they simply focus on different aspects of the situation.

Jamesian veritists hold that our overarching goal is to believe as many truths as possible and disbelieve as many falsehoods as possible. I have

argued that this ought not be our overarching goal. Nevertheless, where we are concerned with what to believe, we face a trade-off between the value of believing truths and the disvalue of believing falsehoods. To a remarkable (and alarming) extent, advocates of Jamesianism ignore the trade-off and focus exclusively on believing truths. But the goal as stated involves a proportion. To settle on a criterion of justification requires weighing the values against each other (see Riggs, 2003). There is no reason to think there is a uniquely best way to assign weights. If the threshold for acceptability is a probability of 0.90 on the evidence, a few false hypotheses will count as justified. If the probability has to be 0.95, even fewer will count. On the other hand, when we raise the standard from 0.90 to 0.95, fewer truths will count as justified. In moving from 0.90 to 0.95, we accept fewer falsehoods and reject more truths. Setting the threshold involves a trade-off. It is widely and rightly conceded that for a hypothesis to count as justified, it must be highly probable given the evidence. That, by itself, does not tell us where to draw the line. Nor does it commit us to the claim that there is a single, optimally correct line to be drawn. A disagreement among experts might be due to differences in their tolerance for epistemic risk. They could conceivably agree about what the evidence is, what weight to attach to the evidence, and which factors are and should be salient, but still disagree because one holds and the other denies that a probability of 0.92 is justification enough.

Where does this leave us? We've seen that the opinions of epistemically responsible agents can diverge along a variety of trajectories. Divergence in and of itself does not show that any of the parties to a dispute is mistaken. Disagreements among experts—those who are competent, conscientious, and suitably sensitive to the level of confidence that their evidence supplies—provide information about the community's current understanding of an issue.

A community of inquiry may have resources to adjudicate some disagreements among its members. It may currently lack resources to adjudicate others. In that case, the disagreement reveals something about the lacunae in its understanding of its subject matter. It may also be in a position to recognize that some legitimate disagreements need no adjudication. Kitcher's point is that the community may best serve its collective epistemic ends by, at any given time, countenancing a range of conflicting views. The reason is largely prudential. When there is a significant chance that a currently disfavored view is true, or that a currently favored one is

false, if the members of the community want to believe what is true, they ought not foreclose inquiry prematurely. Riggs's point is that multiple ways of balancing the value of believing truths against the disvalue of believing falsehoods are permissible. This has two important consequences. The first is that there may be no general answer to the question what balance should be struck. For some issues, avoiding falsehood may be more important; for others, believing truth might dominate. How risk averse epistemic agents should be, then, depends on what they are risking. The risk in question may be practical, as in risking an epidemic by accepting false views about contagion, or purely theoretical, as in risking an epistemically costly mistake, such as one that is likely to prompt a cascade of false beliefs. Moreover, even in the long run, there may be no objective resolution to some disputes. There is no reason to think that there is a single optimal balance.

The value of disagreement for a community of inquiry is that it can serve as a source of insight into the subject under investigation and the available resources for understanding it. Socrates maintained that knowing that one does not know is epistemically valuable. What may be more valuable is recognizing what one does not know and understanding why one does not know it. Suitably expert disagreement can engender such understanding.

Deweyan Deliberation

If free and equal members of an epistemic community engage in Deweyan deliberation, there are no political impediments to understanding the phenomena. But if consensus results from coercion, collusion, or credibility inflation or deflation, the situation is different. If inquirers cannot investigate an issue as they see fit, if they cannot raise objections or gain a hearing for them, if their views are given undue weight or are deprived of the weight they are due, intersubjective agreement provides small reason to think that a conclusion is acceptable, or that it stands up to serious testing.

Arrangements that silence certain voices or deflate their credibility can undermine the epistemic standing of the commitments that are accepted (see M. Fricker, 2007; McGowan, 2009). By omitting or downplaying the significance of particular perspectives, or by unduly emphasizing other perspectives, prejudice and stubbornness can skew matters so that intersubjective agreement does not supply corroboration. If naysayers are silenced, then

the fact that every speaker agrees is not a good reason to think that a contention is acceptable or a method is sound. If the credibility of naysayers' testimony is deflated, then the fact that the balance of what is taken as evidence weighs heavily in favor of a hypothesis is at best a weak reason to accept it. The negative epistemic consequences do not just undermine the epistemic standing (and self-confidence) of those who have been silenced. They affect the entire epistemic community. Without access to the discredited opinions, community members do not know what they are missing. Moreover, the skewing of the evidence that results from credibility deflation undermines the trustworthiness of what evidence they have, for it artificially inflates the credibility of the voices that are heard. Such a situation is epistemically unjust.

An example brings this out. In the 1940s and early 1950s Barbara McClintock published a series of papers contending that she had discovered sequences of genetic material that change position on the chromosome of corn. She was not taken seriously. As a result, she said, "I stopped publishing detailed reports . . . when I realized, and acutely, the extent of disinterest and lack of confidence in the conclusions I was drawing from my studies" (McClintock, 1973). Historians of science disagree about whether sexism figured in the scientific community's disregard of McClintock's findings. The other candidate is a prejudice against studies done on corn (see Keller, 1983; Comfort, 2001). Whatever the reason, she was effectively silenced. Her discovery was not taken up until the 1960s when François Jacob and Jacques Monod discovered that the same transposition occurs in bacteria. Plainly the understanding of genetics was retarded by the failure to give McClintock's discoveries their due. Such epistemic injustice deprives the community of inquiry of data and skews the evidence.

'Free and equal,' here, is a political characterization of a community of inquiry. That community may be located in a wider setting whose members are politically neither free nor equal. A repressive society might, for example, accord greater freedom of thought to its scientists in the context of their investigations than it generally accords to its citizens. If the constraints the society normally places on its citizens are not operative in the context of inquiry, the epistemic commitments inquirers reflectively endorse are legitimately normative for their inquiry, hence for the nonce binding on their practice.[10]

Even so, the idea that the members of communities of inquiry are free and equal inquirers is an idealization. I do not claim that actual communities fully accord with it. I do, however, claim that the closer actual communities come to satisfying these requirements, the more epistemically acceptable their criteria are likely to be. This is because we understand how inequalities of power and limitations on orientation can lead to error. By reflecting on the strengths and weaknesses in the ways the actual communities of inquiry formulate their standards and construct their practices, we contrive a conception of what is epistemically acceptable. The requirement of freedom and equality thus emerges from the method of reflective equilibrium. We draw on, correct, extend, and emend our current best theories and practices until we arrive at a system we can on reflection endorse.

The free and equal requirement is merely a political requirement. It is not a claim that all are equally knowledgeable or equally intelligent. Some hypotheses are plainly untenable; some perspectives are obviously skewed; some methods are demonstrably unsound. Once they are given a hearing, they are promptly and rightly dismissed. Nevertheless, the opportunity to venture a hypothesis and the right to have it assessed on its merits must be real. There must be forums where criticisms can be offered, and where there is an expectation that criticisms will be answered (see Longino, 1990). As Mill insists, "The beliefs we have the most warrant for have no safeguard to rest on but a standing invitation to the whole world to prove them unfounded" (1978, 20). The epistemic value of what Habermas (2001) calls uncoerced conversation lies, I suggest, not so much in the particular agreements it generates as in its propensity to uncover and correct errors.

Since 'ought' implies 'can', the standards the responsible epistemic agent must respect are not those of some idealized, timeless community of rational agents. She could have no idea what those standards are or what it would take to satisfy them. Rather, she must respect the current or reasonably foreseeable standards of a realm of epistemic ends she inhabits. Just what these standards are is not determined a priori. They emerge over the course of inquiry, as the community learns about the subject matter, how to investigate the subject matter, and what sort of understanding of the subject matter is practically possible in the epistemic circumstances, given the resources at hand, and the questions the community wants to answer. Respecting the standards is not the same as being obliged to comply with

them. Rather, respect requires that a responsible epistemic agent either comply with the standards of her epistemic community or supply a cogent reason to revise them, reject them, or restrict their range of applicability.

Nevertheless, the extent to which we have to rely on one another might seem to undercut any claim to epistemic autonomy. Often, it seems, we amateurs are in no position to defend our own beliefs, as they depend on fine-grained epistemic commitments that we are not privy to and would probably not understand even if we were. We defer to the community of particle physicists to specify what counts as evidence for the existence of the Higgs boson, what counts as sufficient evidence, and when such evidence is in hand. We have no choice. Nevertheless, epistemic commitments must be defensible. If my defense of the contention that the Higgs boson exists consists in saying, 'They said so!', it looks weak.

It is not. Although expertise is distributed across the epistemic community, this does not undermine the responsibility of the individual epistemic agent. She confers epistemic authority on those she counts as experts; and she retains both the right and the responsibility to revoke it if the circumstances warrant. She can and at some point should decide that the auto mechanic does not know what he is talking about (even though she does not know what a catalytic converter does), that the investment counselor's advice is not credible (even though she is only dimly aware of what a hedge fund is), that the physician who links autism with vaccines is untrustworthy (even though she has no idea why the incidence of autism is rising). The requirement that autonomous epistemic agents think for themselves is not in tension with the recognition of the extent of our epistemic interdependence. Rather, it means that each agent is ultimately responsible for her choice of experts. As a legislating member of a realm of ends, she confers epistemic authority, thinking that other similarly situated epistemic agents would do the same; and she can and should retract that authority when she considers it no longer warranted. If a contention's defense consists of 'They say so!' backed by defensible reasons for thinking that *their* saying so is trustworthy, it is not a weak defense; it is a strong exercise of epistemic autonomy.

I have argued that epistemic norms are deontological. They are grounded in the epistemic imperative. An epistemic subject ought to accept only what she could reflectively endorse as a legislating member of a realm of epistemic ends—that is, as a responsible citizen/legislator of a united federation

of autonomous epistemic agents. To abide by the imperative requires standing not just in a suitable relation to the phenomena she seeks to know or understand, but also in a suitable relation to other members of the epistemic community. On this view epistemic norms do not, nor do they purport to, ensure that epistemic agents will arrive at the truth. Nor do they contend that failure to reach the truth is always a mark of epistemic inadequacy. But by acting in accord with the epistemic imperative, an agent makes the best use of available epistemic resources.

Epistemic standards owe their normative status to their satisfying the epistemic imperative: they are standards that would be framed and reflectively endorsed by the legislating members of a realm of epistemic ends. Although a realm of epistemic ends is modeled on a scientific community, it obviously involves an element of idealization. As the McClintock example illustrates, we should not assume that actual communities of inquiry consist of free and equal inquirers. Rather than describing the norms accepted or acted upon by an actual community, the ideal emerges from a process of adjudication in which we elaborate, extend, and correct our relevant commitments until we arrive at a conception that is in reflective equilibrium with our other epistemic commitments.

The actual communities we respect serve as touchstones. We treat their policies and practices as initially tenable and attempt as far as is feasible to incorporate them into the constructed ideal. The scientific community is one we respect. It regularly deploys felicitous falsehoods—models, idealizations, and thought experiments. The emendations we saw fit to introduce into our ideal correct for bias and prejudice. They do not undermine the permissibility of relying on felicitous falsehoods. So, absent independent reasons to consider reliance on such devices epistemically deleterious, we should recognize that scientific (and other disciplinary) communities are within their epistemic rights to deploy them.

6 Intellectual Integrity

Epistemic acceptance bridges the divide between the intellectual and the practical.[1] It is practical in that it consists in a readiness to use a commitment in inference or action—that is, a readiness to do something. It is intellectual in that its range is restricted to contexts where the agent's ends are cognitive. Although the practical realm is broader than the moral realm, my account raises the possibility that moral norms figure in epistemic acceptability. This suggestion is strengthened by my characterization of a realm of epistemic ends. Legislating members of such a realm have obligations to one another—obligations to treat one another as free and equal members in a joint epistemic enterprise. These obligations are at least political, if not moral. In this chapter I argue that moral requirements do not just constrain, but infuse responsible disciplinary practice. They not only promote the epistemic ends of a discipline, they are partially constitutive of those ends. I begin with a sketch of what Williams calls 'thick concepts', concepts like 'disreputable' and 'courageous', in which evaluative and descriptive elements are fused. I argue that trustworthiness is such a concept. I then consider how it shapes inquiry. I look at familiar threats and explain how they undermine the epistemic enterprise. This sheds light on what the enterprise is and what it achieves.

Thick and Thin

The widespread conviction that the moral and the factual are disjoint seems to stem from focusing on moral predicates like 'good' and 'bad', 'right' and 'wrong', which appear to have purely evaluative contents. But, as Williams shows, many predicates are simultaneously evaluative and descriptive. They are, in his terms, thick (Williams, 1985, 140–143). To say that a political

maneuver is sleazy is typically to both characterize and disparage it. Occasionally, with a tinge of irony, 'sleazy' can express admiration. But, evidently, its use is never value neutral. 'Good' and 'red', on the other hand, are thin. That an item instantiates the predicate 'good' reveals nothing about what sort of thing it is; that it instantiates 'red' reveals nothing about whether it is any good.

To appreciate the difference, consider the newly coined predicate 'gred', which applies to all and only things that are both good and red. Such a predicate, like 'sleazy', is both descriptive and evaluative. A child's mitten satisfies the descriptive requirement by being red, and satisfies the evaluative requirement by being good. But a gred item's being red has no bearing on its being good; nor has its being good any bearing on its being red. The mitten qualifies as red because of its color; it qualifies as good because of its capacity to keep a small hand warm. The mitten just happens to instantiate the two predicates; it belongs to their intersection. Because we can prize apart the descriptive and the evaluative components, 'gred' is not a genuinely thick term. It is simply a contraction.

The relation between the descriptive and the evaluative elements in thick terms like 'sleazy', 'truthful', and 'trustworthy' is tighter. Thick terms involve a fusion of descriptive and normative elements. They do not admit of the sort of factor analysis that 'gred' does. Although 'sleazy' is a descriptive predicate whose instances we readily recognize, we have no evaluatively neutral way to delineate its extension. We have no way to identify the class of sleazy political maneuvers without intimating that they merit disapproval. If epistemic terms are similar, then we should have no way to identify their extensions without intimating that they are worthy of approval or of disapproval on cognitive grounds.

In *Truth and Truthfulness*, Williams (2002) discusses the ways the thick concept of truthfulness depends on and diverges from its thin descriptive precursor—the concept of uttering truths. I briefly review his discussion because it illustrates how thick terms relate to their less corpulent relatives. He develops a fictional genealogy to show how the virtue of truthfulness could have emerged from the practice of uttering truths. As a first approximation, a truthful person is someone who utters or is disposed to utter truths. We can identify such people without taking a stand on the value of uttering truths, just as we can identify terse people without taking a stand on the value of brevity. The evaluative element enters the picture later,

when we deem uttering or being disposed to utter truths a good thing. At the outset, there is no fusion. The descriptive and evaluative elements can easily be separated. This does not last, because the virtue of truthfulness does not consist exactly in uttering or being disposed to utter truths. It involves both more and less than that.

Obviously, we do not and ought not expect a truthful speaker to utter or be disposed to utter only truths. Everyone makes mistakes. So if anyone is to count as truthful, truthfulness cannot require uttering only truths. It might, however, require saying only what one thinks is true. That would leave room for mistakes but exclude lies. This is a step in the right direction, but it demands too little, for at least two reasons.

First, misleading is as much a violation of truthfulness as lying is. It is possible to utter only truths and, by exploiting implicatures, induce one's audience to form a false belief. If Ben tells his editor, 'I can't get the paper in before Tuesday', he gives her to believe that he can get it in on Tuesday. If he knows full well that there is no chance it will be done before three weeks from Friday, he has not been truthful, even though the sentence he uttered was true.

Second, being a virtue, truthfulness involves responsibility. We would not consider a person truthful if she were unduly careless or gullible in her belief formation. Suppose that, despite her excellent education, Joan believes everything she reads in the tabloids and indiscriminately reports her beliefs. She believes with equal conviction that space aliens landed in Detroit and that Barry Bonds took steroids. Although she uttered the true sentence 'Barry Bonds took steroids', we would not consider her truthful. For she was equally prepared to utter the infelicitous falsehood, 'Space aliens landed in Detroit'. Her utterances are not credible, since she does not adequately filter out infelicitously false claims. Although she says what she believes, she lacks the virtue of truthfulness.

Nor is it enough to impose a responsible belief requirement. Truthfulness does not require uttering every truth one responsibly believes. Some truths are private. When asked about them, Williams maintains, evasions, obfuscations, and even lies are permissible. Your questioner may have no right to know whom you plan to vote for, what your salary is, or how much you weigh. If refusing to answer unduly intrusive questions is not feasible and 'It's none of your business!' is too blunt, Williams contends, a truthful person need not tell the truth. Her interlocutor should recognize that in

asking such questions, she oversteps the bounds within which truthfulness is required.

I would go further. Someone who utters and is interpreted as uttering a felicitous falsehood would normally count as truthful. A regular rider who says, 'It will take you twenty minutes to get to Park Street by subway', would normally be considered truthful even if the transit authority lists the travel time as eighteen minutes. At best, truthfulness requires being true enough in the context of utterance.

Truthfulness, as Williams characterizes it, is intimately related to, but does not supervene on, uttering truths. For the requirements on speaking truthfully diverge from those on uttering truths. The reason for the divergence is practical. Because a more nuanced attitude toward truths is more valuable than uttering all and only things one believes to be true, that more nuanced attitude is the one we should admire and cultivate. Truthfulness does not admit of factor analysis because, without appealing to evaluative considerations, we have no way to delineate the extension of the class of topics about which one should say only what is true enough.

If Williams is right about the relation of a thick concept to its thin descriptive precursor, thick concepts are not a mere convenience. Without such normatively loaded concepts, we could not partition the world as we do. So the idea that thick concepts are simply contractions like 'gred', the idea that that we could make do with morally thin concepts—good, bad, right, and wrong—plus our stock of purely descriptive concepts, is incorrect. Lacking the thick concepts, we would have no way to mark out extensions we are interested in. If there are other thick epistemic concepts, we should expect them to display the same pattern. They should be anchored in, but not supervene on, a descriptive core. They should mark out extensions that we have no evaluatively neutral way to demarcate.

Trustworthiness

A critical consideration in fixing the contours of truthfulness is the idea that truthful people are trustworthy. Within the realms where truthfulness is required, when a truthful speaker says that p, his audience can take his word for it. Trustworthiness too is a thick concept. While truthfulness pertains to linguistic behavior, the scope of trustworthiness is broader. We deem an auto mechanic or a pastry chef trustworthy primarily on the basis

of what she does, not on what she says. Trustworthiness is the genus of which truthfulness is a species.

Williams contends that truthfulness requires Competence and Sincerity (where the upper case indicates that these are quasi-technical terms). Competence is a general requirement, applying to all trustworthy behavior. Sincerity, however, seems restricted to symbolic behavior. Utterances, inscriptions, exemplars, and gestures can be Sincere or Insincere. But Sincerity seems not to apply to other forms of behavior. The mechanic who purposely uses inferior parts has some sort of character defect, but it is not clear that his knowingly installing a faulty gasket can be considered Insincere. Sincerity is, I suggest, a species of the generic concept of being Well-Intentioned.[2] In installing the faulty gasket, the auto mechanic is not Well-Intentioned. To be trustworthy, an agent needs to be both Competent and Well-Intentioned. A Competent but Ill-Intentioned auto mechanic has the ability but lacks the incentive to properly fix the car; an Incompetent but Well-Intentioned mechanic has the incentive but lacks the ability.

How does this bear on epistemology? Because of the division of cognitive labor, epistemic agents need to depend on one another. So epistemic agents should be trustworthy. This involves having and properly using appropriate background assumptions and know-how. Having them can perhaps be construed purely cognitively, but being willing to properly use them is a matter of volition. If agents work together, or depend on each other's actions, utterances, or products, they need to be in a position to trust one another. They need to be able take it for granted that those they rely on are in relevant respects both Competent and Well-Intentioned. I will argue that it follows from the division of cognitive labor that understanding has an ineluctably moral dimension. My main concern is not with individual concepts, but with norms and standards that are at once epistemic and moral. Again, for convenience, I focus on science, with an occasional aside to recognize parallels in other disciplines.

Ethics in Inquiry

Section 7009 of the America COMPETES Act (Federal Register, 2009) requires that all undergraduates, graduate students, and postdocs working on projects funded by the National Science Foundation (NSF) be given 'appropriate training and oversight in the responsible and ethical conduct

of research'. Probably no one denies that scientists should conduct their research ethically;[3] and few would deny that young scientists need to learn how to do so. Nevertheless, many principal investigators apparently regard the requirement as an unfunded mandate—yet another demand on their already strained budgets, time, and attention. Evidently they think that teaching their protégés to be moral scientists is something different from teaching them to be scientists. As a result, generic handbooks and websites have arisen to instruct students how to recognize, avoid, and protect themselves from the more obvious forms of scientific misconduct. Such quick fixes treat moral requirements as addenda rather than as integral to scientific inquiry. This is a mistake. I will argue that there is an ineliminably moral moment in scientific inquiry. This feature of systematic inquiry is regularly overlooked. Ignoring it skews our understanding of the scientific enterprise and its products. If I am right, then although they may not have realized it, senior scientists have been imbuing their juniors with the relevant moral values already. The point of the NSF mandate, then, is not to ask them to take on an additional onerous chore for which they are, or may think they are, unqualified, but to do explicitly and self-consciously what they had been doing tacitly and unconsciously all along. Although the bulk of this chapter focuses on science, the moral dimension I discuss is integral to every discipline. Insofar as inquirers have to depend on one another, they have moral/epistemic obligations to one another. The understanding that emerges is shaped by the recognition of those obligations.

An overarching epistemological objective of science is to develop a comprehensive, systematic, empirically grounded understanding of nature. Although it is not science's only objective, I will call such an understanding its *epistemic goal*. It is common to all the sciences: biology seeks to understand living organisms; meteorology, to understand atmospheric conditions; physics, to understand matter and energy; and so on. Two obstacles get in the way: (1) Nature is enormously complicated. (2) Findings are fallible; no matter how well established a conclusion is, it still might be wrong. I will argue that in order to pursue its goal in light of these obstacles, science incorporates values that are simultaneously moral and epistemic. These values, I will urge, are not mere means; they are integral to the sort of understanding science provides and underwrite its epistemic standing.

Very roughly, the argument is this: Scientific inquiry requires collaboration. That collaboration requires trust. Since trust is unreasonable in

the absence of trustworthiness, scientists need to be, and to consider one another, trustworthy. Since trustworthiness is a moral attribute, scientific inquiry has an inextricably moral dimension.

Scientific inquiry is a social undertaking. Experiments, models, findings, and theories may be fruits of individual efforts, but they are constructed and evaluated by reference to standards devised, endorsed, and enforced by a scientific community. These standards, and therefore the items they sanction, are not arbitrary. They are grounded in a shared conception of the scientific enterprise—its ends, means, liberties, and constraints. They are, that is, generated and reflectively endorsed by legislating members of a community of epistemic ends. Moral values are integral to science. They emerge from and contribute to the scientific community's conception of its enterprise. The practices they engender, the constraints they impose, and the activities they promote figure in its success. Although the values in question may initially be merely instrumental, they become intrinsic when scientists recognize that realizing them not only promotes good science, but in part constitutes good science.

Epistemic Interdependence

Modern science accommodates itself to the complexity of nature through a division of cognitive labor. Rather than starting from scratch, each scientist builds on previously established findings, uses methods others have devised and tested, employs measuring instruments others have designed, constructed, and calibrated, and analyzes data using mathematical and statistical techniques others have validated.[4] Research teams rather than solitary investigators frequently generate results. Such teams can do more than any one of its members could do alone, not only because they can cover more ground, but also because team members typically have different cognitive strengths and often have different areas of expertise. Collectively they bring to bear a broader range of talents and training than any individual scientist possesses. This is not to say (although it might well be true) that it would be impossible for an isolated individual to come up with an empirically adequate understanding of some aspect of nature.[5] My claim is only that the sort of understanding modern science supplies is rooted in epistemic interdependence. Science as we know it is a cooperative endeavor.

Because it depends on a division of *cognitive* labor, scientific collaboration occurs in a particularly harsh epistemic climate. Collaborators need not know each other well or understand each other's contributions. The paper reporting the sequencing of the human genome had over 250 authors, working in institutions on four continents (Venter et al., 2001). Most of the authors were in no position to vouch for one another's intellectual character or monitor one another's work. Whether or not they were in the same venue, many were probably at best dimly aware of what some of their coauthors contributed. Although few scientific papers have so many authors, such a lack of proximity and close personal ties among members of a research team is common. Moreover, all scientists build on the findings of, and deploy methods developed by, predecessors they do not personally know. And even when they work in close physical proximity, the intellectual distance between collaborators who differ in expertise can be vast. Statisticians often have no clear idea how bench scientists generated the raw data they are expected to analyze. Nor do the bench scientists know precisely how or why the statisticians analyze the raw data as they do. A scientist does not always know the reasons behind a conclusion a collaborator drew; and if she were told, she might not fully appreciate their force (Hardwig, 1991).

Such epistemic interdependence yields significant benefits. It extends epistemic reach: More data can be gathered, more experiments run, more matters investigated, more factors considered, more perspectives accommodated than would be possible for a single researcher working alone. It stabilizes findings: Individuals are subject to bias, inattention, carelessness, and wishful thinking, as well as to limitations in background knowledge and scientific imagination. If the result of an investigation is the product of many minds, individual foibles can cancel out. What one team member overlooks, another may notice; what one slights, another may focus on; what one never learned, another may have studied in depth. It solidifies results: because of team members' differences in expertise, the quality of evidence and range of interpretations available through teamwork is apt to be better than the quality and range available to an individual (Hardwig, 1991). Collectively they can advance understanding more than any one of them could do alone.

But epistemic interdependence creates vulnerability. To accept a conclusion is to be prepared to use it as a basis for cognitively serious inference and

action, and to be prepared to give others to accept it as well. To the extent that one scientist does not know or understand what another did, she is in danger of accepting and basing her own investigations on unwarranted results. If she does, she may pursue fruitless lines of inquiry and overlook promising ones. If she coauthors a paper contaminated by a collaborator's scientific misconduct, her professional reputation and livelihood are at risk. Epistemic interdependence thus requires trust. Scientists can reasonably depend on one another's contributions only if they can be confident that those contributions are scientifically acceptable. Perhaps personal acquaintance grounds trust among a small number of investigators who work closely together. But the requisite epistemic interdependence extends across the scientific community, so personal ties are insufficient. Trust in one's fellow scientists must be impersonal. A major function of the infrastructure of science is to secure the bases of impersonal trust. Scientific misconduct betrays that trust.

Misconduct

The most flagrant forms of scientific misconduct are fabrication of data, falsification of findings, suppression of results, and plagiarism. Fabrication and falsification misrepresent what was done; plagiarism misrepresents by whom it was done; suppression conceals what was found. The obvious moral objection to all such conduct is that it is dishonest. But what, if anything, does this have to do with its being bad science?

To *fabricate* data is to issue a report of an investigation that was not conducted (or not conducted for the cases reported). To *falsify* a finding is to report that an investigation had a result that it did not actually have. Both fabrication and falsification are lies. Thus, one might think that they subvert the epistemic goal by introducing infelicitous falsehoods into science, representing them as true. That would obviously be bad science. This explanation assumes that a contention grounded in falsification or fabrication is infelicitously false. That need not be so. Fabrication and falsification misrepresent the epistemic status of a claim; they purport that it has backing that it does not have. Such misrepresentations lack justification. But a statement that lacks justification may nevertheless be true or true enough. Fabricated data may be just what would have been found had the appropriate experiment been run; a falsified finding may be true or true enough even though no (or unacceptably little) evidence confirms it.

In fact, scientists who falsify or fabricate rarely intend to introduce infe-licitous falsehoods into science. Because science is self-correcting, they could not reasonably expect to succeed, at least not in the long run. Rather, they seek to incorporate into science what they take to be a truth, but one for which they do not have, or do not yet have, evidence. This makes the prob-lem trickier. If the conclusion they purport to have established is true, one might argue, our understanding of the phenomena under investigation is enriched rather than degraded by its acceptance. And if the perpetrator has good instincts about where the truth lies, his falsified or fabricated conclu-sion might well be true. If it is, what is the problem?

If truth is the end, and scientific research a mere means to it, falsifi-ers and fabricators with good instincts or good luck simply take shortcuts. They arrive at the end without deploying the standard means. They skip the intermediate steps of acquiring and evaluating empirical evidence. If all that matters is truth, how a truth was arrived at should make no differ-ence. But mere truth is not what science is after. A science is not just a compilation of discrete bits of information about a topic that happens to be accurate; it is a comprehensive, systematic understanding of a range of phenomena, based on carefully acquired, scrupulously recorded, empirical evidence. The tie to evidence is crucial; for evidence supplies the ground-ing that underwrites epistemic standing. Conclusions backed by evidential support are conclusions scientists have reason to accept; conclusions, even true conclusions, without such support are conclusions they lack reason to accept.

This explains what makes the falsified or fabricated 'finding' scientifi-cally objectionable. Although the alleged finding may be true, in the absence of adequate evidence one would be unwise to depend on it. It is not trust-worthy. To accept a conclusion that lacks adequate backing is to take an epistemic risk. And to put other members of the community in a position where they are taking such a risk unawares is to fail in one's obligations to them. That is what proffering a falsified or fabricated finding does. It con-stitutes a betrayal of science because scientists have undertaken a commit-ment to proffer only results that members of the realm of epistemic ends can, by their own collective lights, trust.

The counterpart to publishing reports of research that was not done is refraining from publishing reports of work that was done. *Suppression of results* thwarts science's goal in much the way that falsification and

fabrication do. Obviously a scientist is not obliged to publish the results of every failed experiment or every fruitless or frivolous line of thought. Doing so would create an information glut of no use to anyone. But if a well-designed, scrupulously conducted investigation was carried out to establish the truth of a scientifically significant hypothesis, the fact that it failed is something the scientific community deserves to know. Suppressing such a finding is in effect a lie of omission.[6] It misleads the community, because suppressing the result unjustifiably implicates that the question it investigated is still open, that a particular approach to the question is still viable, or that a previously accepted finding has not been discredited. It implicitly invites investigators to pursue what is now known to be an unproductive line of inquiry or to rest their research on what has been found to be an untenable base.[7] This suggests that the standard journal practice of refusing to publish null results is more problematic than it appears. The failure to disclose that a promising line of inquiry went nowhere can impede scientific progress. Had the null result of the Michelson–Morley experiment been suppressed, physics would have suffered.

Even the best science affords no guarantees. However scrupulously performed an experiment is, its result may be false. There is always some risk in accepting any finding. The standards of a scientific community constitute its consensus about what risks are worth taking. But whatever the standards, the goal would be subverted if science were permanently saddled with the errors of its forebears. Science therefore incorporates methods for rooting out received errors. One is that results that cannot be reproduced or that do not stand up under further testing and elaboration must be excised from science. But it is not enough to simply declare a previously accepted finding to be false or unwarranted. To learn as much as possible from their mistakes, scientists want to know what went wrong. Rather than simply deleting an erroneous conclusion, it is desirable to backtrack to the source of the error. This is central to the objectionability of *plagiarism*.

The content of a plagiarized report may be, and I suspect typically is, beyond reproach. The problem is that it is not the product of the scientist who purports to have produced it. Again, this is clearly dishonest. But what makes it bad science? The acceptability of a finding is independent of the identity of the scientist who generated it. Indeed, if the work is beyond reproach, any competent scientist could have done the research and written it up. This almost makes it look as though science should have no objection

to plagiarism. Why should scientists care who did the research, so long as it satisfies their standards?

Giving credit where credit is due encourages good work. So an obvious objection to plagiarism is that failing to credit productive scientists while crediting unproductive ones threatens to drive the wrong scientists from the field and to direct research funding to lazy, inept, unscrupulous, or unproductive investigators. Plagiarism, then, is objectionable for the practical reason that it interferes with an effective incentive system by making it difficult to identify and reward those who are doing good work. This might be called the department of human resources objection to plagiarism.

Is there any epistemic reason to credit the individuals who actually did the research and to object when others try to steal their credit? That is, does the misappropriation of other people's work itself damage the pursuit of science's epistemic goal? An affirmative answer stems from fallibilism. No conclusion is certain. However scrupulous scientists are, however well tested their methods, and however well-honed their standards, conclusions still might be, and sometimes are, wrong. That being so, science incorporates mechanisms for detecting and correcting errors. The overwhelming majority of such errors are honest mistakes. When a mistake is discovered, one might think, the remedy is simply to excise it. Withdraw the paper and admit that there is currently no good reason to accept its conclusion. But if the research was responsibly done and satisfies current scientific standards, a question remains: what went wrong? Rather than merely deleting the error, it is desirable to track down its source. Was it due to sloppiness, bad luck, a mistaken assumption, or an unreliable method? Are the science's standards too lax or their applications uneven? To maximally profit from mistakes that have previously passed muster, it is advisable go back and look at the raw data, the interpretations, and the analyses to find out where the misstep occurred. The aim is not only to excise the error but to correct the assumptions, methods, practices, or standards that led to it. Scientists do not want to keep making the same sort of mistake.

Plagiarism stymies backtracking to the source of an error. The scientist who purports to have done the research did not do it. So he cannot provide information about what was done and what went wrong. Fallibilism gives rise to a demand for corrigibility. Plagiarism unduly limits corrigibility; it undermines the self-correcting character of science.

The discussion so far has shown that the dishonesty embodied in falsification, fabrication, plagiarism, and suppression of important findings is not an accidental feature of practices whose scientific objectionability lies elsewhere. Such dishonesty thwarts science's epistemic goal. The practices in question are bad science, then, precisely because they are dishonest. And their being dishonest makes them untrustworthy. An obligation to be trustworthy thus lies at the heart of science.

Disciplines differ in the sort of epistemic interdependence they display. Historians often work alone. Their books and scholarly papers are apt to be single authored. Even so, they depend on archivists, cartographers, and archeologists; and they build on what other historians have established. And history, like every discipline, is a realm of epistemic ends, whose commitments are jointly crafted and collectively endorsed because they foster the community's ends. So historians, like scientists, are beholden to the norms and practices of their intellectual community and can engage in conduct that betrays their obligations to that community.

Plagiarism can occur in any discipline, and regardless of field, can impede backtracking to discover the source of errors. Falsification and fabrication exhibit different profiles in different disciplines. In history and archeology, for example, fabrication often takes the form of forgery. A document or artifact is represented as having a provenance it is known not to have; and on the basis of that provenance, is claimed to be an exemplar of a period, source, or region to which it does not belong. Falsification occurs when, for example, archeologists seed a site, planting authentic artifacts in locations where they were not in fact found. Since archeological understanding involves an appreciation not just of what is found, but also of where it is found, the arrangement of artifacts in the seeded site yields potentially misleading information about the civilization under study.

History faces problems of selection. To generate an understanding of an era, a historian has to decide what events to include and what descriptions of those events to provide. Although there is considerable latitude, such choices are not unbounded. Epistemic irresponsibility can consist in lying about, omitting mention of, or over- or underemphasizing specific aspects of events. Holocaust-denying histories of World War II are irresponsible because they intentionally misrepresent history. Other intellectually irresponsible histories of the period are more circumspect. They issue no denials

but simply neglect to mention the Holocaust. Yet other histories distort through their choice of grain or terminology. A history that acknowledged the Holocaust and accurately reported the number of victims would still be misleading if it neglected to mention that the victims were victims of genocide. So would one that acknowledged the genocide but failed to identify its targets. Conceivably, highly specialized histories, for example, histories of the naval war in the Pacific, could justify such an omission on grounds of irrelevance, but any general history, and any history that purports to do justice to the war in Europe, could not justifiably omit mention of or downplay the importance of the 'final solution'. Because the Holocaust is so important, omitting it would inevitably skew the understanding of any history of the period that failed to give it its due.

The issue of omissions, like the issue of suppression of results, is tricky. I picked an egregious omission to make the problem vivid. But historians face delicate decisions about what to include and how to describe it. Typically there is a range of acceptable solutions. My point is not to provide a rule for settling such questions, but to emphasize that scholars face obligations not just to tell the truth, but also not to mislead. Omissions and terminological sleights of hand can mislead. How and when they do so is not always obvious. The developing norms that govern a realm of epistemic ends delineate where the lines are drawn.

As in the case of the hard sciences, in the humanities and social sciences, moral and epistemic obligations fuse. What is wrong with forging a document or seeding a site or omitting or sanitizing an important fact is not only that it is in effect a lie, but that it subverts the ends of the discipline, leading its members to accept and purvey unwarranted claims. It thus betrays the trust of the other members of the community. Insofar as the miscreant accepts the norms of the community, it also constitutes a sort of self-sabotage.

Even scrupulously done, properly documented research that exhibits no hint of overt misconduct is suspect if the investigators have a *conflict of interest* (Lo & Fields, 2009). Where such conflicts occur, agents have an extradisciplinary interest that threatens to deflect them from the discipline's goal. If, for example, a scientist has a financial stake in the outcome of a study or an ideological commitment to things turning out a certain way, he may be apt to weigh the evidence differently than an unbiased investigator would. Conflicts of interest arise at the institutional level as well. Since the pharmaceutical industry has a strong financial interest in the outcomes of

clinical trials, there is a legitimate concern that the trials it conducts or the reports it generates might be biased in favor of finding new drugs effective.[8] Commercial research in any area has a higher-order conflict of interest. The epistemic imperative requires that all members of a community of inquiry have free and equal access to the evidence for a claim. Any member of the community should be able to challenge a finding. But commercial enterprises have trade secrets. To protect those secrets, they do not make the details of their research publicly available. Whether the claims they make about their products are epistemically acceptable is open to question. Privatization of inquiry comes at considerable epistemic cost.

A conflict of interest exists whenever there is a danger of undue influence from an extradisciplinary interest. It does not require that anyone actually be unduly influenced. So the existence of a conflict of interest is compatible with complete absence of wrongdoing on anyone's part. An investigator with a conflict of interest may in fact weigh the evidence exactly as an unbiased investigator would. Even so, the possibility of bias undermines his credibility.

This raises the question of why conflicts of interest are objectionable. One might think that it would be enough to bar the existence of undue influence. Why should the mere risk of undue influence be a problem? Here again the answer turns on trustworthiness. Undue influence by extradisciplinary interests corrupts the discipline and deflects it from its epistemic goal. If a scientist does not know whether another scientist was unduly influenced, she does not know whether his result is trustworthy. And if she harbors doubts about its trustworthiness, she will not and should not trust it. The risk that a scientist's or institution's extrascientific interest deflects from the goal provides grounds for doubt. Consciously or unconsciously, investigators may be influenced by factors that do not withstand scientific scrutiny. Given that risk, others ought not be confident that their results satisfy the standards of the scientific community. The mere risk of undue influence by extrascientific factors thus undermines confidence. It is enough to disincline others to accept assurances that the work is sound.

Epistemic Responsibility

Scientific misconduct thwarts the epistemic goal because it undermines collaboration or corrigibility. This suggests that besides abiding by the

prohibition against specific forms of misconduct, members of a community of inquiry bear additional obligations to one another. Overt misconduct of the sort examined above is not the only threat to collaboration and corrigibility.

Science is subject to epistemic requirements that constrain experimental design, data gathering, analysis and assessment. Hypotheses should be formulated sharply enough to admit of empirical testing. Experiments should be designed to yield data capable of confirming or disconfirming the hypothesis under investigation. Irrelevant and obviously misleading evidence should be discredited. Data should be painstakingly gathered and carefully recorded in a timely fashion. They should be displayed in such a way that their significance and their limitations are manifest. Analyses should be rigorous, relevant, complete, and apt. Assessments should be judicious. Research reports should be grounded in a comprehensive, fairminded, current understanding of the topic and the methods for investigating it. They should acknowledge the assumptions made, the limitations of the approach being used, and the potential threats to validity. Previously published relevant work should be cited. Such requirements are familiar; their satisfaction yields conclusions there is reason to accept.

So far, however, there is nothing overtly moral about the requirements or the benefits of satisfying them. The items on my list seem to be purely epistemological requirements. The moral element enters when we realize that in publishing a result, a scientist issues an open invitation to the community to rely on it and assures its members that they can do so with confidence. Results that justifiably inspire confidence must be epistemically responsible. In addition to intellectual honesty, competence and conscientiousness are mandatory. In publishing a result, a scientist implicates not only that it was arrived at honestly, but also that it satisfies the epistemic standards of the community. That is, she implicates that the work is competently done. Because scientists must depend on one another, the epistemic requirements are moral requirements; if scientists fail to satisfy these epistemic requirements, they fail in their moral obligations to the community.

Carelessness, fickleness, obtuseness, and bias, then, are not just epistemological failings; in the context of inquiry, they are also moral failings. Results infected by such faults are honest mistakes. Confirmation bias leads an investigator to overlook or underrate evidence that tells against a hypothesis she favors. She is not being dishonest; she genuinely believes

that her evidence supports her conclusion. But inasmuch as she can, by being more rigorous, avoid confirmation bias, her mistake is a culpable error. The same holds for carelessly taking or recording data, overstating one's results or understating their limitations, and so on. By inviting her colleagues to accept a conclusion that by their own lights they ought not accept, a scientist does them wrong.

Fickleness is different. An intellectually fickle investigator flits from one set of commitments to another without due regard for the underlying issues. One day she is a gradualist about evolution; the next day she accepts punctuated equilibrium. Her flaw is not that she changes her mind; it is that she does so cavalierly. She changes her mind too easily. Her pronouncements are unstable. Since they are so easily abandoned, her compatriots would be unwise to accept them on her say-so. The gullible investigator has a similar flaw. He believes whatever he hears, at least so long as the information is remotely credible. Perhaps he would not go so far as to believe that space aliens landed in Detroit, but he might accept the contention that because the gene for blondness is recessive, blondes are going extinct. If he spreads the word about the imminent extinction of blondes, he is intellectually irresponsible. For the information about how recessiveness operates is readily available and ought not in this context be ignored. Insofar as these flaws could be avoided by being more scrupulous in gathering evidence and justifying inferences, the agents are culpable. They are willfully ignorant. If they present their claims as scientifically backed, they do their fellows wrong.

In agreeing to work together, people engender obligations to one another—obligations whose content is determined by the objective they are jointly pursuing. Ceteris paribus, they owe it to one another to discharge those obligations. In inquiry, the goal is understanding. So the obligations are in the first instance epistemic; they pertain to generating and sustaining the understanding that inquiry seeks. But because inquiry is a joint epistemic venture, the epistemic obligations are also moral obligations. Investigators owe it to one another to satisfy the epistemic requirements, and to make it manifest that they do so.

This pattern is not peculiar to science, or even to inquiry. Members of other collectives, such as athletic teams and orchestras, are under similar obligations to one another. In the first instance, the obligations of a quarterback are athletic; those of a bassoonist, aesthetic. But the quarterback owes it to his teammates to discharge his athletic obligations well; he is

something worse than a poor athlete if he carelessly or negligently lets down his side. And the bassoonist owes it to her fellow musicians to perform well; she is something worse than a poor musician if she neglects to learn her part or is cavalier about playing in tune. The commitments team members make to one another are akin to implicit promises; because the common knowledge of those commitments creates a basis for legitimate expectations, one does the others wrong if one does not abide by them.

Arguably, in inquiry there is a significant difference in scope. The members of a research team, like the members of a football team or an orchestra, consist of a limited number of identifiable individuals who have well-defined obligations to other people they know. But in publishing findings, a scientist effectively invites the entire scientific community to join her team; since she provides assurance that her findings satisfy their epistemic standards, she is under an obligation to proffer findings that they all can be confident of. Indeed, to the extent that the rest of the population depends on the findings of science, in publishing her results, a scientist offers her assurance that everyone can have confidence in them. The same holds for scholars in other disciplines. To publish a conclusion is to make that conclusion public. It is to invite readers or auditors to accept the conclusion with the assurance that it meets the relevant standards of acceptability.

A scientific community consists of practitioners who share a discipline. They are bound together, and think of themselves as bound together, by a network of commitments that govern their professional lives. They are joint legislating members of a realm of epistemic ends. They justify particular verdicts by reference to the rules and standards of their practices, and justify those practices by reference to the ends of their science. In performing this dual justificatory function, a scientific community is largely self-legislating. As I argued above, it makes the rules, constitutes the norms, and sets the standards that constrain and guide its activities. It determines what counts as an experiment, a confirming instance, a statistically significant result, a representative sample. Its norms and standards are not arbitrary. They are designed, and continually redesigned, with an end in view—namely developing and confirming a systematic, explanatory account of the phenomena, an account that is grounded in empirical evidence and that underwrites nontrivial inferences, arguments, and perhaps actions regarding those phenomena. Acceptability derives from satisfying the interlocking network of standards science sets for itself, standards that are grounded in an evolving,

collective understanding of the scientific enterprise and of the difficulties to be encountered in pursuing its goals.

Although the standards apply throughout scientific practice, their general, impersonal character comes to the fore with publication. In publishing, a scientist undertakes an obligation to be worthy of her community's trust. A critical question is what standards a work has to meet to vindicate the open invitation. Ideally, perhaps, we'd like to demand that the research report's content be accurate: the claims that purport to be true are in fact true; those that purport to be felicitous falsehoods are in fact felicitous. But this is both too much and too little to ask. Science is fallible; it has no way to guarantee that findings are accurate. Moreover, as we have seen, mere accuracy is not enough. The scientist needs to convey her assurance that her research, whether the result is accurate or not, satisfies current, reflectively endorsed standards. Because those standards embody the best way the discipline has devised to generate and validate understanding in its field, results that satisfy those standards are currently acceptable.[9]

Assurance

To understand what a scientist proffers in publishing a paper, we need to distinguish between warrant and the assurance of warrant. Warrant is a first-order relation between evidence and conclusion. A conclusion is warranted when the evidence renders it sufficiently probable (as always, measured by current standards). A conclusion could be warranted even if there was reason to doubt that it was. This might, for example, be the case if the connection between the evidence and the conclusion obtains but is not obvious. An assurance of warrant is a second-order assessment of the strength of the connection between the evidence and the conclusion. Warrant, then, is a function of the force of evidence; the assurance of warrant, a function of the weight of evidence. The assurance of warrant is what grounds confidence in first-order assessments (Adler, 2002, 251–252).

To appreciate the difference, consider the following example. Bob takes a quarter from his pocket, looks it over and sees nothing odd about it. Concluding that it is a fair coin, he judges that the probability of its coming up heads when flipped is 0.5. In light of the prior probability that an arbitrarily chosen coin is fair and his cursory inspection of the coin, his conclusion is warranted. Now suppose he flips the coin 1,000 times and records the

result of each flip. Not surprisingly, the coin comes up heads just about half the time. So he again concludes that the probability of its coming up heads is 0.5. Should we consider his epistemic situation unchanged? Should we think that he is epistemically no better off for having run the test? The test supplied no reason to alter his original probability assignment. That assignment was already warranted. There is little, if any, evident epistemic gain in his first-order assessment. Still, he is plainly better off. What he gains is not warrant, but a right to be confident. After the test, he is in a position to be more confident in the probability he assigns (Adler, 2002). The test result provides assurance that his first-order assignment is correct.

It is a principle of jurisprudence that justice not only must be done, it must be seen to be done. Something similar holds in science. Results not only must be warranted, they must be seen to be warranted. That requires that they be shown to be warranted. For scientists to be in a position to confidently rely on an investigator's work, it must be manifest that the work satisfies community standards of acceptability. So rather than just issuing a press release announcing their findings, scientists publish a peer-reviewed paper that spells out (admittedly, rather tersely) their background assumptions, research design, measurements, analysis, and results. Such papers must be public and transparent. Publicity demands that results be available to the entire community; transparency demands that the reports provide sufficient detail that other members of the community can recognize why they ought to be accepted. They exemplify that the work they report satisfies the relevant standards.

I have been speaking as though the reason for making it manifest that the standards have been met is entirely other-directed. An individual scientist owes it to her colleagues to proffer only results that satisfy the standards they reflectively endorse. This is part of the story. But the satisfaction of those standards also gives the scientist herself reason for confidence in her findings. Because the standards function as a stay against carelessness, bias, and wishful thinking, they provide her with assurance that her work is worthy of her own reflective endorsement.

The scientific community sets standards whose satisfaction not only warrants conclusions, but exemplifies that they are warranted. Some requirements on acceptability—such as those concerning the size and constitution of the evidence class, the standards of statistical significance and power, the proper sort of analysis—bear directly on warrant. These are first-order

requirements. Others, by making it manifest that results hold regardless of who generated them or why, pertain to the assurance of warrant. They mandate that findings be intersubjectively observable; experiments, reproducible; data, unambiguous; research reports, peer reviewed. If an item is intersubjectively observable, it does not matter who in particular observed it; any competent scientist could have done so.[10] If an experiment is reproducible, the acceptability of its result does not turn on the character or intellectual endowments of the scientist who performed it; had any competent scientist done it, the result would have been the same. And if data are unambiguous, the identity of the interpreter is immaterial; all competent scientists would interpret the data in the same way.

One might think that these requirements bear only on warrant, not on assurances of warrant. If so, the fact that an experiment is reproducible affords only first-order evidence that its result is correct. Granted, if you ran the experiment a second time, you would get more first-order evidence for its conclusion. Rather than 64 data points, you would have 128. So the probability that the conclusion is correct would increase, at least slightly. But if the original experiment yielded a publishable result, it already provided enough evidence to warrant its conclusion. The additional evidence provided by the second run, although not unwelcome, would by no means be mandatory. It would not do much to increase the warrant of the already warranted conclusion. And if the original experiment can yield enough evidence to warrant the result, no more first-order evidence is needed. What does the reproducibility add?

Reproducibility obviates the need to take matters on faith. Doubters can go back and run the experiment for themselves. Similar points hold for the requirements that data be unambiguous and that phenomena be intersubjectively observable. When it makes no difference who generated a result, when any competent scientist could have done so, the personal foibles or intellectual idiosyncrasies of the investigator are irrelevant to the result's acceptability. This indifference makes science impartial.

One might worry that by rejecting deliverances that are not intersubjectively observable, experiments that are not reproducible, and interpretations that are idiosyncratic, science throws out potentially valuable information. No doubt it does. But even though an irreproducible result, a quirky interpretation, or idiosyncratic observation may in fact be correct, one would be unwise to count on it and irresponsible to give others to

count on it. Similarly for findings shrouded in secrecy. If a scientist cannot or will not publicly document what she did to achieve her result, she cannot responsibly invite others to accept her assurance that they can count on it. Without such documentation, they have no reason to believe that it satisfies standards they collectively endorse.

Peer Review

I have claimed that because of their interdependence, members of the scientific community have moral obligations to one another. One might agree that they are accountable to one another but still doubt that there is anything especially moral about that accountability. Given that their work is subject to peer review, it might seem that an individual scientist's motivation for conforming to community standards is like a student's motivation for not cheating when a proctor is looking over her shoulder or a driver's for not speeding when a traffic cop is pointing a radar gun his way. The chances of getting caught are high enough and the disutility of getting caught great enough that enlightened self-interest provides sufficient reason to obey the rules. If so, there is no need to invoke distinctively moral obligations. But science is not Big Brother. It does not monitor every experiment to ensure that its standards are met. Although the scientific community collectively sets standards for how investigations are to be carried out, how evidence should be gathered, presented, and analyzed, and how research should be reported, peer review can do relatively little to enforce the standards.

To be sure, peer review provides a measure of quality control; but its scope and therefore its powers are limited. Its immediate objects are documents—research reports and grant proposals. Because reviewers are knowledgeable about their field, peer review is a good way to assess the originality, importance, and epistemic adequacy of the research those documents describe. It can reveal that, for example, an experiment did not adequately control for factors it ought to have controlled for or measure magnitudes it ought to have measured, since the authors spell out what is, or is to be, controlled for and measured. It can pinpoint gaps in presuppositions, or weaknesses in experimental design, since the authors describe their presuppositions and experimental design. It can reveal that the paper overstates or understates its findings, since the authors state exactly what they take the findings to be. It can tell whether the authors have taken appropriate note of previous

relevant work.[11] In doing all these things, peer review assumes that the research is accurately reflected in the documents. But peer review is typically ill-equipped to ensure that the research reported was actually done, that the results listed were actually found, or that the alleged authors actually wrote the paper. The most it can do to uncover fabrication or falsification is to assess plausibility, either of a single paper or of a body of work.[12] The most it can do to uncover plagiarism is to rely on reviewers who know the literature well enough to recognize the original source. For obvious reasons, peer review is in no position to assess suppressed results. My point is not to disparage peer review. It is to note that the function of peer review is neither to check for nor to preempt the need for scientific integrity; it is rather to assess the quality of scientific work in a context that presupposes scientific integrity.

Fostering Integrity

If the requirements of science were mandates imposed from without, the limitations on peer review would open the door to a free rider problem. A scientist might recognize that although the moral/epistemic requirements should generally be satisfied, science would not totter if a few investigators cut corners, any more than the subway system is jeopardized if a few riders fail to pay their fares. So, she might think, in neither case does the threat to the institution provide a reason why she in particular should conform to the rules, provided that most people do. This would leave room for it to be rational to defect. But the situation is different for members of a realm of epistemic ends. Because they make the rules that bind them, they cannot violate those rules without violating their own principles. The problem is not just that a miscreant scientist could not justify her behavior to other members of the community. It is that, having reflectively endorsed the rules she is flouting, she could not justify her behavior to herself. This does not ensure that there will be no free riders, but it does ensure that free riding by legislating members of a realm of ends is not rational. In a realm of ends, a free rider's behavior is inconsistent with her own principles.

Publicity and transparency foster scientific integrity. Because it is common knowledge that only work that at least ostensibly satisfies the scientific community's requirements will be accorded scientific standing, practitioners have a strong incentive to do only work that at least ostensibly satisfies

the requirements. Typically the easiest and most straightforward way to at least ostensibly satisfy the requirements is to actually satisfy them. So scientists have an incentive to produce work that satisfies the requirements. This hardly handles every case. But it may go some way toward explaining why scientific misconduct is not rampant.

Another incentive is this: Scientists recognize that the requirements are not arbitrary. They are instrumentally valuable because their satisfaction is conducive to realizing the epistemic goal. Inasmuch as scientists want to realize the goal, they have reason to employ the requisite means. As Williams notes, Watson and Crick wanted to discover the structure of DNA; they did not want merely to be credited with the discovery. No doubt they were personally ambitious. But "their goal [was] fame, above all fame and prestige in the scientific community, and that [would] come from the recognition that they have done good science" (Williams, 2002, 142). To realize their ambition, they had to satisfy biology's standards.

An even deeper incentive lies in the recognition that satisfying the requirements is not merely conducive of the goal; it is at least partially constitutive of the goal. Being comprehensive, impartial, evidentially grounded, and reflectively endorsed are properties of the sort of understanding that science seeks. Such an understanding is subject to, and holds up under, critical scrutiny, where the grounds for criticism and bases for assessment have been tempered through a succession of increasingly refined and rigorous processes of self-correction. Moreover, since that understanding is open to extension, correction, and refinement, a reflectively endorsed finding is not considered fixed or final, but is an acceptable springboard for further inquiry. Scientists have an incentive to satisfy the requirements, then, because the satisfaction of those standards is built into scientific understanding. That is, in Kantian terms, they have an incentive to conduct their research in ways that respect, rather than merely conform to, the requirements.

Nevertheless, the mutual trust that I have argued sustains the scientific community may seem fragile, and relying on it may seem naive. Even if, qua scientists, they have reason to uphold the standards, scientists wear other hats as well. They may be entrepreneurs with a financial stake in the outcome of their experiments; they may be parents who need to feed their families; they may be ambitious neurotics whose sense of self depends on the (earned or unearned) respect of their peers. One might wonder, then,

whether the incentives to cut corners outweigh incentives to behave with integrity.

If talented, amoral scientists emerged full-grown from the head of Zeus, with all the factual knowledge and skill needed to ply their trade, the incentives I identified might well be too weak. Normally, scientists do not attempt to replicate one another's results; and when they do, an irreplicable result is apt to be construed as an honest mistake. So although there is enormous disutility in being found guilty of scientific misconduct, if his research reports are plausible, a culpable scientist can reasonably doubt that he will get caught. The worst he need realistically fear is that he will be found to be mistaken. An adroit, amoral scientist might readily recognize this and, being devoid of scruples, consider the risk worth taking.

Even if such a scientist granted that in general the best means to achieving the goal is to satisfy the requirements, this would not obviously provide him with an incentive to do so if he thought highly enough of his own intuitions. He might rationalize his falsifying or fabricating on the ground that although in most cases the requirements should be satisfied, his particular case constituted an exception. He has a good enough feel for the subject that he can just tell how the experiments would come out. I conceded earlier that if truth is the end and the scientific method a mere means, it may well be that a scientist with good intuitions (or luck) could achieve the end without employing the standard means. An amoral scientist may believe that the only disincentive to cutting corners is the possibility that his conclusion will be found to be false. But, he might think, his conclusion probably is true, and if it is found to be false, it is apt to be consigned to the realm of honest mistakes. Again, it might seem, the personal risk he takes is slight.

Science Education

Luckily scientists do not emerge full-grown from the head of Zeus. People become scientists by studying science. And the moral values I have discussed are instilled in the course of their education. Even if a brilliant, amoral senior scientist could readily contrive a plausible, fraudulent result, doing so is not easy. For a beginning student, the best way—very likely the only way—to ostensibly satisfy the requirements is to actually satisfy them. A novice is apt to get caught if he fakes, fabricates or plagiarizes his work. He

is also apt to get caught if he is careless, inattentive, or lax. Learning to do scientific research is learning to do honest, truthful, careful research. It is not learning to do research, with honesty, truthfulness, and conscientiousness tacked on as afterthoughts. So, even if he starts out as amoral as his Zeus-born senior counterpart, the novice has a strong, purely self-interested incentive to conform his behavior to the requirements. He is simply unable to cut corners and produce plausible results. What he learns in learning to be a scientist is to satisfy the requirements. He does not learn that it is normally a good idea to conduct the experiments one purports to have conducted. He learns that scientific research consists in conducting those experiments.

As his education progresses he develops the habit of conforming to the requirements; doing so becomes second nature to him. In this respect, his moral education as a scientist is Aristotelian. He develops scientific integrity by behaving with integrity, whatever his initial motives for doing so. As he learns more about the subject he is studying, he gains respect for the requirements. He comes to recognize that their satisfaction is conducive to the epistemic goal, and eventually, that it is in part constitutive of the epistemic goal. That is, he learns what scientific understanding is. So the reason that the trustworthiness required of scientists is not excessively fragile and that the expectation that scientists will be on the whole trustworthy is not naive, is that the grounds for it are so deeply embedded in science that they cannot help but be inculcated in scientific education.

This suggests that the education future scientists standardly receive already fosters scientific integrity. It invites the idea that nothing more need be done. Sadly, that is too optimistic. Some scientists are guilty of misconduct. Some students have no clear sense of where the boundaries are drawn. How can science education foster integrity?

When something has become second nature, the agent automatically and unthinkingly does it. So most scientists probably do not even consider flouting the requirements. This is both good and bad. It is good in that they simply do the right thing; it is bad in that they may fail ever to consider why it is the right thing to do. If they do not know, for example, why data must be recorded in a timely fashion, or what exactly is problematic about conflicts of interest, they may be ill equipped to withstand temptation, to adjudicate competing claims when all the desiderata cannot be

simultaneously satisfied, or to recognize what their obligations are when they detect misconduct. By highlighting aspects of a practice that had been tacit, scientists put themselves in a position to articulate the relevant factors, discover their basis and their strength, subject them to scrutiny, and consider whether they are worthy of reflective endorsement. And they put themselves in a position to deepen science education by sensitizing students to often unrecognized aspects of scientific practice that are critical to its success. So, in addition to forestalling explicit scientific misconduct, the NSF requirement may prompt a reorientation to the practice of scientific inquiry that can augment practitioners' understanding of the discipline and its results.

Conclusion

Throughout this chapter, I have emphasized the obligations that members of the scientific community bear to one another, and emphasized the ways these obligations infuse the epistemic goals of science. Because scientists need to trust one another, they have devised institutional arrangements to secure the bases of trust. In publishing a paper, a scientist issues an open invitation to the community to accept her findings and use them as a springboard for further investigation. She gives her colleagues her assurance that the standards they collectively endorse have been satisfied.

My discussion may suggest that science is insular. But the fruits of science are used for more than further research, and they are used by nonscientists as well as scientists. We take medicines, drive cars, eat genetically engineered vegetables. We sit on panels and juries where we hear scientists testify as expert witnesses (see Brewer, 1998). In doing so, we rely on the deliverances of science. The open invitation the scientist issues thus is wide open; it is issued to the world at large, not just to the scientific community. In publishing a finding, speaking at a conference, or testifying in court, she conveys her assurance that anyone—layperson or expert—can rely on her results. The point is a delicate one. She does not assure that the results are true, only that they are scientifically well founded.

As I have indicated, these points generalize. Intellectual integrity is required in every discipline. Investigators are expected to be scrupulous and conscientious. They should adhere to the methods, canons of evidence,

and norms of acceptability reflectively endorsed by their discipline; or they should publicly articulate and justify the reasons for challenging them. The methods, canons, norms, and epistemically acceptable reasons for challenging them can vary from one discipline to another. But their role in the disciplines is constant. Although the standards to be satisfied are set by the community of inquiry, the moral obligation to behave with integrity is utterly general.

7   Objectivity: A Bulwark against Bias

Despite what I've said thus far, worries about objectivity might remain.[1] To address these worries, we need to consider what objectivity is and why it is valuable. On one interpretation, objective judgments are nonperspectival representations that carve the mind-independent world at its mind-independent joints. Clearly the epistemological position I have been developing cannot claim that the commitments it countenances display that sort of objectivity. This is just as well. I consider that sort of objectivity a chimera. It is not something we can achieve (or, I suspect, even make sense of). Nor is it something we need to achieve to serve our epistemic purposes. But it does not follow that objectivity is a chimera. I will argue that my position countenances a robust form of objectivity that is procedural, perspectival, and either impersonal or impartial.

Utter Objectivity

Some think that for a contention to be objective is for it to express, reflect, or represent what is there, regardless of what anyone or any group thinks or does. That is, it is to reflect what Williams (1978) calls the absolute conception of reality. Such objectivity is, Nagel (1986) maintains, what the view from nowhere delivers. Let us call contentions that satisfy this standard *utterly objective.* Products of and statements about attitudes, institutions, and artifacts reflect human interests and goals. They involve views from somewhere. They are excluded from the utterly objective realm, regardless of their truth-value or warrant. But, Williams, Nagel, and their followers maintain, natural science is different. It discovers and explains what is, in Williams's term, *there anyway*, most of which would have been there even if

rational organisms had never evolved. It and its findings are utterly objective. Utter objectivity is a consequentialist notion. What makes a procedure utterly objective is that its products represent aspects of the way the world is anyway.

Science has been stunningly effective at providing an understanding of nature. No other approach even comes close. Much of nature is as it is regardless of what, if anything, anyone thinks about it. Perhaps, then, it is reasonable to think that science represents the way the world is anyway, that at least in the ideal end of inquiry it will (would?) do so completely and without bias or distortion, and that current science approximates the ideal. If so, scientific representations are utterly objective: what they say or otherwise represent does not depend essentially on users or potential users with distinctive interests and points of view. The semantics of such a representation is thus independent of and logically prior to its pragmatics.

A *representation* is a putatively denoting symbol. Among the familiar sorts of representation are verbal descriptions, mathematical equations, pictures, diagrams, and maps. Such symbols need not actually denote; their semantic character is determined by their being the sort of symbol that would, if the world obliged, denote. 'Phlogiston' is as much a denoting symbol as 'oxygen' is. In one sense, any representation depends on users or potential users with distinctive interests and points of view. Something would not be a representation unless it was used or could be used to represent, and it would not be a representation of a particular thing unless it was used or could be used to represent that very thing. There would be no representation of a particular item if no one had interests that would be served by producing one. The sentence 'Tibet is mountainous' would not represent the topography of a particular geographical region if speakers did not call a particular land area 'Tibet' and did not call a particular sort of terrain 'mountainous'. But the choice to correlate particular labels with particular referents is effectively arbitrary. The very same information would have been conveyed by the sentence 'Murble is morbish' if 'Murble' had been the English word for Tibet and 'morbish' the English word for mountainous. The user-independence required for utter objectivity is independence except for the making of such arbitrary choices. The contention that scientific representations are utterly objective is a contention that, beyond arbitrary choices, they exhibit no dependence on users or potential users.

Van Fraassen (2008) poses a challenge to this picture: if scientific representations—models, equations, maps, graphs, and the like—are to

perform their scientific functions, they cannot be utterly objective. Cartwright (1983) makes similar arguments about particular scientific models. Like van Fraassen, she regards this feature not as a defect of models but as an insight into the sort of epistemic access they supply. Van Fraassen's argument runs deeper. It is not just that this or that model, or models in this or that science, cannot be utterly objective. Rather, in order to afford the sort of understanding of nature that it does, science cannot consist exclusively or predominantly of utterly objective representations. Science as we know it must deploy perspectival representations. It ineliminably involves views from somewhere. For science is an epistemic practice. To make this out, we need to consider the nature of perspective.

Getting Perspective

Linear perspective is a miracle of Western art. First described by Alberti (2011), it is a method for representing spatial depth on a flat surface. Lines orthogonal to the picture plane converge at the vanishing point, and the depicted sizes of objects are proportional to their actual size and to their ostensible distance from the viewer. The pictorial effectiveness of linear perspective in works like Raphael's *School of Athens* is obviously of great aesthetic interest. Van Fraassen maintains that linear perspective should be of equally great interest to philosophy of science. With the development of projective geometry, it became possible to prove that properly drawn linear perspective representations are rigorous geometric projections. Because cross-ratios are invariant across changes in orientation and origin, drawings in perspective convey objective information about constancies in the appearances items present from different points of view. They are not utterly objective, though, for they still depend on points of view. Perspective drawings are indexical: they represent how things look from *here* (for some value of 'here'). Their indexicality does not, however, make them subjective. Cross-ratios are determinate matters of fact about projective relations, not mere matters of opinion. Van Fraassen concludes that a linear perspective drawing of an actual scene is a measurement, a mapping.

There are inherent limitations on what any single representation can represent. Every representation is a product of selection. The representer has to choose what to represent, what aspects of the represented item to capture in the representation, and what level of detail to represent.

Despite the fact that these are matters of choice, when it comes to scientific representations, it is possible to make mistakes. Simpson's paradox arises because fine-grained statistical regularities are obscured at a coarser grain. Although the pattern of graduate admissions across the university suggested that Berkeley discriminates against women, the admissions patterns of each graduate department indicated the contrary. Settling the issue requires knowing which statistics to use (Cartwright, 1983, 37). To use a coarse-grained representation to answer questions that require a finer grain is to make a mistake, even if all the features displayed in the coarse-grained representation actually obtain. Likewise, it is an error to use a fine grain when a coarse grain is needed.

Because they depend on a specific origin and orientation, perspective drawings have further limitations. From a given vantage point, one object may *occlude* another. If the horse is behind the barn from the picture's point of view, then the picture's representing the barn precludes its representing the horse. Nor is this always a matter of big things occluding little ones. If the mouse is close to the picture plane and the barn is far from it, the perspectival representation of the mouse may occlude the barn. A representation is *implicitly noncommittal* with respect to a property if it makes no commitment as to whether the represented object has that property. A stick figure is implicitly noncommittal with respect to its subject's girth. It simply does not go into the matter. A representation is *explicitly noncommittal* with respect to a given property if its representing the having of one feature precludes its taking a stand on the having or lacking of another. A representation of a man wearing a hat is explicitly noncommittal with respect to whether he is bald, because representing him as hatted makes it impossible for the picture to commit itself on the question of his baldness. To be sure, the horse, the girth, and the hairline could be represented in a perspective drawing. But to represent the horse requires removing the barn or changing the perspective. To represent the girth requires a more rounded figure. To be committal with respect to the man's baldness requires omitting the hat.

Indexicality, occlusion, and noncommitment do not either severally or jointly entail that there are things that cannot in principle be represented perspectivally. They entail limitations on what a single perspectival representation can represent. But because of these limitations, there cannot be a single, comprehensive perspectival representation that represents

everything from a single point of view. The 'God's eye view' cannot be a point of view.

The pictures painted by Renaissance artists are, of course, of (putatively) visible objects, and the space they (purport to) represent is (putatively) physical space.[2] These restrictions are philosophically incidental. Early perspective drawings had only one vanishing point. But it is straightforward to create works with multiple vanishing points. *Le Déjuner sur l'Herbe*, for example, is in two-point perspective. Thus the number of vanishing points is optional. Features at higher levels of abstraction than size, shape, and distance can be represented perspectively, if the requisite structural relations are preserved. So the restriction to visible features is optional as well. The space need not even be physical.

A logical space is a multidimensional array of possibilities open to the items that occupy the space. To locate an item in a logical space is to determine which of the possibilities defined by that space it realizes. To represent an item in a logical space is to represent it as having a particular position in the array of possibilities the space marks out. Representations in a logical space, like representations in a physical space, can be perspectival. They can show how occupants of that space appear from a certain vantage point. And they can do so with no loss of rigor. In what follows, I take the term 'perspectival' to refer to any representation that represents how things appear from a particular point of view.

With the restrictions to physical space and visible features lifted, it is evident that science could avail itself of perspectival representations. It could generate and deploy a host of perspectival drawings, diagrams, scale models, and maps. There is plenty of evidence that it does so. Still, one might think, only overtly pictorial or diagrammatic representations—the sketch of a harmonic oscillator, the diagram of the double helix, the tinker-toy model of the protein—could be perspectival. Many scientific representations are mathematical models—systems of equations. It may be hard to imagine how any extrapolation from linear perspective could characterize to them. This may suggest that the visual models that admit of a perspectival interpretation are mere heuristics. They are visual aids that help us imagine the phenomena, but not essential elements of science. The mathematical laws and models—the equations—are the true scientific representations, and they are utterly objective.

   Things are not so simple. Analytic geometry provides the resources to interpret geometry algebraically. Geometric truths are provably equivalent to algebraic ones. Contemporary Cartesian geometry demonstrates that equivalence goes both ways (van Fraassen, 2008, 41). Algebraic truths are mathematically equivalent to geometric ones. If we define an appropriate space, pretty much everything we can characterize mathematically can be spatialized. Indeed, we do not even need to do the reduction. The equivalence shows that the equations themselves, whatever their ostensible form, are construable as spatial representations. To be sure, the space need not be three-dimensional physical space. All that is needed is that there be an n-dimensional space of alternatives that embeds the mathematical model. Nor, of course, does this show that the spatialized representations are perspectival. But the mere fact that they are presented as equations provides no reason construe mathematical models as nonperspectival. In principle, then, equations are as capable as pictures and diagrams of bearing a perspectival interpretation.

The View from Nowhere

Plainly, not all pictorial or diagrammatic representations are perspectival, even in the extended sense in which I am using the term. The Cartesian coordinate system provides a familiar and elegant framework for non-perspectival representations. Cartesian representations have no vanishing points; parallel lines in every dimension remain parallel. Although Cartesian representations locate represented objects by reference to an origin and a direction, the choice of origin and direction are arbitrary. "The chosen frames of reference, the co-ordinate systems, are inessential to the geometry taken in and by itself" (van Fraassen, 2008, 69). Such representations are not indexical. One need not locate oneself in the space of the representation to understand what and how it represents. Nor are the representational powers of Cartesian systems limited by occlusion or explicit noncommitment. Rousseau's picture of a tiger "is explicitly non-committal about the [number of the] tiger's stripes, because it represents the tiger as obscured by leaves, and this precludes it from representing all the tiger's stripes" (Lopes, 1996, 118–119). A Cartesian representation, though it would not look like a tiger, could easily circumvent the difficulty. Let the x-axis represent stripes, and the y-axis represent leaves in the environment. The only available positions

are integer values along either axis. Then where a stripe is not obscured by leaves, $y = 0$; where it is overlapped by leaves $y > 0$, with the value of y indicating the number of leaves overlapping a particular stripe. Where there are leaves but no stripes, $x = 0$. Such a graph represents the overlap of stripes by leaves, but nothing is occluded. It is expressly committed to the number of stripes on the tiger, regardless of their relation to leaves.

Why shouldn't we think of the logical spaces of scientific representations as Cartesian spaces? Then, aside from the few scientific representations that are expressly perspectival, we could construe scientific representations as utterly objective. The mathematical models could be spatialized—we could represent them in a Cartesian coordinate system—but the results would still be utterly objective. There would be no reason to deny that they represent the way the world is anyway.

It is undeniable that science can generate Cartesian representations. It is undeniable that such representations are as likely to be accurate and adequate to their subject matters as any other representations we might produce. So perhaps we should consider them utterly objective. In the limit, if not now, they will embody aperspectival information about the way the world is anyway. If science's function is merely to mirror nature, the claims of ideal science could bear an interpretation under which they reveal how things are anyway. But the fact that science *could* generate aperspectival representations gives us no reason to think that it *does* or that it *should*. As van Fraassen argues, such a construal does not do justice to science. Indeed, under such a construal we have no reason to trust the findings of science. Accuracy and adequacy to the subject matter are not enough.

A View from Somewhere

Science is not something that just happens to us; it is something we do. To do it, we need to use scientific representations. Use is a pragmatic matter; and to make use of anything an agent needs to properly locate herself with respect to it. This is evident in the relation between science and technology. To develop an alloy that resists metal fatigue, a metallurgist has to be able to recognize signs of metal fatigue. To do that, she must adopt a perspective from which she can discern metal fatigue if it is present. This will not, of course, be a matter of merely looking at distressed samples. It will involve measurements, many of them made with technologically complex

measuring devices. It will involve subjecting samples to stress tests, many of them technologically mediated. The measurements and test results will exemplify features that put her in a position to say, 'This is how metal fatigue (or its absence) looks from here'. Assuming she is scrupulous, she will run series of tests that yield different perspectives on the phenomena. She will check one appearance against another, drawing her conclusion from what she observes from a variety of points of view. She does not, because she cannot, solve her problem using the view from nowhere. For she has to know how metal fatigue is manifested.

This point must be conceded; but maybe it is just a point about technology, not about pure science. It is to be expected that when we put something to use we need to adopt a perspective. But, one might think, science is not technology. Pure science may be utterly objective even though deliverances of technology cannot be.

As van Fraassen points out, however, scientific results are established by testing and experimentation, and they are in principle always open to further testing. Testing is as indexical and perspectival as technological deliverances. If a scientist wants to ascertain whether semiconductors operate in a magnetic field, she must run a series of experiments and take measurements. To design the proper experiments and take the proper measurements, she must adopt a perspective on the phenomena—she must figure out how the phenomena would present themselves under various circumstances, how they would look from various points of view. Indeed, the distinction between testing and technology is spurious, since the testing devices are products of technology, and every technological application is in principle a test. Even as simple a measuring device as a tape measure owes its status to considered judgments about the appearances things present—for example, that the items a tape measure measures are not affected by the fact that they are being measured. Hence the fact that something that measures 32 cm is a reliable indication that it is ≈32 cm long.

Van Fraassen's argument, then, is this: science affords epistemic access to nature. That access is achieved through experimentation and measurement. The only way for an observer to perform the experiments or make the measurements is to locate herself in the logical space of the phenomena. Thus, her stance is indexical and perspectival. She has to think, 'If φ is going on, these are the appearances it will present under these test conditions'. That is, if φ is going on, this is how it will look from here. This locates the

observer in the logical space she is evaluating; it affords a perspective on how things should look from here (for some value of 'here').

Although we could take the representations produced by science to be utterly objective, doing so would divorce them from the empirical methods that generate and confirm them. In that case, however, science would provide no reason to believe or accept them. To interpret the representations as utterly objective is to cease to consider them scientific. For science to do its epistemic job, it requires measurement. Measurement is always indexical and perspectival. Hence for science to do its epistemic job, it must involve indexical, perspectival representations.

To be testable, science must use perspectival representations. Those representations are objective in that they contain information that is invariant across representations of the same object. They are testable in that multiple representations of the same objects from the same perspective yield equivalent information, and in that that information can be accessed from multiple perspectives. But they cannot plausibly be construed as embodying the view from nowhere or the way the world is anyway. They are not utterly objective.

Procedural Objectivity

Still, it is clear that in some sense scientific claims are objective. Following Arthur Fine (1998) and Heather Douglas (2004), I suggest that the sort of objectivity we seek is procedural. In the first instance, then, it is procedures that are objective. Findings are objective because they result from or are confirmed by objective procedures. Objective procedures are epistemically valuable because they promote trustworthiness. They are devised, tested, and certified by epistemic communities, who understand their domains, their disciplines, and the available and appropriate means for investigating the phenomena. Objectivity emerges from the self-reflective activities of epistemic communities. It is neither mere correlation of an opinion with mind-independent facts nor a matter of pure consensus. Procedural objectivity is nonconsequentialist. The objectivity of procedures is logically prior to the objectivity of their objects. What justifies calling a particular result objective is that it is the product of an objective procedure.

Trustworthiness might seem to derive from a source. Being someone of great moral and intellectual integrity, Bill would neither lie nor intentionally

mislead nor convey epistemically dubious information about an important topic. He would not assert that *p* unless he was and took himself to be in a position to back up his claim. Nevertheless, his integrity is not enough to make him worthy of confidence. For Joan to responsibly take his word that *p*, she needs to know or reasonably believe that he is a man of integrity. But we live in a world populated by strangers. We do not personally know everyone whose opinions we might—indeed, must—draw on. Some are trustworthy; some are dishonest; some are unduly cavalier in forming their views; some are careless in conveying them. Often we do not know whom to trust.

To alleviate this problem, we devise and deploy procedures whose outputs are worthy of confidence. If such procedures are properly carried out, their results are trustworthy, regardless of the moral and/or intellectual character of the performer. A postage scale affords excellent reason to think that the envelope weighs 0.5 oz. no matter who is doing the weighing. The results of Marsh's test afford excellent reason to think that a substance contains arsenic so long as a competent chemist performs the test. Such procedures, rather than the particular individuals who perform them, are the sources of trustworthiness (Fine, 1998; Douglas, 2004). The point is not that the procedures are reliable; it is that we have good reason to trust them. They have been devised, calibrated, and validated to satisfy the standards of a realm of epistemic ends.

Objectivity, so understood, is not equivalent to and does not ensure truth. Hal's jingoistic conviction that his American car gets better mileage than his neighbor's foreign car may in fact be true; but his opinion is not objective, since his reasons are political rather than automotive. Nor is Fred's true belief that his investments lost value today, since his conviction is grounded in his horoscope rather than up-to-date economic information. Nor is Maria's belief that there are more Muslims than Christians, based as it is on a coin flip. In none of these cases does the procedure used properly align with the conclusion drawn. Procedural objectivity neither assures nor is assured by utter objectivity. There are objective procedures for determining whether Fred's investments are down even though money is a human institution; whether Joe's car gets better gas mileage than his neighbor's, even though cars are human artifacts; whether there are more Muslims than Christians, even though religions are institutions, religious beliefs are attitudes, and religious observances are practices. Moreover, a

conclusion arrived at objectively may nevertheless be false. A titer test for Lyme disease yields false positives in the relatively few cases where the patient's blood contains antibodies that the test cannot distinguish from those standardly produced by the bacterium *Borrelia burdorferi*. The test has a margin of error. Still, since the margin of error is small, such a test is objective. Although it yields a few false positives, it aligns well enough with the phenomenon in question.

It might seem that trustworthiness is a matter of reliability. What validates the titer test and invalidates the coin flip is that the titer test is, and the coin flip is not, highly reliable. There are at least two difficulties with this position. The first is brought out by a variant on BonJour's (1985, 38–40) example. Suppose Marie has ESP. She is subject to reliable extrasensory deliverances. In 96.2 percent of the cases where she has such a deliverance, it is accurate. She is also subject to the normal range of hunches and intimations, and her track record with these is no better than anyone else's. Moreover, there is no phenomenologically salient difference between her deliverances of ESP and her hunches. Nothing in her experience enables her to tell them apart. Nevertheless, the etiology is different. The deliverances of ESP result from a reliable perceptual mechanism; the hunches result from unreliable mechanisms. (Possibly an fMRI could reveal the difference if only we knew where to look.) Marie is aware that scientists believe that ESP does not exist. So although her genuine extrasensory perceptions are reliable, she has no reason to trust them. She has no way to distinguish them from her unreliable hunches, and has good reason to suspect that there is no such thing as ESP. She does not and should not consider her ESP deliverances trustworthy. Given our available procedures, it is not procedurally objective. Evidently, mere reliability is not enough. Minimally, for the process to be trustworthy, the subject should have reason to think it is reliable.

Even that seems inadequate. A second worry brings this out. An alarming number of therapeutic regimens work for no known reason. An effective treatment for bladder cancer, for example, involves flooding tumors with live bacteria. Somehow this kills the cancer cells. The statistical evidence is strong; the therapy is effective. But, not unreasonably, patients feel epistemically insecure about trusting their lives to a course of treatment that sounds unsanitary and whose effectiveness cannot currently be explained. This suggests that knowing *that* a process is reliable is not enough. We want

to understand (at least roughly) why and how it works. Otherwise any confidence we place in it is precarious.

Evidently, what we seek from objectivity is a stay against bias, idiosyncrasy, and chance. Objective procedures eliminate or control for the personal element and for randomness. Those that eliminate the personal element are impersonal; those that control for it are impartial. The assumption driving the quest for objectivity is that bias, idiosyncrasy, and chance are likely to skew results. That being so, they are untrustworthy.

My discussion does not purport to be a comprehensive account of the factors that can underwrite procedural objectivity. My goal is simply to describe aspects of practices we consider objective to show how communities of inquiry devise and validate them, and why the role of such communities does not impugn their claim to being objective.

Impersonality

Impersonal procedures and devices are, or can be made to be, automatic. They deliver what Porter (1995, 7) calls 'mechanical objectivity'. A motion detector turns on the lights without anyone having to fumble for the switch. It could become a measuring device if a counter were attached to record the frequency with which it does so. A thermometer reflects changes in temperature whether or not anyone notices the changes. The impersonality of such devices and procedures is an achievement, not just a fact of nature. It pays to see how it comes about.

One area where mechanical objectivity is rampant is measurement. We time reactions, weigh objects, calculate areas, and so on. To do this, we require magnitudes—numbers that can be used as bases for comparison of members of a set. So we need to define a magnitude. This is largely a pragmatic matter; it depends on what we want to do with the information we glean. There is plenty of room for choice. But the choice is not completely arbitrary. A viable magnitude must be suitably invariant under the conditions of interest. No one would be happy with the definition of the inch as the length measured by an elastic tape measure. Its readings would be too variable to be worthwhile. But absolute invariance is not required either. Weight is a function of mass and gravity. An object whose mass is constant varies in weight depending on the force of gravity it is subject to. For interplanetary assessments, comparisons of weight are apt to be confusing or

misleading; mass would be a preferable magnitude to use. Still, weight is a fine magnitude so long as we limit ourselves to a single gravitational field, and do not seek too much precision. Although there are small differences in the force of gravity across the Earth, for many purposes they are negligible. For mundane terrestrial uses and most scientific uses, weight is an acceptable magnitude. It is invariant enough.

Once we have a magnitude, we need a way to measure it. The units may be arbitrary. Perhaps the length we call an inch was originally chosen because it was the distance around a man's thumb. That, if true, is a historical curiosity that in no way affects the acceptability of the unit. What matters is that it equips us to make acceptable measurements. An acceptable measurement procedure should insure sufficient interrater agreement, sufficient test–retest agreement, and its results should accord with other acceptable ways of measuring the same magnitude. That is, so long as the item being measured does not change, different competent measurers should obtain pretty much the same result, different measurement procedures should yield pretty much the same result, and a single measurer, making multiple measurements should obtain pretty much the same result. 'Pretty much', because measuring devices are never exact. There is always a small margin of error.

It might seem that once we have defined a magnitude, the procedures for measuring that magnitude are fixed, at least in cases that do not involve proxies. But that is not so. It is one thing to specify what magnitude is to be measured and another thing to specify how it is to be measured.

Town halls in eighteenth century Europe were likely to display a bushel vessel, valid for the region. If anybody questioned the accuracy of any particular bushel, its contents could be poured into the official one to see if they were equal. But this was by no means the end of the matter. Everybody knew that grain could be packed more densely by pouring it from a greater height, and for certain purposes the method of filling might be specified in contracts or by law. (Porter, 1995, 24)

Even where we have the magnitude, we need to specify how to properly perform the measurement.

Nor need the magnitude be defined prior to the development of the measuring procedure. Temperature as a magnitude emerged with the development of the thermometer. The process of development was iterative. It involved a series of steps, each refining and correcting previous ones. A crude thermoscope enabled scientists to improve on everyday sense impressions

as of colder and hotter, yielding an ordinal scale. This enabled them to make further investigations and refinements, eventually concluding that water's behavior was invariant enough that its freezing and boiling points could anchor the scale. They had a choice. They could have, and for a while some of them did, attempt to use other substances—such as alcohol—to anchor the temperature scale. The early 'fixed points' were only relatively fixed. It is not the case that water boils at the same temperature come what may. But the chosen points were relatively fixed, giving scientists a more stable platform to build on (Chang, 2004). The process of devising an adequate measurement device and a magnitude to be measured went hand in hand. They were informed both by what the community sought to do and by the opportunities and obstacles that the world presented. What resulted was a reflective equilibrium that balances, among other things, a magnitude, a measurement practice, a device, and an epistemic goal.

Things are even more complicated when proxies are involved. When it is impossible or impracticable to measure a variable of interest directly, investigators may resort to proxies: variables that correlate with the variables of interest. Climatologists cannot go back in time and measure earlier conditions directly. So they study such things as tree rings and ice cores, investigating preserved phenomena that, they believe, correlate with earlier climate conditions. A core sample of glacial ice, amassed over centuries of snowfalls, contains different isotopes of hydrogen and oxygen at different levels. Since these correlate with differences in the temperature when the snow originally fell, trends in the distribution of isotopes are indicative of changes in ambient temperature. Such proxies are only as good as the correlations they depend on. So the results of investigations that rely on proxies are only as objective as the procedures for establishing the correlations. And these in turn depend on the background theories of the communities devising and deploying the proxies.

The calibration of measuring instruments is similar. In some cases, to be sure, the process is simple, in theory at least. Two seemingly identical measuring devices should be tweaked until their readings are brought into accord. Since whatever the one measures, the other does too, and in exactly the same way, calibration is a matter of minor adjustments. But when it comes to the cross-calibration of instruments, things are different. A balance scale should yield (to a close approximation) the same weight for a given object as a spring scale. To bring this about requires considerable

ingenuity, grounded in an understanding of how each of the devices is sensitive to the weight of an object.

Even the most mechanical of experimental results are not as impersonal as they might appear. A result may simply register on a measuring instrument. But for this to happen, someone had to devise an instrument capable of capturing such a result (see Chang, 2004). A scientist had to devise an experiment that would produce the result. This may have involved a good deal of configuring and calibrating apparatuses. Someone had to set the threshold of measurement, ensuring that the result is not statistical noise. None of these scientists works in epistemic isolation. All of their efforts must be justifiable to the community of inquiry, hence must satisfy community standards. Otherwise there would be no reason to think that the read-out registers anything, that the course of events that occurred in and around the apparatus constituted a test, that the result is worth taking seriously.

Once we have the raw data, we are apt to subject it to statistical analysis. Statistics is the science that controls for chance. It appeals to infinite populations to bracket potentially misleading peculiarities in actual (finite) distributions. It ignores outliers on the grounds that they are not representative of the phenomena of interest. It sets standards for significance and power to rein in the dangers posed by false positives and false negatives. It develops a variety of regressions to disclose subtle patterns in the data.

The impersonality of mechanical objectivity is an achievement. It is the fruit of considerable epistemic labor concerning the phenomena, the proper ways to investigate it, and the available resources for conducting investigations. Methods have to be validated, measures defined, proxies identified, instruments invented and calibrated, thresholds set. Each of these steps is answerable to the standards of a community of inquiry. When and only when this is satisfactorily done can the mechanical procedure yield an objective result.

We should not think, however, that mechanical objective is a mere product of consensus. The world has to oblige. Only some items display the sort of invariance required for measuring devices. Only some materials can be manipulated so as to make instruments that can take and hold their calibration. Only in some areas are there suitable proxies available for variables we cannot access directly. And so on. The products of mechanical procedures are not 'out there' to be read off nature. Nor are they just products

of agreement. They are fruits of careful deliberation and manipulation, grounded in an understanding of the object of inquiry and the available resources for investigating it. (See, among others, Cartwright, 1983; Hacking, 1983; Douglas, 2004.) This is a familiar point, but it is worth emphasizing to make clear that even the most mechanical or impersonal of objective procedures is not a mere matter of mirroring nature.

Impartiality

Impersonality is achieved by off-loading judgment to rules, standards, and techniques whose proper application ensures the acceptability of results. That requires setting limits on precision and on the number of dimensions and fineness of distinctions to be considered. This means, of course, that the community of inquiry had to exercise a good deal of judgment in setting the standards, formulating the rules, and devising the techniques that produced the results. There remain areas where we are unwilling or unable to off-load judgment completely. We want or need to preserve the possibility of more complicated, fine-grained, nuanced descriptions and assessments than available mechanical procedures afford. At the same time, though, we do not want to open the floodgates to idiosyncrasy, bias, and chance.

We thus devise methods that are impartial but not impersonal. When a method is impartial but not impersonal, people—qualified people—apply it and generate results. What makes the method impartial is that it does not matter who in particular those qualified people are. Such methods ground their findings in interrater agreement, where the criterion for acceptability is agreement among members of the community rather than answering to some external standard.

An example of an impartial procedure is the judging of competitive diving. Each dive is to be assessed along multiple dimensions that are spelled out in advance. In each of these dimensions a dive can be done well or badly. And what it is to do well or badly in a given dimension is also spelled out in advance. These specifications remove a good deal of room for idiosyncrasy and bias. The judge is not permitted to ignore the extension of the diver's legs. Nor is she permitted to give or take away points for the grace with which the diver climbs the ladder. But further, more nuanced assessments need to be made within the domains set out by the explicit standards. For example, the rules specify that entry to the water should be

vertical, with the diver's body straight and her toes pointed. When these conditions are not met, judges are instructed to deduct points. How many points is left to each judge's discretion. There the judge is instructed to use her own judgment (USA Diving, 2013: 14). That discretion is needed is no surprise. Individual judges have to decide just how far from the specified ideal a dive was in each of several respects. Although the rules tell them what dimensions to attend to, they cannot and do not try to tell judges exactly how to assess each divergence from the ideal.

The judge is a trained judge. She is not someone randomly recruited off the street. The way she was trained involves having her initial verdicts calibrated against those of others who are already expert judges.[3] The newly appointed diving judge acquires the tacit knowledge that guides more experienced judges. This is not the end of the story, though. A diving competition has multiple judges—typically seven or nine—and to obtain the overall rating for a dive, the officials drop the highest and lowest scores and average the rest. So a judge's personal assessment has a role to play, but it is not an overriding role. If her assessment is out of sync with the assessments of her peers, it will simply be dropped.

Whether the results of diving competitions are objective may seem epistemically insignificant. But what goes on in judging such competitions is similar to what goes on in scientific coding—the process of classifying some observational outcomes as alike, others as unlike. This is often the first step in assessing evidence. Suppose an investigator wants to know whether infants recognize their fathers' voices. A plausible experiment would expose infants to a variety of men's voices, filming and coding their reactions. Multiple trained investigators would look at the films and code what they observe. A critical issue is to decide what sort of reaction indicates recognition. There will presumably be clear cases—ones where any trained observer would recognize that the infant plainly did or plainly did not react differentially to her father's voice. For such cases, interrater agreement and test–retest agreement are easily achieved. But there are apt to be borderline cases. Should the baby's turning slightly in the direction of her father's voice count as a positive response or not? This is a judgment call in much the same way that the diving judge's assessment of whether the diver's toes were sufficiently pointed is a judgment call.

As in diving, what is important is that all assessors use the same standard. It would muddle the experiment if some assessors treated the borderline

cases as positive and others treated the same cases as negative. Coders have to calibrate their judgments against each other. Everyone coding the same data should deliver the same verdicts on at least most of the cases. And a single coder should deliver the same verdicts when she looks at the film a second time. The procedure for classifying cases is impartial. Although epistemic agents are ineliminably involved in the assessment, we have devised methods to control for bias and idiosyncrasy.

The standards may take the form of rules. If so, assessors need criteria to determine how to tell whether the rules have been followed, and what margin of error there is for following the rules. They may also take the form of exemplars. Rather than an articulate or articulable rule, we may rely on cases that exemplify the features we seek to match. The novice diving judge and the novice infant-behavior coder might be better served by being shown clear cases. They would see what a perfect dive or a recognitional response is supposed to look like. They might be presented with foil exemplars to see what should count as a definite mismatch. This would enable them to figure out how far an instance can deviate from the exemplar without being defective. (Arguably, this is what Kuhn's normal science exemplars do [Kuhn, 1970]. They show rather than say what is to count as acceptable normal science.)

The Value of Objectivity

Objectivity as I have construed it depends on the availability and the applicability of public standards of assessment. Despite my talk of magnitudes, measures, and the like, objectivity allows for loose and flexible standards. That students are to turn their papers in on time is an objective requirement, and one that can be objectively measured, even though 'on time' is less than precise. Whether half an hour late counts as being on time may be undecidable without stipulation. But three weeks late definitely does not satisfy the standard. Moreover, the standard may have an implicit ceteris paribus clause that allows for exceptions in the case of unforeseeable emergencies. If so, it would be an impartial but not an impersonal standard. But if there was widespread agreement about what did and did not count as an unforeseeable emergency, objectivity would not be compromised. We could, of course, be more precise if we wished. How precise should we be: on time to the minute? the second? the nanosecond? This way madness lies. For

the more precise standards would be harder to apply but no more useful. In setting objective standards, we need to consider what they are supposed to accomplish. Insisting that they be objective goes but a little way toward answering that question.

Following van Fraassen, I have argued that epistemically useful standards need to be answerable to evidence. We should expand this slightly to allow for answerability to reasons more generally. This is necessary if we want a conception of objectivity that comprehends mathematics as well as empirical inquiry. The satisfaction of procedurally objective standards cannot purport to reflect the absolute conception of reality. They are always products of a view from somewhere. But they are products that stand up to scrutiny when viewed from somewhere else, so they are not purely subjective.

It might seem, then, that we should conclude that objective procedures are designed to secure consensus in a community of inquiry. Since everyone in the community has the requisite level of expertise, it might seem, they will all agree on what the result of an objective procedure is and what it portends. This is plausible, but even for science, it is not quite right. As we saw in chapter 5, science has reason to promote a measure of disagreement (Kitcher, 1990). Within limits, some scientists can responsibly accept a risky new theory while others are more conservative and stick to the old one. Objective procedures leave room for responsible agents to disagree. Rather than fostering agreement per se, I suggest, procedural objectivity constitutes a venue in which both agreement and disagreement are possible.

8 Irreconcilable Differences

Felicitous falsehoods admit of objectivity in art as well as in science. Because aesthetics is often considered the province of subjectivity, in this chapter I take up the role of reasons in aesthetics. I argue that aesthetic judgments are procedurally objective. This procedural objectivity does not ensure consensus, even in the long run. This raises the question whether, or to what extent, consensus should in general be an epistemic desideratum.

The Role of Reasons

Jim likes caramel ice cream; Kim dislikes it. Jim might try to talk Kim into trying it again, but he can offer her no reason to think that caramel ice cream tastes good. He can, of course, say that he likes it. He can add that lots of other people do too. These are sociological facts, and Kim, we may assume, has no quarrel with them. She may recognize that she is in a gustatory minority. But such facts do not get anywhere near the heart of the matter. They give her *no reason* to think that caramel ice cream tastes good. There is no such reason. Some people like it; some people do not. That's the end of it. There is no room for argument because there is nothing to argue about. For a divergence in attitudes to be a genuine disagreement, it seems, there must be some matter of fact about which parties hold opposing views. Whatever their differences with respect to caramel ice cream, Kim and Jim do not disagree. Reasons have no purchase. What we call *tasting good* is entirely subjective.

Things are different in factual realms. There are genuine disagreements about how many Olympic medals Jesse Owens won, whether Neanderthals buried their dead, whether electrons have exact locations, whether

there is milk in the fridge. Opinions may diverge; evidence may be sparse or inaccessible; but in genuine disputes, opinions are supposed to be responsive to reasons.

As we have seen, reasons are not mere expressions of personal conviction; they are considerations that should weigh with others (Korsgaard, 1996b). Ordinarily, a speaker can responsibly adduce as reasons only considerations that she accepts.[1] These are her reasons. Her goal in adducing them is to convince or persuade her auditors of something she is committed to. Reasons thus provide answers to an audience's (often implicit) question: why should we think *that*? In effect, a provider of reasons is saying: these are the factors that weigh with me; they should also weigh with you. A wholly satisfactory resolution to a disagreement consists of conclusive reasons.

A speaker's considering her reasons conclusive is, of course, not enough. Her audience may find (and be right to find) her reasons too weak to be convincing. The audience may have defeaters in their doxastic systems that undermine the significance, relevance, or adequacy of the considerations she adduces. She may have overestimated their strength or bearing on the case. Moreover, not all reasons even purport to be conclusive. A speaker may adduce reasons that are, and are mutually acknowledged to be, relatively weak. They are at best indicative. There is, she says, some reason to believe that the butler stole the spoons or that the bacterium is resistant to antibiotics, but we need more evidence or better reasons before we consider the case closed.

In paradigmatically factual cases, there tends to be widespread agreement about what counts as a reason and, skeptical worries aside, about what would count as a compelling reason, if not a conclusive one.[2] Maybe the evidence about Neanderthals is too sparse; maybe it always will be. Maybe the models, methods, or mechanisms needed to determine whether an electron is a sharp-edged particle or a cloud have not been developed; maybe they never will be. Still, what it would take to resolve paradigmatically factual disagreements generally is clear, or can be made clear. Where there is a genuine disagreement, reasons have a purchase. On this picture, the point of adducing reasons is to resolve a dispute by establishing once and for all which side is right. That reasons function to settle factual disputes is plain. A critical question will turn out to be whether this is their sole or primary function.

So far, we've surveyed purely subjective cases that are unresponsive to reasons, where there are differences in taste but no real disagreements, and factual cases where differences of opinion are genuine disagreements that are in principle resolvable by appeal to public, intersubjectively shared reasons. Discussions about art seem to want to have it both ways. Aesthetic responses are grounded in personal sensory and emotional experience. They seem clearly to be matters of taste. That being so, it might seem, what holds for ice cream should hold for art. Some people like serial music; others loathe it. Some are revolted by disturbational art; others are intrigued. These are sociological truths about subjective responses. We can dispute the sociology if we like, but not the responses. The difficulty is that when it comes to the arts, we do not exercise the open-minded tolerance we show to those with different tastes in ice cream. As Hume (2007) recognizes, we think some people are better judges of art than others. As Kant (1987) insists, judgments about art make a claim to universality. 'I like caramel ice cream' carries no implication that you should like it too. '*The Nightwatch* is masterful' carries the implication that you should think so too. Of course, you might not. It seems then that with respect to the arts, there is something to disagree about. Kant and Hume focus on aesthetic *assessment*—in particular on judgments about whether an item is beautiful or sublime. But the same blend of subjectivity and putative universality is characteristic of *interpretations* of works of art. So I take their points to apply to interpretation as well.

Although there is something right about Kant's claim, its scope has to be restricted. If a critic says that a work is beautiful or banal or original or inept; that it is an expression of the moral bankruptcy of bourgeois society or a commentary on the plight of the working class or a reaction against the stultifying restrictions of the diatonic scale, she implicates that other suitably sensitive, suitably knowledgeable audiences should find it so. In saying that Schönberg's music is beautiful, or that it is a limiting case of tonality, she does not implicate that even deaf people should agree. Nor does she implicate that teenagers who listen only to heavy metal music should share her assessment. Rather, she implicates that all suitably experienced, suitably sensitive listeners to classical music should agree with her. She is in effect appealing to the other members of her specific realm of epistemic ends.

The critic's implicature invites a question: why should we, the members of her realm of ends, think *that*? The critic obligingly supplies reasons to

support her assessment or interpretation of the work. The problem is that another equally knowledgeable and sensitive critic takes and defends a diverging view of the work. He too claims restricted universality for his judgments, appeals to the same restricted audience, and supplies reasons to support his judgment. A dearth of reasons is not the problem. What makes controversies about the arts philosophically curious is that they seem intractable and interminable. Each party can supply reasons, but the reasons are and seem fated to be inconclusive.

To be sure, paradigmatically factual disputes may be irresolvable. But typically we can say what it would take to resolve them. Or if we cannot, we at least consider our failure to be an outstanding problem for the field. When Joe disagrees with Flo about whether there is milk in the fridge, it is perfectly clear what would settle the issue: open the fridge and look. Being in the grocery store, they are in no position to do that, which is why the issue remains unsettled. Nevertheless, they know exactly what would resolve their disagreement. When paleontologists argue about whether Neanderthals buried their dead, they know what it would take to settle the issue positively, at least. Find something that is incontrovertibly an undisturbed grave that contains remains that are incontrovertibly Neanderthal. Then find another one at a second site, thereby indicating that the original discovery was not a fluke. Negative facts are harder to establish. But indirect evidence about Neanderthal cultural practices and comparisons with other hominid cultures might strongly indicate that they did or that they did not bury their dead. A balance of indirect evidence pertaining to cultural development might settle the matter. Maybe it wouldn't. We may never know. The evidence required to answer the question might not be available. Still, parties on both sides can agree about what direct or indirect evidence would resolve the dispute.

Aesthetic Disputes

Not so in aesthetics.[3] There, disputes seem to be interminable. Roger Fry (1952, 13) maintains that the key to understanding *Le Compotier* is to appreciate how Cézanne constitutes mass and volume out of subtle gradations of color. Clement Greenberg insists that the key is to recognize how, by distorting shape, Cézanne emphasizes "the ineluctable flatness of the picture plane" (1966, 103).[4] Each goes on to explain how and why the features he

points to make sense of the work. In so doing, each marginalizes features the other points to. And each does so by the simple expedient of showing that good sense can be made of the work without adverting to those features. We have no idea what would carry the day. What sort of reason, if it could be found, would settle the issue between them?

Both Hume (2007) and Goodman (1978) provide resources for explaining the apparent intractability of aesthetic controversies. Hume believes that aesthetic judgments are procedurally objective. But making an aesthetic judgment is an extraordinarily delicate matter. The requisite procedure is hard to perform. It requires high levels of discrimination, good sense, and imaginative acuity. The judgment must be made in propitious circumstances, where the observation of the item being judged is unimpeded and the observer is sufficiently experienced, and is currently free of prejudice and sensory impairment. Making an accurate aesthetic judgment is on this view rather like doing an extraordinarily delicate scientific experiment, such as those conducted deep underground to block the myriad, barely detectable surface-level vibrations that would interfere with accurate measurement. In short, although there is a fact of the matter as to whether a work of art is beautiful, Hume believes, making an accurate aesthetic judgment is exceedingly difficult. The judgment itself is highly refined, and it is hard to tell in any given case whether the conditions for making it correctly have been satisfied. It is, therefore, not surprising that even sensitive, knowledgeable critics disagree. Still, there is an objectively right or wrong answer; and (skeptical worries aside) in principle a suitably situated, suitably sensitive critic could supply conclusive reasons for her judgment. Where there is disagreement, at least one of the critics is impaired.

Goodman intimates otherwise. Given the symptoms of the aesthetic, divergences in interpretation and assessment are to be expected. "Where we can never determine precisely just which symbol of a system we have or whether we have the same one on a different occasion, where the referent is so elusive that properly fitting a symbol to it requires endless care, where more rather than fewer features of the symbol count, where the symbol is an instance of properties it symbolizes, and may perform many interrelated simple and complex referential functions" (Goodman, 1978, 69), even viewers who satisfy Hume's exacting standards for being competent, impartial judges who are judging in propitious circumstances, are apt to diverge in their interpretations and assessments of a work.

Critical discussions bear this out. Fry and Greenberg discuss the individual dabs of paint that make up *Le Compotier*. Fry notes that they are parallel to one another; that they are nearly rectangular; that there is no attempt to blend one into the next; that they slant from right to left; and that they do so regardless of the contours of the shape being depicted (1952, 45). As he interprets the work, these are significant features of the dabs of paint. Greenberg ignores the specifics, but takes the dabs to constitute "a mosaic of brush strokes that . . . call attention to the picture plane" (1961, 52). On Greenberg's reading, the fact that the dabs are discrete units like mosaic tiles is syntactically significant. But the shape and direction of the dabs, and the indifference of their shape and direction to the shape of the contours they figure in, are secondary at best. Fry takes the direction of the brush strokes to function aesthetically; Greenberg evidently does not. Greenberg takes the exemplification of the flatness of the picture plane to function aesthetically. Fry apparently does not.

An interpretation of a work is a constellation of considerations that bear on the work and seek to illuminate or make sense of it. These considerations, either individually or collectively, supply reasons that constitute a particular reading and perhaps assessment of the work. So, for example, Greenberg's claim that the paint dabs are like mosaic tiles supports his point about the centrality of the flatness of the picture plane, because mosaic tiles not only are flat but achieve their effects at least in part by means of their flatness.

The diverging interpretations are both viable. Each makes sense of the work it pertains to and highlights features that, once identified, strike us as significant. There is no plausible way to say that any of them is *obviously* in error. According to Goodman (1968), we should expect divergence of critical opinions not (or not just) because interpreting and assessing a work of art is hard, but because a certain indeterminacy of interpretation follows from the structure of the symbols that constitute a work of art and the symbol systems they belong to. On his account, the works themselves admit of multiple, divergent, but equally tenable interpretations.

It might seem that this analysis consigns aesthetic differences to the ice cream realm.[5] I said that parties to a genuine disagreement have to be talking about the same thing. If Goodman is right, a single work of art can be construed as any of several symbols, and bears an interpretation because of the symbol it is. There seems then to be nothing—no thing—to disagree

about. Construed as one symbol, it has property *p*. Construed as another symbol, it lacks property *p*. Attempting to disagree would be like trying to argue with a Frenchman about whether 'chat' refers to a conversation or to a cat.

Things are not so simple. For works of art are, or are realized in, objects, inscriptions, and/or events—objects, inscriptions, and/or events that symbolize, to be sure, but still objects, inscriptions, and/or events. An object, inscription, or event is something about which we can disagree. So it is possible to disagree about how best to interpret the object that is a painting by Cézanne, or the inscription that is a play by Shakespeare, or the motion of a body in space that is a dance by Martha Graham. It is even in principle possible to disagree with a Frenchman about whether an inscription of 'chat' refers to a conversation or to a cat. The question is: what is to be gained by doing so? Why bother?

If such controversies can be settled by adjudication, the answer is obvious. We seek an objective answer to a specific question. To get one, we need to determine which interpretation is best supported by reasons. This is how Stadler reads the dispute between Fry and Greenberg. Assuming that if one is right the other is wrong, she finds for Greenberg, saying that his interpretation enables her to appreciate "the gawky ellipse of the compotier's opening, flatter below, arched above"; Fry's does not (Stadler, 1982, 202). Her approach reflects a familiar view of the point of argument. As Nozick puts it, "The terminology . . . is coercive: arguments are *powerful*, and best when they are *knockdown,* arguments *force* you to a conclusion. . . . Perhaps philosophers need arguments so powerful that they set up reverberations in the brain: if a person refuses to accept a conclusion he *dies*" (1981, 4). Nozick acknowledges not only that arguments are never that strong, but also that there is something ridiculous about the aspiration, even if we set aside the macabre idea that the goal of argument is to kill the opposition. If someone doesn't mind being (or being called) irrational, he can disregard the force of argument entirely. Still, the idea that reasons should compel is familiar, and many think it right.

It is implausible in aesthetics. Even if we concede that Greenberg's interpretation makes better sense of the odd geometry of the compotier, he has not given anything like a knockdown argument for his interpretation of the painting. There are aspects of the painting—the massiness of the apples, for example, for which Fry seems to give a better account. To adjudicate

between the two would seem to require delineating and justifying the delineation of all the aesthetically important aspects of the work, establishing a grading scale to evaluate their relative importance, then grading the two interpretations of the work on that scale. That, of course, would require a third interpretation that identified the important properties, and justified the weighting of their relative importance. Its adequacy would have to be judged in turn.

Reason Giving

We seem embarked on a regress, with no end in sight. Do we have to take the first step? That depends on the purpose of reason giving. Why do we go in for it? If our goal is to convince others of the truth of *p* by the deployment of epistemically sound means, then both the point of reason giving and its value are plain. Reason giving transmits the grounds for epistemic entitlement across the community. It transfers justification for a belief. In discussions about art, however, such a goal seems spurious. We expect serious aesthetic debates to be irresolvable. But if there is no reasonable expectation of defeating the relevant alternatives, which would be required to convey epistemic entitlement to one's claim, isn't the enterprise idle? Is aesthetic interpretation akin to the project of squaring the circle? Why should we even start if we recognize that we'll never succeed?

It might seem that we can evade this conundrum by shifting the terms of the debate. Rather than the disagreement between Fry and Greenberg being a disagreement about whether the key to *Le Compotier* is the way Cézanne constitutes mass out of color or the way he emphasizes the flatness of the picture plane, it is a disagreement about whether the work is *best seen* as one where mass is composed out of color, or one where the emphasis lies in the flatness of the picture plane. In that case there is, arguably at least, a fact at the heart of the dispute: how is the picture best seen? Then reasons perform their standard function of supporting or undermining a statement of fact. If this is so, then our mistake was that we mistook what the debate was about.

There is something attractive about this move. We need a basis for dismissing unacceptable interpretations. Some, we want to insist, are simply false. Under no remotely plausible interpretation is *Le Compotier* a portrait of the Pope. Because such an interpretation simply does not engage with its

object, it is easily discredited. The problem comes with interpretations that contain no falsehoods, but are inappropriate or inadequate. They may be uninformative or unenlightening. They may focus on unimportant features or fail to account for important ones. An interpretation of *Le Compotier* that said only that it is a picture of a bowl of fruit is inadequate. It sheds virtually no light on the work. An interpretation that faults the work for failing to address the destruction of the rain forests is misguided. There is no reason to think that it ought to address that issue. If we insist that a correct interpretation says how the work is best seen, we have grounds for dismissing such interpretations. They simply fall short of the mark.

But we go too fast in assuming that the choice is between saying that there is a unique best interpretation and saying that all interpretations are on a par. The omitted alternative is to say that there may be—and indeed often are—multiple equally acceptable interpretations and a vast number of unacceptable ones. It might seem that this points to an easy way out of our conundrum. If multiple interpretations of a work are acceptable, it is plausible to consider them all true. Then we can simply conjoin them. If *p* is true and *q* is true, then *p* & *q* is true. So if '*Le Compotier* underscores the ineluctable flatness of the picture plane' is true and '*Le Compotier* manifests the constitution of pictorial mass out of color' is true, then '*Le Compotier* underscores the ineluctable flatness of the picture plane and *Le Compotier* manifests the constitution of pictorial mass out of color' is true. There's no disagreement. The two critics are just stating different, but mutually consistent truths.

Although logically impeccable, this is unhelpful. For even if the result of conjoining Fry's and Greenberg's interpretations is a truth about the painting, it is not a good interpretation of the work. The two interpretations pull in different directions. An interpretation contends that its object exemplifies certain features. It draws attention to those features and highlights their significance. The reasons it advances make the case that we should find it worthwhile to see the work as exemplifying them. Different interpretations highlight different aspects of the work, drawing attention to them. They provide a way to read the work as exemplifying those features.

Exemplification is selective. Exemplifying some features of an item requires marginalizing or downplaying others. In interpreting *Le Compotier*, Greenberg downplays the significance of features of the work that do not bear on the flatness of the picture plane. In so doing, he takes a risk. If

features he overlooks are too significant to be ignored, his omissions will tell against the adequacy of his interpretation. Under his interpretation, Cézanne's creating volume and mass out of color is relatively insignificant, if not off the table entirely. Has he omitted something too significant to ignore? Any given interpretation involves a balancing act. What is gained and what is lost by choosing to highlight or omit certain features?

If there is exactly one correct interpretation, the stakes are high. But they are nonnegligible in any case. The goal of an interpretation is to make sense of a work; if something salient is omitted or marginalized, the goal may not be met. Still, the idea that a single interpretation should do justice to every significant aspect of a work is unduly ambitious. Just as no literal point of view can disclose every visible aspect of an object, no single interpretation can accommodate every seemingly significant aspect of a work. Features that one interpretation marginalizes may be highlighted by another. Fry's interpretation presents a different perspective from Greenberg's, taking the flatness of the picture plane to be relatively insignificant and the role of gradations in color in constituting pictorial mass and volume to be central. We should doubt that there is a uniquely best interpretation of a work because every interpretation perforce omits or downplays aspects of the work that another interpretation may show to be significant.

Something similar occurs in science. Different models make manifest different features of the target. Diverging models afford different perspectives on the same reality. And it is no surprise that different perspectives reveal different aspects of that reality. There is no optimal model for the same reason that there is no optimal perspective. Every perspective, in revealing some things, inevitably occludes others. Science and art are not so far apart as they might seem.

The role of reasons in aesthetics is not primarily to convince one's audience that the interpretation they constitute is the one true interpretation of the work or that it is the single best way to read the work. Nor is the role of reasons in justifying a scientific model to convince one's audience that the model provides the single best way of representing the phenomenon. It is to highlight features and display their significance. In neither case ought the reasons purport to exclude alternative viable perspectives. Fry and Greenberg articulate reasons that lead us to discern different features of *Le Compotier* and enable us to discern different aspects of reality via our engagement with the work. In proffering reasons, each in effect invites us

to read the work in a particular way. Each intimates that doing so will prove rewarding. Their intimations are tested by our responses. Do we find the interpretations enlightening? Do they illuminate aspects of the work, its connections to other works, its connections to the world beyond the work that we otherwise find bewildering or obscure? Do they enable us to make sense of the work? As I see it, an interpretation is closer to an invitation than to a thesis.

Our reflections on aesthetic disagreements show that differences of opinion should not automatically be construed as adversarial. In adducing reasons to support her opinion, the critic should not be thought of as seeking to best her rivals in a winner-take-all competition. Rather, as Dewey (1916) urges, we should treat those whose interpretations diverge from ours as resources for extending our epistemic range. They evidently see something in the work that we as yet do not. If we seriously entertain their point of view, we may come to discern something of value. This is not to say that every perspective will be rewarding. Some, like that of the Holocaust denier's, can be quickly dismissed. But by accessing the opinions of others and the reasons for those opinions, epistemic agents gain access to insights and perspectives that they would otherwise miss.

In the last few pages, I have focused on aesthetics, because aesthetics is a realm where reason giving and disagreement coexist. It is therefore a good place to look for a role for reason giving other than to win an argument. We have found one. Opposing opinions can make different aspects of their shared object salient. When these are backed by publicly available, publicly assessable reasons, they can make a case for the epistemic value of considering those features important. Opposing opinions backed by reasons can thus enhance our understanding of their object. There is no reason to think that the arts are the only venue where this occurs. Indeed, the highlighting of relevant features is the way resolvable disagreements often get resolved. And it is why we find advancement in understanding in fields like paleontology where in principle resolvable disagreements may never be resolved. We understand more about the Neanderthals and about our understanding of them when we appreciate the force of the arguments for and against the claim that they buried their dead.

Practices that involve the giving and accepting of reasons, I suggest, are procedurally objective. A conclusion that emerges from a chain of reasons gains its epistemic status from the process that led to it. It is credible to

the extent that those reasons in that epistemic environment back it up. It might seem that this notion of objectivity sets the bar too low. If even art criticism qualifies as objective, then, one might think, objectivity doesn't count for much. I think this is wrong. A good interpretation shows us how to look at a work and how to look at the world through a lens that the work provides. It thereby has the potential to enhance our understanding. The interpretation itself is subject to dispute. Reasons can be adduced for and against it. Perhaps the reasons are not Nozick's knockdown reasons. But it is not obvious that the measure of an argument's objectivity should be the power of its punch. If we are going to take the admissibility of incompatible alternatives to discredit the claim to objectivity of criticism, we seem forced to discredit the claim to objectivity of the sciences that admit alternative incompatible models. It seems preferable to recognize that any field that admits of reason giving and sets standards for the appropriateness, accept-ability, and weight of reasons is one that admits of objectivity. Only then can questions arise about how strong the reasons are or need to be.

9   Exemplification in Understanding

The epistemological position I have been developing does not privilege truth.[1] This leaves room for symbols whose contribution to understanding is independent of truth-aptness. Here I argue that understanding is rife with such symbols. Rather than representing, they exemplify. Although I have mentioned exemplification in previous chapters, it needs more explicit treatment. After explicating exemplification, I present case studies to illustrate some of the ways exemplars figure in ethical and mathematical understanding.

Examples are ubiquitous. Philosophers adduce examples to support their theories. Students learn from examples and display their learning by providing examples. Where standards (for, e.g., intonation, proper stance, or poise) are inarticulable, perhaps the only effective way to learn is to take as an example someone who does the thing well. If a student purports to understand a theory—be it Darwinism, socialism, consequentialism, or whatever—but can provide no examples of how the theory applies, her claim is at least suspect. Why are examples important? A single example is, after all, statistically insignificant. So, it might seem, the ability to provide an example should count for virtually nothing. But often it counts for a lot. The reason, I suggest, is that the example *displays* an understanding of the topic. It is not just an instance, it is a telling instance.

Earlier I argued that my position is not a pure coherence theory, since tenability requires that a system answer to something outside itself—our antecedent commitments about the topic, the methods for finding out about the topic, and the standards for justifying our conclusions about the topic. Nevertheless, my position might seem too close to a coherence theory to be palatable. Plato (1997) maintained that the difference between doxa and episteme is that, unlike doxa, episteme has a tether. The tether I focused on

in *Considered Judgment* connects the elements of a system to antecedently held epistemic commitments. The community's understanding must be, in a suitable but flexible sense, sustained by what its members were already committed to. This ensures that they are not making up their accounts out of whole cloth, but it may seem to provide no strong reason to think that their conclusions answer to the facts. I continue to consider the tie to antecedent commitments important. But we also need a tether that connects the system to the facts it pertains to. Ever since Plato, epistemologists have assumed that the tether must somehow link the justification for a belief to its truth maker. As Gettier cases show, such a link is elusive. Here I focus on a different tether—exemplification. This is the connection between a sample or example and whatever it is a sample or example of. I will argue that it plays a central role in epistemology, accounts for the epistemic functions of models and idealizations, and enables us to appreciate the epistemic contributions of symbols that are not truth apt.

The Nature of Exemplification

When an item serves as a sample or example, it exemplifies: it functions as a symbol that makes reference to some of the properties, patterns, or relations it instantiates (Goodman, 1968; Elgin, 1996). In spelling this out, let us take a tolerant approach to properties, recognizing a property corresponding to each extension an item belongs to, regardless of whether that extension is semantically marked or metaphysically privileged.[2] A property, then, is that which all members of a particular extension share. Patterns and relations receive analogously tolerant treatment. Let us call anything that exemplifies *an exemplar* and all of an item's properties and all of the patterns and relations it figures in *its features*.

   Such tolerance is metaphysically innocuous. There is no denying that every item belongs to indefinitely many extensions. The members of any given extension have something in common—namely, membership in that extension. Some extensions strike us as homogeneous. The members of the class of trombones have in common that they are all trombones. That is a class worth marking out, so we introduce a label for the feature that they share. Other classes seem more motley. There is an extension consisting of the Washington Monument, the atomic number of lead, and my calico cat. Unsurprisingly, we have no label for their common feature. Nevertheless,

my cat shares something with the Washington Monument and the atomic number of lead. Their joint membership in that extension is the feature that they share.

Exemplified features may be static or dynamic, monadic or relational, and may be at any level of generality or abstraction. Because each item has a plethora of features, its capacity to exemplify is broad. It is not boundless, however, for exemplification requires instantiation. Only a bit of Swiss chocolate can be a sample of Swiss chocolate; a bit of Belgian chocolate, however similar in flavor or constitution, is disqualified. Moreover, instantiation alone is not enough. The authentic Swiss chocolate that Jon ate yesterday did not exemplify Swiss chocolate or any other feature. It merely satisfied his sweet tooth. To exemplify, an item must refer to the feature in question, and must do so via its instantiation of that feature.

An exemplar can simultaneously exemplify multiple features. The candy might exemplify being Swiss chocolate, rich in flavor, and an expensive indulgence. But an exemplar cannot simultaneously exemplify all its features. It can point up, make manifest, display, or convey some of its features only by marginalizing, downplaying, bracketing, or overshadowing others. Even if a sample problem in a logic text is the only problem in the book that mentions a dog, it does not (at least under its standard interpretation) exemplify the property of being the only problem that mentions a dog. In a context designed to exemplify a logical form, the problem downplays its being the sole mention of a dog. Exemplification is selective.

Because an item can in principle exemplify any feature it instantiates, the range of features it has the capacity to exemplify is vast and heterogeneous. Remei Capdevila's (2009) discussion of Alighiero Boetti's *Rosso Gilera, Rosso Guzzi* brings this out:

Boetti's piece consists of two square, nearly identical panels, one next to the other, whose meager distinction from each other is a slight variation in their red paint, and the raised names and code numbers that identify the different paints, which are inscribed on the panel. Like a color palette, Boetti's piece juxtaposes two different kinds of reds, and in that way it is possible to distinguish between them. In a certain sense, then, the work functions as a paint sample; they exemplify two different synthetic reds, with their commercial codes (60 1232 and 60 1305) and names ("Rosso Guzzi" and "Rosso Gilera"). However the work does not only function as a simple color sample, but exemplifies other properties that a chip of paint in a color palette does not exemplify. "Rosso Guzzi" is the red used to paint Guzzi motorcycles, and "Rosso Gilera" is the one used for Gilera motorcycles, the two rival Italian motorcycle

manufacturers. Put side by side, the two panels not only exemplify a slight difference in color but stand for the divide between passionate advocates of each brand. That is to say, since each kind of red possesses the property of being used to paint a specific kind of motorcycle, they can further exemplify the two different brands, and via a chain of reference, the rivalry between the companies. In addition, since the difference in reds is barely noticeable, the piece can further symbolize the negligible distinction that sustains this rivalry. (130–131)

She could have gone on in this vein. Building on the exemplification of the rivalry between the two groups of motorcycle aficionados, perhaps the work exemplifies features of rivalry in general, including how, for there to be a genuine rivalry, rivals must be very much alike. Although there is plenty of rivalry between professional baseball teams, and between Little League teams, there is no rivalry between the Boston Red Sox and the Lexington T-Ball Tigers. She might have interpreted the work as exemplifying that, objectively speaking, there is little basis for choice among the rivals. Perhaps it exemplifies that in the end all rivalry is sibling rivalry.

Capdevila continues:

The illusion of difference is also exemplified by the fact that both paints come from the same manufacturer, as the serial numbers indicate. And the work's symbolization is not exhausted yet. Boetti's piece is made out of synthetic commercial paint intended to lacquer vehicles instead of common fine arts materials, and in that way, the artistic properties of this paint—glossiness, brightness, viscosity, or the dripping left when applying it on the panel—are exemplified. This can also symbolize the introduction of mass production and consumer goods into the art world. (2009, 130–131)

Here she moves to a different domain. That the paint is ordinary commercial paint used to paint vehicles exemplifies an intimate connection between art and commerce, and the vanishing line between the aesthetic and the commercial. It connects art to the market (and not just—or even mainly—the art market). That features like the brightness, glossiness, and viscosity of vehicular paint are exemplified in a work of art highlights features we might find it worthwhile to attend to when we encounter them in their standard use.

A further interpretation is available, which moves from the economic to the political realm. The paint is not an esoteric substance encountered only in ateliers; it is the sort of stuff you might find in a garage. This exemplifies a democratization of the arts. The material for making art is readily available to pretty much anyone. If the paint on the canvas is mundane

and readily available, and an interpretation of the work draws on a rivalry between motorcycle aficionados, then art need not be the province of an elite. It is open to all. Evidently a single item can, in the right context, exemplify any and many of its features, enabling the interpreter to forge a variety of epistemically valuable connections across a variety of domains.

Exemplification is not just a device for underscoring salient features or supplying emphasis. It often highlights and affords epistemic access to features that were previously disregarded, even to features that are semantically unmarked. It points up a feature, enabling us to mark out its contours and appreciate its significance. It equips us to recognize that feature when we encounter it in other contexts. Exemplification, then, is not just a vehicle for conveying salience; it is a source of salience. Moreover, the features an exemplar highlights may be difficult to discern. An effective exemplar can marginalize conspicuous features in order to exemplify elusive ones. The most conspicuous feature of a manufacturing process may be its deafening din, while in the context of a safety inspection what the process exemplifies is a barely detectable vulnerability to sabotage. The vulnerability then stands out, while the noise, although still deafening, fades into the cognitive background of things to be ignored.

Being symbols, exemplars require interpretation. In the context of the inspection, does the manufacturing process exemplify vulnerability to sabotage, susceptibility to unscheduled downtime, the prospect of prompting a major insurance claim? The very same security flaw might instantiate all of these. In treating that flaw as an exemplar, the question is: an exemplar of what? Some exemplars belong to regimented systems, where the structure of the system supplies or highly circumscribes the answer. Others are ad hoc. Then interpretation depends on the context within which they function and the way they function in that context.

Even apparently straightforward, highly regimented cases are not so simple as they seem. A customer peering at a commercial paint sample card might seem thereby to know exactly what color Benjamin Moore jonquil yellow paint is. It's the shade on the card. Such cards constitute a regimented system if anything does. Still, knowledgeable consumers recognize that the colors on the sample cards symbolize not an exact shade but a range of colors around the shade appearing on the card. (That is why paint companies recommend that you buy all the paint you will need for a job at once. The paint from different batches that satisfy the recipe for being

Benjamin Moore jonquil yellow will rarely be visually indistinguishable.) In interpreting the card, consumers need to determine what range of discriminable shades qualifies as jonquil yellow. The cards often provide resources for answering this question as well. Typically a sample card does not just exemplify a single color. Nor does it provide exemplars of an arbitrary collection of colors. Such a card usually consists of a sequence of exemplars of closely related shades. In figuring out the range of the exemplar of jonquil yellow, consumers rightly reason that it includes all the Benjamin Moore yellows that are closer to the exemplar of jonquil yellow than they are to the exemplars of the closest alternatives—perhaps goldenrod, sunflower, and saffron. Although some vagueness remains, relevant contrasts help to fix the reference of an exemplar.

Ordinarily, folks do not wander into paint stores out of idle curiosity about available paint colors; they want to choose paint for a job—maybe to paint the house. So they need to project from the sample to a larger extension it represents. Let us say that a sample that exemplifies a particular feature *typifies* the extension comprised of all and only things that instantiate that feature. Such items *match* the exemplar. The paint sample then typifies all and only those things whose color falls within the range of Benjamin Moore jonquil yellow. By exemplifying a feature—by highlighting or displaying that feature—an exemplar affords epistemic access to it. Someone who properly interprets the exemplar is in a position to recognize the feature in question.

Because paint samples belong to regimented systems, knowledgeable consumers have little difficulty determining what features they exemplify and what extensions they typify. But not all exemplars are so readily interpreted, for not all exemplars belong to regimented systems. One of the powers of exemplification as a referential device is that we can improvise exemplars at will. Simply adducing something as an example typically makes it one. Cycling to work can exemplify concern for the environment; a can of gourmet cat food can exemplify the propensity to unduly pamper pets; a slipshod student paper can exemplify indifference to the course, and so forth. Treating something as an example normally makes it one.

Interpretation can be tricky, particularly in the absence of a regimented system. An ornithologist identifies a bird as an example of a junco. He may do nothing more than point to it, expecting his audience to figure out what extension it typifies—that is, what other birds are juncos. Or he may

underscore its size, coloration, markings, and the shape of its beak. Had he ignored the bird, it still would have had all these features; but it would not have symbolized them. By pointing it out, he exploits features the bird had anyway, bringing the bird to exemplify them. Perhaps he tells his audience which features to fixate on; perhaps not. If all goes well, they can now recognize a junco when they see one. But despite his tutelage, they may be uncertain how to interpret the exemplar. How closely do other birds have to match the exemplar to belong to the typified extension? Do the eye markings matter? Does the plumage vary with season or age? Do juncos exhibit sexual dimorphism? To complicate matters further, the same item can exemplify different features in different contexts. A free sample of laundry detergent that exemplifies its cleaning power in a marketing campaign exemplifies banality in an assessment of that campaign. A bird that exemplifies being a junco in one context exemplifies being a nonmigratory bird in another.

When exemplars are products of improvisation, interpreters have no regimented system to rely on. They draw on context, background assumptions, and, where available, collateral information instead. In communicative settings like the one above, their interpretive stance is Gricean (see Grice, 1989). They assume that the cooperative principle is in effect, hence that they have the resources needed to interpret the exemplar correctly. If so, the feature exemplified is relevant to the communicative exchange; it is capable of becoming salient to the audience via that exchange, and its being made salient will promote the objectives of the exchange, or will be sufficiently important to justify abruptly changing the subject.

In investigative settings, interpretation relies on an analogue of the cooperative principle that involves confidence in the study's design. In effect, interpreters assume that nature is cooperative. Assuming the study is well designed, the result—whatever it turns out to be—will exemplify something informative, accurate, unambiguous, and relevant. Needless to say, such confidence can be misplaced. Some results exemplify not hitherto unknown features of the phenomena, but inadequacies in the study's design. Results that were expected to fall within a given range fall outside it, or they are ambiguous, or they fail to achieve statistical significance. This is disappointing. But it should be no surprise that the assumptions that guide interpretation are sometimes unfounded. Exemplars are not always effective or properly interpreted, but when they are effective and properly

interpreted, the reasons are similar to those that explain the effectiveness and proper interpretation of other symbolic exchanges.

Focus on pedagogical and commercial samples may suggest that exemplification is a conduit for what is already known or understood, rather than a source of knowledge or understanding. This is incorrect. A mining inspector extracts air samples from a mineshaft to find out something no one yet knows about the distribution of gases in the mine. If the samples are properly taken, he has reason to believe that the gases his samples exemplify are typical of the gases in the mine. A pollster questions potential voters about their attitudes toward immigration. If the sample is suitably chosen and the questions well framed, she acquires new information about public opinion on the topic. There are, of course, no guarantees. Both cases involve projection from a limited sample. It is a brute fact of inductive life that even well-chosen, well-taken samples are sometimes misleading. But if the samples are well taken and the background theory is well founded, there is epistemically good reason to project to a wider class of cases.

An appreciation of the feature and its instantiations can be woven into our developing understanding, stabilizing or enhancing its equilibrium. Alternatively, an exemplar may threaten equilibrium, making manifest that things are more problematic than we previously thought. Either way, there is an epistemic gain. A well-chosen sample or example can either strengthen an account or disclose an unanticipated weakness.

Given what I've said, accurate projection might seem assured. The sampler need only project to the extension the sample typifies. The problem is to figure out which extension that is. The mine inspector's air sample typifies all the air in the mine that shares the chemical features exemplified by the sample. That's trivial and unhelpful. The critical question is which air is that? Is it the air throughout the mine? Or is it the air at a given level in the mine? Or is it the air where the mine crosses a particular seam of ore? Or is the sample a statistical fluke, exemplifying features it alone possesses? These are standard and often hard questions about how to characterize evidence in order to make valid projections from it.

One might wonder whether the air sample or the poll's results are evidence at all. A plausible criterion of evidence is that evidence pertaining to p consists of whatever bears on whether p is the case. If so, epistemically inaccessible items qualify as evidence. So do items no one has any reason to consider evidence. Suppose 83 percent of the cases of absolute pitch have a

genetic basis. No one knows what the genetic basis is or even suspects that there is a gene for absolute pitch. Obviously, genetic screening would be useless in attempting early identification of children with absolute pitch, since investigators would not know what to look for. Still, on the above criterion, the presence of the gene would qualify as evidence. Although such unknown factors bear on whether p, they are epistemically inert. They provide no reason to believe, accept, or even suspect that p. We might conclude that this discredits the criterion. Alternatively, we might distinguish between *being evidence* and *functioning as evidence.* Then we can concede that the gene was evidence all along, even though there was no feasible way to access that evidence. Not much hangs on the choice between these alternatives. Either some items that are in fact evidence do not function as such, or to qualify as evidence, an item must, in the epistemic circumstances, be capable of functioning as such. The critical point is that for something to function as evidence for p, it should provide a reason to believe, accept, or at least suspect that p. For something to function as evidence against p, it should provide a reason to disbelieve, reject, or doubt that p. This, I suggest, requires that to function as evidence, an item must exemplify its connection to the hypothesis it bears on. Once investigators know enough about DNA, a strand of hair found at the crime scene can function as evidence linking the suspect to the crime. The hair exemplifies a genetic pattern instantiated by the subject and few, if any, others.

Making an Example Of

Goodman (1968) maintains that anything that instantiates a feature can exemplify it. In principle this may be so, but in practice things are not so simple. For not all instances of a feature can constitute *good* or *effective* exemplars. A feature may be camouflaged: a tiger who blends into the surrounding jungle instantiates being striped, but would in that circumstance be unlikely to exemplify stripedness. (It might, however, exemplify how stripes—even orange stripes—can camouflage in a jungle of green.) A feature may be occluded: a bald man wearing a hat is ill positioned to exemplify his baldness. A feature may be overshadowed: the fearsome timbre of a lion's roar may block its effectively exemplifying its pitch. An instance may be a borderline case. Even though a chartreuse fire hydrant is a vivid instance of greenness, its color is too close to yellow to be a good exemplar

of greenness. An interpreter would likely take it to typify the members of a class consisting of some greens and some yellows rather than all of the greens and none of the yellows. It would not be a good tool for teaching a young child colors.[3]

What converts an instance into an exemplar? Sometimes conversion requires no more than pointing. What is a junco? That's one! So long as it is clear that we are concerned to identify a normal member of a species, no more may be required. Sometimes we need to exclude distractors. The roof of a crocodile's mouth is a distinctive shade of coral; but a paint company would be ill advised to recommend that potential customers peer into a crocodile's mouth to see that color. Crocodiles are so rare and so dangerous that any glimpse of the roof of one's mouth is unlikely to make the color manifest. We could not see it long enough or well enough, and would be unlikely to attend to it carefully enough (or, if we did, to survive long enough) to decide whether it is the color we want to paint the den. It is far better to create a lasting, readily available, easily interpretable sample of the color—one whose function is precisely to make the color manifest. Such a sample should be stable, accessible, and have no properties that distract attention from the color.

Considerable stage setting is often required to generate an effective exemplar. Since shape is always accompanied by color (Plato, 1997), we cannot hope to provide a completely colorless exemplar of a dodecagon. We might, however, tell our interlocutors to ignore the color and focus on the shape of our exemplar. Or we might provide several, differently colored dodecagons, and induce them to focus on the salient common property. Either way, we set the stage to orient the audience to the feature of interest. In yet other cases, we need to tamper with the substance being sampled. An experiment designed to discover whether water is compressible would not use ordinary rainwater or tap water. Such liquids contain impurities. Were we to effect compression of such a liquid, we would not be able to tell whether the water or the impurities accounted for the solution's compressibility. Rather, the experiment would begin with distilled water—water from which, as far as we can tell, all impurities have been removed.

Before we can remove irrelevant factors, we need to do some analysis: we need to conceptualize the item in question as comprised of components—those we seek to exemplify, and those we do well to set aside. Analysis is often straightforward. Our prior understanding of the domain frequently

enables us to identify the relevant factors. Rainwater = H_2O + impurities. But things are not always so simple. For a different experiment we might want to distinguish between the components of rainwater that are due to pollution and those that are not. In that case, pollen would be a component of rainwater rather than an impurity. In yet other cases we might take rainwater to consist of whatever liquid falls from the sky. Then even sulfuric acid would be a component of rainwater. As our interests and purposes change, we may need to reconceptualize the phenomena to devise an effective exemplar.

The proper way to analyze depends on what we are trying to do. Linguists factor verbal behavior into competence and performance. Competence is stable; performance fluctuates. Given this factorization, they can investigate a variety of verbal phenomena effectively. They can, for example, bracket performance errors to exemplify features that remain invariant across language use. Or they can bracket underlying competence to identify patterns of performance errors common in a given speech community or at a given stage of linguistic development. In assessing fluency, however, they might consider the competence–performance distinction idle. They might do better to partition the domain so as to exemplify the frequency and distribution of surface errors, being indifferent to underlying linguistic structures.

Differentiating between signal and noise is commonplace. It is a matter of distilling meaningful information out of the total available information on a topic. But what is signal in one context is noise in another. Ordinarily, a questioner wants to know the answer to her question, and a teacher wants to know whether her student knows the answer. In these cases, the content of the answer is the signal. How quickly the respondent replies counts as noise. But psychologists often focus on reaction times. For their purposes, it doesn't matter what the respondent answers. The content of her answer is noise. The signal is how long it takes her to answer. To distinguish between signal and noise requires a criterion for what makes information meaningful. The criterion is keyed to the question at hand. If the question is a new one, inquirers may need to reconceptualize the phenomena or repartition the domain to devise an appropriate criterion.

Such reconceptualization can highlight features that obtain but are obscured under more standard characterizations. Rather than calling an enclosed curve an ellipse, it might be fruitful to consider it a perturbed circle. The curve is an ellipse, just as the raindrop contains both sulfuric acid

and pollen, and Boetti's painting contains two rectangles. But the items admit of other descriptions that may, in a given context, be more illuminating. The critical question is what features of the item we want to be in a position to focus on. Since exemplification is selective, to treat an item as an exemplar is to selectively disregard some of its features.

The novice who emulates the performance of an expert treats that expert as an exemplar. She identifies the features of his behavior that she thinks account for his success. That is, she factors his complex behavior into elements, selectively disregards those she considers irrelevant to his effectiveness, and sets herself to reproduce the rest. It is neither necessary nor possible to reproduce his performance in its entirety. Some of the features of that performance are irreproducible—the fact that it was done on Tuesday the fifth of June, or done by someone born in Detroit, for example. Others are simply irrelevant. There is no reason to think that someone's tennis game would improve were she to change the color of her bedroom curtains or to reorganize her library to match his. Yet others are at the wrong grain. Possibly revising how far she stands from the baseline would make some difference to her serve, and it would not be surprising if copying the expert in this respect is a good idea. But it is probably not beneficial to insist on emulating his position to the nearest millimeter. Instead of attempting to slavishly reproduce everything she discerns, she interprets the expert's behavior as exemplifying the features she takes to constitute, underwrite, or contribute to his excellence. If she is successful in identifying and reproducing enough of the features that figure in his success, her performance is apt to improve.

Berkeley's criticism of Locke's discussion of abstract general ideas brings out the power of selectivity. According to Locke, "The general idea of a triangle . . . must be neither oblique nor rectangle, neither equilateral, equicrural, nor scalene, but *all and none* of these at once" (Locke, 1984, Book IV, ch. 7, sect. 9, quoted in Berkeley, 1957, §13). Berkeley balks. He can, he insists, frame no such idea. Nor does he think that anyone else can. The definition of the abstract triangle is, as Locke admits, inconsistent. How are we supposed to frame an idea of such a thing? How can it figure in proofs? But if we cannot frame such an idea, how is it possible to prove theorems that hold of triangles in general? Berkeley's answer is this:

Though the idea I have in view whilst I make the demonstration be, for instance, that of an isosceles rectangular triangle whose sides are of a determinate length, I

may nevertheless be certain that [my proof] extends to all other rectilinear triangles of what sort or bigness soever. And that is because neither the right angle, nor the equality, nor the determinate length of the sides are at all concerned in the demonstration. It is true the diagram I have in view includes all these particulars, but then there is not the least mention made of them in the proof of the proposition. It is not said the three angles are equal to two right ones, because one of them is a right angle, or because the sides comprehending it are the same length. Which sufficiently shows that the right angle might have been oblique, and the sides unequal, and for all that the demonstration would have held good. (Berkeley, 1957, §16)

By selectively disregarding the rightness of the angle, the equality of the sides, and the exact length of the sides, Berkeley brings the particular triangle he is working with to exemplify features it shares with all other Euclidean triangles. It affords epistemic access to geometrical features of Euclidean triangles in general.

One might think that exemplification is not necessary to achieve this goal: it suffices that the precise measures of the sides and the angles are ignored. This is not so. The isosceles right triangle is *capable* of symbolizing generally because these particulars can be selectively disregarded. But to be capable of symbolizing generally is not enough. A picture of my cat does not become a picture of cats in general merely because in discussing it I omit mention of the distinctive color of her eyes and the markings on her fur. Rather, Berkeley insists, "An idea which, considered by itself, is particular, becomes general *by being made to represent or stand for* all other particular ideas of the same sort" (Berkeley, 1957, §12, italics mine). The mode of representation is exemplification. The isosceles triangle in his proof not only *is* an enclosed three sided figure on a Euclidean plane; his proof is an effective general proof because *the triangle refers to its being such a figure*. By so doing, it stands for other such figures, and discloses properties they share.

Exemplars need not be propositional. Boetti's painting and Berkeley's triangle, as well as the multitude of samples and examples we encounter every day, lack truth-values. Still, it is worth emphasizing that the mode of symbolization can be crucial to the exemplification of a particular feature, and that propositional modes do not always have the edge. In 1854, a cholera epidemic broke out in a London neighborhood. John Snow, now recognized as one of the fathers of epidemiology, traced the epidemic to its source by plotting cholera deaths on a city map that included the locations of water pumps and major buildings. His plot demonstrated that the deaths

were centered around the Broad Street pump. To be sure, the denizens of the neighborhood shared more than a water pump. They breathed the same air, bought food at the same markets, were likely bitten by the same fleas, and so forth. So the proximity of the pump would hardly seem to favor the hypothesis that cholera is water borne. Snow's map displayed the significance of the pump, because it also disclosed that people associated with two local institutions did not contract cholera. Workers in the local brewery, who drank beer rather than water, were unaffected; so were denizens of the nearby workhouse, who drew water from a different pump. The two institutions supplied a natural control for testing the hypothesis that water from the Broad Street pump was the source of the epidemic (Tufte, 1997). Snow's map exemplifies the connection between the cholera deaths and the pump. It shows that the pump is at the center of the epidemic. A comprehensive list of the addresses of the victims, the locations of London water pumps, breweries, and workhouses, and the admission that brewery workers prefer to drink beer would not have made anything like so clear a case, for it would not have made the proximity to the pump perspicuous.

Case Studies

What I've said so far has perhaps made it plausible that exemplification is epistemically important. It highlights features we might otherwise overlook. This might seem to be merely a matter of directing attention. But exemplification plays a far more central role in understanding. Let's look at some cases.

Ethics

Suppose our goal is to understand the wrongness of lying. This might mean a variety of things. We might want to understand *why* lying is wrong, or *what makes* lying wrong, or *when* or *to what extent* lying is wrong. These are all legitimate questions. But I am after bigger game. I want to understand how lying's being wrong fits into the fabric of human life. Satisfactory answers to each of the foregoing questions will supply part of the answer but, I suggest, only part. Although it is widely agreed that lying is generally wrong, there is considerable controversy about why. Kantians locate the wrongness of lying in its being a violation of the categorical imperative; consequentialists, in the undesirability of the outcomes that result. Perfectionists focus

on the defective moral character of the liar; rights theorists, on the violation of the hearer's rights to truthful information. And so forth.

Some philosophers, such as Strevens (2010), maintain that to understand the wrongness of lying is to derive a proposition such as 'Lying is generally wrong' from the true moral law that grounds it. Each of the positions I've listed, we may assume, can provide a suitable derivation from its chosen basis. None has achieved consensus that *its* basis is *the real basis*. This is not by itself problematic. It could just be evidence that ethics is hard. I have a different worry in mind. It is that any such derivation—even the one grounded in the real basis, if there is such a thing—is incomplete. It omits important aspects that other theories supply. To see it, suppose we concede that, say, Kantians are right. What grounds the wrongness of lying is that lying typically violates the categorical imperative. Nevertheless, it remains true and not unimportant that liars are morally defective; that lies tend to produce a preponderance of undesirable consequences; that lies violate people's rights; and so forth. Even if these truths cannot be derived from the categorical imperative, if we want to understand the wrongness of lying, they ought not be overlooked.

I suggest that to understand the wrongness of lying is to appreciate how lying's generally being wrong permeates human relationships, practices, and institutions. This involves more than just recognizing how it fits into the moral order. Other aspects of lying's wrongness stem from the cognitive, emotional, social, and economic costs of lying and being lied to, as well as the costs of designing and implementing safeguards to protect against lies. Our understanding might incorporate an appreciation of the mechanisms to secure the bases of trust among colleagues and strangers, and/or mechanisms to limit the consequences of trust's being betrayed, and so forth. It might be sensitive to the ways lying fosters distrust, undermining the credibility of even those who tell the truth. It might even recognize how the wrongness of lying figures in the natural order—for example, in the evolution of language and of creatures capable of using language. To understand the wrongness of lying, of course, we also need to appreciate how and when lying is morally permissible, and how and when, permissible or not, it is advantageous. The issue is extraordinarily complex. An understanding of it ramifies in numerous, sometimes surprising directions.

Each of the moral laws, and each of the social, institutional, and biological accounts I've mentioned, captures something significant about the

wrongness of lying. Even if, as Kantians maintain, the reason why lying is wrong is independent of the positive or negative consequences of telling lies, it is hard to deny that lying often has bad consequences. The same holds for its figuring in a morally defective character. It is hard to deny that these features matter. If we want to understand the wrongness of lying, we should not lose track of such points. So an account of understanding that stops with grounding relations is inadequate. To understand the wrongness of lying evidently involves appreciating the myriad ways lying's being wrong infiltrates multiple, disparate domains. Even if some of the factors I mention are only peripherally related to whatever it is that *makes* lying wrong, we should not cavalierly dismiss them. A comprehensive understanding of the phenomenon would do well to accommodate such peripheral features.

I have cast a very wide net, and claimed that the phenomenon of interest extends across and, no doubt, well beyond its reach. Maybe moral wrongness resides at the center of the net, but prudential and epistemological tendrils extend outward. How can we hope to grasp it? As I've described it, the problem is what William James (1983, 448) called a 'blooming, buzzing confusion', with no obvious way to corral it. I've mentioned a bunch of disparate facts, each of which bears some connection to our topic, many of them mediated by controversial—or at least controverted—theories. It might seem that the effort to understand the wrongness of lying is hopelessly demanding.

Rather, I suggest that each of the plausible moral accounts of what makes lying wrong, and each of the plausible social, economic, and biological accounts, affords some insight into the issue by exemplifying a different way in which lying is deleterious. Kantianism exemplifies the failure of respect inherent in lying; consequentialism exemplifies the largely negative consequences; perfectionism exemplifies the defectiveness of the liar's moral character; and so forth. Each highlights features that the others obscure. The several accounts of what makes lying wrong are incompatible. They cannot all be true. But each plausible account is true enough to be capable of exemplifying an important feature of lying's deleteriousness. To focus on only one account—even if it is the true and ultimate account of morality—would leave us with an impoverished understanding of the wrongness of lying.

It does not follow that anything anyone might say or think about the wrongness of lying has some place in my capacious account. Because

truthfulness is so central to human interactions, there is a lot that an understanding of the wrongness of lying might incorporate. But there is plenty that it rightly excludes. The Pinocchio-inspired contention that lying's wrongness engenders unattractive nose elongation has no place in an understanding of the topic. I suggest that any account that exemplifies deleterious features or consequences of lying is apt to exhibit some understanding of the topic. So do accounts that show why lying is sometimes advantageous or even admirable. But accounts that allege that lying has features it does not have, or that features of lying are deleterious (or, indeed, advantageous) when they are not, embody a misunderstanding of the topic.[4] They purport to exemplify features that lying does not in fact instantiate or features that lying instantiates but are irrelevant to lying's being wrong.

Mathematics

One might think that the plausibility of my claim derives from the lack of consensus about the ground of lying's wrongness. Perhaps the epistemic value of diverse exemplars stems from the sorry state of ethics. I disagree. A similar pattern obtains in mathematics, a field characterized by overwhelming consensus. Mathematical theorems are proved and valid proofs are considered beyond dispute so long as the axioms and rules remain in effect. But what exactly is the function of proof?

Fermat's last theorem is: For $n > 2$, $x^n + y^n \neq z^n$. It is true. By 1990, it had been tested computationally for values of n through many millions. By then, probably no one who considered the matter worried that it was false. Still, mathematicians weren't satisfied. They demanded proof. Evidence, even overwhelming evidence, wasn't good enough. Why not? One seemingly plausible answer is they want a right to be sure. But doesn't all that evidence confer a right to be sure? By 1990 there was far more evidence of the truth of Fermat's last theorem than there is for almost any empirical contention we count as knowledge. No one denied that. Still, in mathematical circles, Andrew Wiles's proof of the theorem was a cause for celebration. This suggests that the proof of a theorem is doing something other than supplying sufficient reason to think that the theorem is true. What might that be?

One possibility is that what is wanted is not just sufficient reason to think that the theorem is true, it is sufficient reason to think that it is

necessarily true. Empirical evidence, or the quasi-empirical evidence that testing cases computationally supplies, is evidence of truth, but not of necessary truth. The problem with this proposal is that Fermat's last theorem is a mathematical claim. We already know that if it is true, it is necessarily true. There's no chance that although true, it's just a coincidence that it turns out to be true, or that, although true in this world, there are other possible worlds in which it is false. The goal of proof, it seems, is neither to engender a nonfortuitously justified belief that a mathematical proposition is true nor to engender a nonfortuitously justified belief that it is necessarily true. Something else seems to be going on.

There are additional reasons to buttress this suspicion. We value proofs of what might be called Moorean mathematical claims, such as $2 + 3 = 5$. They are, like Moore's commonsensical claims, contentions we are completely convinced of independently of proof (see Moore, 1993). Indeed, it is overwhelmingly likely that if we had a putative proof that such a claim was false, we would reject the proof and the proof procedure that generated it rather than reevaluate the claim. Still, it is lovely to see how such a claim follows from the Peano axioms.

I suggest that the function of proof is not mainly to afford reason to believe that a theorem is true or that it is necessarily true. It is to exemplify a series of mathematical relations that show it to be true. Demonstration is exemplification. The function of a proof is to make the necessary truth of a particular mathematical proposition manifest—that is, to exemplify its necessary truth and to exemplify relations between its truth and other mathematical truths.

To see this, consider Pythagoras's proof of the irrationality of $\sqrt{2}$. (This is a very simple, lucid, and justly famous proof. I include obvious steps, and explications in order to make my point clear.)[5]

The background:

1. A *rational number* is a ratio, a/b where a and b are integers.
2. Odd numbers always have odd squares. (This has been previously proven.)
3. Common factors cancel out, so every fraction is equivalent to a fully reduced fraction.

The proof:

4. Let a/b be a fully reduced fraction. (This begs no question since common factors cancel out.)
5. If $\sqrt{2}$ is irrational, $2 \neq (a/b)^2$.

6. $2 \neq (a/b)^2$ is equivalent to $a^2 \neq 2b^2$.
7. Assume for reductio: $a^2 = 2b^2$. (That is, assume that $\sqrt{2}$ is rational.)
8. Then a^2 is even (since it is equal to $2b^2$).
9. Therefore a is even (since the square of an odd number is always odd).
10. Since a is even, $a = 2c$ (where c is some integer).
11. Therefore $2b^2 = a^2 = (2c)^2 = 4c^2$ (from 5 and 8 and $((2)^2 = 4)$).
12. Therefore $b^2 = 2c^2$ (by dropping $a^2 = (2c)^2$ and dividing by 2).
13. Therefore b^2 is even (since it is equal to $2c^2$).
14. Therefore b is even (since the square of an odd number is always odd).
15. But if both a and b are even, then they have 2 as a common factor.
16. Therefore a/b is not a fully reduced fraction, which contradicts the hypothesis.

QED.

The proof consists of a chain of exemplars, each highlighting a feature that would have to obtain if $\sqrt{2}$ were rational. Some (like the fact that the square of an odd number is always odd), are drawn from previously proven background information. Others are products of earlier steps in the proof. Many are equations. a^2, $2b^2$, $(2c)^2$, and $4c^2$ are mathematically equivalent. In Fregean terms, they differ only in their mode of presentation. Despite their equivalence, the choice among them is not arbitrary. The modes of presentation matter, for the expressions exemplify different features of their ostensible referent. For example, that $b^2 = 2c^2$ makes it manifest that b^2 must be even. And in light of the fact that only odd numbers have odd squares, this makes it manifest that b must be even. Had another equivalent expression, such as $(a^2 + 14 - \sqrt{25} - 3^2)$, replaced $2c^2$ in step 10, b's evenness would have been obscured. Because the proof must make b's evenness manifest, such a substitution would be unacceptable. At different points in the proof, different features need to be exemplified. The proof proceeds by a series of transformations, each step exemplifying something that had been instantiated but not exemplified previously. By highlighting consequences implicit in the assumption that $\sqrt{2}$ is rational, proof as a whole thereby exemplifies the impossibility of $\sqrt{2}$ being rational.

The mode of representation a proof uses to characterize the items it draws on is critical. For a sequence of mathematical propositions to constitute a proof, the mathematical entity or operation must be represented in a way that exemplifies the features that make manifest the connection between the premises, axioms, lemmas, and definitions and further elements in the chain, and ultimately to the conclusion.

Mathematics values multiple proofs of the same theorem. If truth were the only issue, one valid proof would suffice. Possibly a second proof would give a modicum of corroboration; but beyond that, further proofs would be otiose. Gauss devised eight proofs of quadratic reciprocity. A vast number of additional proofs have been developed since. Why go to all that trouble? Surely not because after the first few proofs mathematicians were still harboring doubts about the theorem's truth! Fans of grounding maintain that the only proof whose apprehension constitutes an understanding is the proof from the *real* metaphysically basic axioms (see Strevens, 2010). Neither position is plausible. Nor is either consonant with mathematical practice. Rather, I suggest, the reason multiple proofs of the same theorem are valuable is that they exemplify different mathematical relations.

There are numerous proofs of the Pythagorean theorem. Pythagoras's proof relies on the properties of squares; Bashkara's proof, on properties of right triangles; Garfield's proof, on properties of trapezoids. Each contributes to our understanding of geometry by showing how geometric truths about different shapes interweave. Recently, Zimba (2002) devised a trigonometric proof of the theorem. This was something mathematicians thought could not be done. They believed that since $\sin^2 a + \cos^2 a = 1$ presupposes the Pythagorean theorem, any trigonometric proof of the theorem would be circular. But Zimba's proof by subtraction does not rely on $\sin^2 a + \cos^2 a = 1$. So the role of the Pythagorean theorem in the proof of $\sin^2 a + \cos^2 a = 1$ is irrelevant. Zimba's proof thus exemplifies that the relation between Euclidean geometry and trigonometry is more complicated and less stratified than was previously supposed. We increase our understanding of the theorem by seeing how it figures in different mathematical domains, how it follows from and leads to different theorems. Our understanding of the warp and woof of the fabric of mathematics is enhanced by each of the multiple proofs. Just as different features of the moral realm are exemplified by different accounts of the wrongness of lying, different features of the mathematical domain are exemplified by different proofs of the theorem.

Exemplification's role in mathematics extends beyond its function in proofs. It enables us to appreciate a major advantage that the Arabic numeral system has over the Roman numerals. We are taught in school that one of the limitations of Roman numerals is that the notation did not have a numeral for 0. This sensitized us to the bewilderment that would emerge should someone attempt to subtract V from V. But 0 plays another important role

in Arabic numerical notation. It serves as a place holder. Roman numerals had no need of such a place holder. Other early notations just used spaces between numerals. This makes it difficult to discern the difference between, say, 1 14 and 1 14. So 0 as a placeholder is an enormous convenience (see Mazur, 2014). All of this is true. But there is another, conceptually deeper benefit. With Roman numerals, you can count only so far before you have to introduce another letter—C or M or whatever. This prompts the question: what happens to the numbers when the alphabet is exhausted? The system of Roman numerals supplies no clue. The Arabic system, having the capacity to use 0 as a placeholder, exemplifies the unboundedness of the integers. They are never exhausted; you can just keep counting.

Conclusion

In this chapter I've argued that exemplification is an important vehicle for epistemic access. It acquaints epistemic agents with features of things they had been unaware of or whose significance escaped their notice. Exemplars such as textbook examples and commercial samples are conduits for the transfer of understanding or knowledge. But exemplification also figures as a source of understanding. Different approaches to a problem can highlight different aspects of it. The proof of a theorem exemplifies, and enables epistemic agents to understand, mathematical relations that were latent in the axioms. Multiple exemplars exemplify different aspects of the same phenomenon, enabling epistemic agents to appreciate how that phenomenon figures in the greater scheme of things.

10 Embodied Understanding

In this chapter, I look specifically at the role of exemplification in dance. By so doing, I hope both to reinforce my earlier argument that the arts are vehicles of understanding and to highlight the ways in which symbols that are not and do not purport to be true can be vehicles for understanding. This is crucial for what follows. In the next two chapters, I will argue that such symbols are ubiquitous in the sciences.

Dances as Symbols

Swan Lake is beautiful. It is delicate, graceful, enchanting. Martha Graham's *Night Journey* is not. It is riveting, harrowing, horrifying, often ugly. Yvonne Rainer's *Trio A* isn't even that. Being utterly pedestrian, it does not play on the emotions at all. But it is intriguing. Taken together, these three dances raise questions: What is dance up to? What does it do and how does it do it? *Night Journey* discredits the thesis that the end of dance is beauty. *Trio A* discredits the thesis that the end is affective engagement. Possibly dance as such has no end. Different works and different genres pursue different ends. But whether or not dance as such has a telos, questions arise: What does this particular dance do? How does it do it? And why?

I will argue that beauty, ugliness, motion, and emotion are among the means dance uses to embody and convey understanding. This is not to say that dance has only epistemic functions. We may value dances for their own sake, or for the pleasure they give, or for other reasons as well. What I claim is that one important function of dance is epistemic. Dance enriches our lives at least in part because it enables us to understand things differently than we did before.

Like Goodman (1968), I believe that dances consist of symbols. This claim is not, on the face of it, problematic. Dance critics regularly speak of a genre's vocabulary or idiom. If this way of speaking is accurate, choreographers draw on the resources of a symbol system to create their works. In that case, to glean an understanding from a dance, spectators must interpret those symbols correctly. They must recognize what the dance symbolizes. They must know how to read the dance. If they do, they understand what the dance conveys. But for the dance to convey understanding, rather than merely to be understood, more is required. The spectators must also have reason to accept what the dance conveys.

This is problematic. How can dances provide such reasons? Even if dances are comprised of symbols, they present no arguments. They do not make statements, express propositions, or assert that the world is this way or that. Some, like *Swan Lake* and *The Firebird*, have a narrative structure. But they are fictions. And fictions do not make literal assertions about the way the world is. Moreover, much of what seems significant in such works does not figure directly in the plot. So, evidently, something else is going on besides telling a story. Other works, like *Trio A* and Merce Cunningham's *Points in Space* are nonnarrative. They do not stand in a representational or denotative relation to the world. They are not truth apt.

To make my case, I need to undermine the idea that only symbols that figure in arguments—that is, only symbols that have a propositional structure or are components of items with a propositional structure can advance understanding. If nonpropositional items can advance understanding, then the thesis that dance advances understanding has some chance of being correct. Let's look at some cases.

Consider the counterexample. If Mike asserts, 'All swans are white', all it takes to refute him is one black swan. No words need be spoken. The bird alone is enough. The black swan contributes to Mike's understanding of ornithology by demonstrating to him that his belief about swan plumage was false. It may, of course, do more. Perhaps it has a ripple effect, prompting him to revise a cluster of associated beliefs.

As van Fraassen (2008) argues, perspective also figures in our understanding of the world. By adopting a different perspective, we come to see familiar items in new ways. We thereby appreciate relationships between them that we previously had overlooked or underemphasized. For example, the shift from a third-person to a first-person perspective may be crucial to

appreciating the close connection between belief and assertion. The assertion 'It is raining and Kate does not believe it is raining' is unproblematic. There are, after all, many facts that any given person does not believe. But 'It is raining and I do not believe it is raining' is an instance of Moore's paradox. It is not something I can reasonably assert. The first-person perspective, but not the third-person perspective, thus affords reason to think that assertion is intimately connected with belief.

A third case is pattern detection. Even if all the evidence is in hand, understanding is enhanced when a pattern emerges. In such a case, although the facts were previously known, the relations between them were not perspicuous. Snow's plot made manifest that virtually all the cholera victims of the 1854 epidemic got their drinking water from a single source (Tufte, 1997). It led to the obvious, but at the time radical, conclusion that contaminated drinking water was spreading the disease.

One might object that all of these *can* be captured in propositions. So, one might think, there is an implicit argument. It's not the black swan, then; it is the proposition: 'Here's a case that shows that your hypothesis is false: [insert black swan]'. It's not the perspective, then; it is the proposition 'Looked at from this point of view, you will see that the following is unassertible [insert an instance of Moore's paradox]'. It's not the pattern; it is the proposition 'This plot shows that the data cluster around a single point [insert map]'. Although one can frame such arguments, they do not paraphrase away or capture in propositions the items in question. Those items have been embedded into propositions. But they do their cognitive work independently of such embedding, and are only worth embedding because of their prior cognitive status. In Wittgenstein's (1961) terms, they show rather than say. They exemplify.

Propositions are standardly held to be that which truth-apt sentences express. Two sentences express the same proposition just in case they are co-intensional. Moreover, it is held, the co-intensionality condition is sometimes satisfied; some pairs of sentences are co-intensional. If 'vixen' and 'female fox' are synonymous, then 'Vixens are fierce' and 'Female foxes are fierce' are co-intensional. They express the same proposition. Any world in which vixens are fierce is one where female foxes are fierce; any world in which female foxes are fierce is one where vixens are fierce. Recently, metaphysicians have argued that propositions are too coarse-grained to do the work philosophers want done. One problem stems from the human

capacity for ignorance and error. Even if 'vixen' and 'female fox' are synonymous, it is entirely possible for Harry to believe that female foxes are fierce without believing that vixens are fierce. Even if 'London is beautiful' and 'Londres est belle' express the same proposition, Pierre can believe one but not the other (Kripke, 2011b). Propositions, as standardly characterized, are not sufficiently fine-grained to reflect belief contents. Such contents are evidently hyperintensional.

Some, such as Bealer (1998), reject the co-intensionality criterion and maintain that propositions are hyperintensional too. If so, perhaps we might say that the contents of dances and other works of art are propositional. This would require a rather drastic departure from tradition in at least two respects. First, it would recognize that propositions need not be truth apt. It is hard to see what the truth conditions on *Points in Space* or *Trio A* could be. Second, it would recognize that there are cases where a symbol is so fine-grained that no other symbol could have the same content. Such symbols are semantically dense (see Goodman, 1968). Although the notion of a proposition can be stretched, I am not confident that hyperintensionalists want to stretch it this far. In any case, our concern is with what dances exemplify, not with their claims—if any—to be propositional.

Dances highlight certain properties, rendering them salient, and thereby affording epistemic access to them. Classical ballet, for example, literally exemplifies properties such as grace, delicacy, and beauty; and metaphorically exemplifies properties such as love and longing, weightlessness and ethereality. Martha Graham's modern dances metaphorically exemplify psychological properties such as grief, regret, horror, and hope. They literally exemplify that the body of the dancer has a certain weight—that it is subject to literal as well as metaphorical gravity. George Balanchine and Merce Cunningham, choreographers who bridge the modern—postmodern divide, created works that exemplify properties of dance itself, movements of dancers in time and space. They also exemplify properties like vitality, dexterity, and sinuousness. The works of the choreographers in the minimalist, postmodern Judson Dance Theater exemplify properties of ordinary pedestrian movement. Rather than exemplifying properties stereotypically associated with dance, they exemplify activities like walking, running, carrying a mattress, and climbing over a barrier.

I have characterized the properties dances exemplify using monadic predicates. This might suggest that they can be instantiated in an instant. But the important properties in dance are typically dynamic. They emerge and develop over time and across space. Jocasta's convulsive grief, Odette's ethereal grace, Cunningham's jittery counterpoint, and Paxton's prosaic walk are spatiotemporally extended. This, I suggest, is crucial to their epistemic functions.

A Look Back

The Judson Dance Theater consisted of a group of minimalist postmodern choreographers who sought to pare down dance to what they took to be its essence: human bodies moving in space. They sought to democratize dance, to eliminate its elitist, distancing, off-putting qualities. They had no interest in storytelling, in transcendence, in illusion of any kind. Their dances consist of mundane, pedestrian, nonstylized, uninflected movements of the sort you can see on the street. Yvonne Rainer expressed her choreographical ideals in her over-the-top "No Manifesto" of 1965:

No to spectacle no to virtuosity no to transformations and magic and make believe no to glamour and transcendency of the star image no to the heroic no to the anti-heroic no to trash imagery no to involvement of performer or spectator no to style no to camp no to seduction of spectator by the wiles of the performer no to eccentricity no to moving or being moved. (Rainer, 1995)

Once all these aspects of dance are excluded, what is left, the minimalists maintained, is the essence of dance—movement as such.

What might be the value of such a dance? We see people walking, running, climbing over barriers, carrying loads every day. Why should we go to a performance to watch them? Why should we buy a ticket? Sally Banes (2003, 3–5) suggests that the answer lies in defamiliarization, a process by which what is familiar is rendered strange. The idea is this: when something is familiar, we are so accustomed to it that we do not really look at it or attend to it. A passing glance enables us to recognize it as what it is and then move on. Defamilarization heightens awareness of things that are so obvious that we routinely ignore them. We walk, run, climb, and see others doing so without giving it much thought. When we carry a mattress, we do give it thought. We are painfully aware that carrying a mattress is

hard. It requires continually readjusting our bodies to accommodate the awkwardly shifting center of gravity of the bulky, heavy, unwieldy burden. But we are intent on the task—we want to get the mattress moved. So we attend to the task and not to our doing of it.

The Judson Theater dancers put us in a context where we attend to the physical intelligence that goes into such mundane activities. We notice and attune ourselves to the minute, intricate, muscular adjustments involved in keeping one's balance while schlepping a mattress. We notice the rise and fall, the small and large physical adjustments that it takes to walk or run across the floor. The dances thus exemplify features that mundane motion instantiates but that we, either makers or observers of that motion, routinely ignore. The exemplification is literal. The dancers exemplify features of walking by walking. They exemplify features of climbing by climbing. On the one hand, their message seems to be 'What you see is what you get'. On the other hand, they put spectators in a context where they can ask, 'Well, what do we get?' and see, perhaps for the first time, what was before their eyes all along. By making us aware of the physical intelligence of ordinary, mundane movement, the Judson Theater's dances increase our awareness and advance our understanding of ourselves as organisms capable of locomotion.

Arguably, they do something more. Although many of the discrete, component movements in Rainer's *Trio A* are ordinary movements that pretty much anyone could do, it is not the case that the complex movements that they are part of are things that just anyone can do. As Jill Sigman says, "Moving a head one way and feet another is difficult enough, but switching quickly from head to feet to other body parts is even more challenging. . . . Furthermore, some of the movements are simply difficult to accomplish. One passage requires slowly rising into relevé on one leg and repeatedly alternating legs. Another involves squatting and extending the left leg fully to the back, then bringing it under the torso and through to the front without losing one's balance" (1998, 164–165). This undercuts the idea that the uninflected, unvirtuositic dances are just 'slices of life' brought indoors and presented in such a way that we can attend to ordinary movement for its own sake. Nevertheless, the dances exemplify ordinary movements, and present the more complicated movements in the same uninflected way as they present the ordinary ones. Perhaps they thereby exemplify that the

ordinary is continuous with the extraordinary, or that the ordinary is itself extraordinary.

According to the 1960s minimalists' ideology, there is something phony or inauthentic about dances that purport to be something other than what they are. So it is perhaps not surprising that exemplification figures prominently in their works. Since exemplification requires instantiation, a symbol can exemplify only what it is—that is, only features it has. This suggests that what is characteristic of postmodern minimalist dance might not be characteristic of other forms of dance.

Ideologically, Merce Cunningham and George Balanchine are not far from the minimalists. They too eschew narrative and psychological expression. They too want to pare dance down to its essence. But rather than taking the essence of dance to be *human bodies* moving in space, they take it to be *dancers* moving in space. So their works exemplify features of dance itself. Dancerly forms, movements, and patterns are exemplified in their works. *Points in Space*, Cunningham (1986) maintained, is about dance; it is not about anything else. Unlike the Judson group, Balanchine and Cunningham do not purport to restrict their range to movements anyone at all could do. They recognize that dancers have extraordinary physical and expressive abilities, and are willing to exploit the full range of those abilities. Thus they take properties like grace, virtuosity, suppleness, and the illusion of weightlessness to be suitable candidates for exemplification. So are abstract geometrical and kinematic patterns that can be realized only by trained, talented dancers.

Again the question is: what is the cognitive value of such a dance? A dance about dance could embody and convey an understanding of the art form. It could show what dance (or perhaps ballet, or perhaps a certain style of modern dance) as such does and how it does it. It could make manifest that dance is not just, or not mainly, a sort of entertaining pantomime for telling fairy tales without words—for conveying something that could be better done in words. Balanchine seems to suggest that classical ballet starts, 'Once upon a time, there was a prince who fell in love with an enchanted swan . . .' Now take away the prince, take away the swan, take away the love and betrayal, take away the enchantment, and what do you have left? The answer to that question, arguably, is what Balanchine makes manifest in his ballets.

Still, there is something irritatingly self-indulgent about artists' talk of exploring the limits of their medium. One wants to reply, 'Yes, yes, I can see why artists working in a medium and art students studying a medium need to care about the limits of the medium. But why should the rest of us care? What sort of understanding does such an exploration yield for us?' Later, I will suggest reasons to think that Balanchine, Cunningham, Rainer, and their colleagues provide acceptable answers to these questions. For now, however, let us look at their predecessors.

Modern dance tends to respect nature. Human bodies present themselves as human bodies, not snowflakes or enchanted swans or ethereal spirits. They are subject to the laws of physics and psychology, and sometimes ground down to the earth by the forces acting on them. Modern dances are more likely to present dancers writhing on the ground than leaping improbably through the air. Psychology is central. In Martha Graham's works we see the outward manifestations of mental states, expressing fear, joy, elation, and revulsion. Rather than relying on established conventions, as classical ballet does, they create their own meanings—they constitute the symbols that convey their content.

One way is through narrative. Knowing, as we do, the story of Oedipus, we have resources for interpreting the gestures in *Night Journey*. But *Night Journey* is no mere pantomime of *Oedipus Rex*. It presents Jocasta's mindset at the moment of her suicide, something Sophocles left out. It reviews the joys and sorrows and eventual horror of her life with her son/lover/husband Oedipus. It displays the guilt, revulsion, and self-loathing that make suicide the only option. It does so through tensions and releases, conventional gestures and newly invented ones—motions that express tenderness, repugnance, love, and profound regret. The dance intimates that the true tragedy of *Oedipus Rex* is Jocasta's. To have borne a son and, loving him, to have (inadvertently and with the best intentions) brought it about that such a horrifying fate could befall him is to have utterly betrayed one's obligation as a mother. To have loved a man and put him in a position where all he could feel for you or himself is loathing and disgust is to have utterly betrayed one's obligation as a lover and a wife. *Night Journey* enhances our understanding of *Oedipus Rex*. It presents the story from a novel point of view, and reveals features that are not salient in Sophocles. It modulates and elaborates the understanding we glean from the play, convincing us that there is more to the story than the fate of one man with a tragic flaw.

Whatever insight into the human condition we gain from the original play is extended and ramified through this new interpretation.

Modern dances, such as *Night Journey*, are not insular in the way the works of Rainer, Balanchine, and Cunningham are. They point beyond themselves and appeal to resources drawn from the outside. They refer to things that are apparently not integral to dance. They may tell stories to express feelings. They elicit and express emotions. I have argued elsewhere that emotions exemplified in the arts are vehicles for understanding (see Elgin, 1996, 2007b). Still, there is a worry. It seems that Martha Graham could give an utterly compelling performance of *Night Journey* even if she happened that day not to be feeling the amalgam of love, regret, revulsion, and self-loathing that the work expresses. The emotions expressed by the work evidently need not be felt by the dancer.

This, however, presents a difficulty for my position. Exemplification, as I keep insisting, requires instantiation as well as reference. If the dancer does not experience the emotions in question, how can the dance instantiate them? This is a tricky question that, I will argue, reveals something important about the philosophical significance of dance. But before I take up the problem, let us take a brief look at classical ballet.

Classical ballet is practically the antithesis of Judson Dance. It says 'yes' to many of the items listed in Rainer's "No Manifesto." Yes to spectacle, illusion, and make believe. Yes to magic, virtuosity, glamour, transcendancy of the star image. Emphatically yes to the seduction of the spectator. Classical ballets tell stories, magical stories about enchanted princesses and evil magicians. They rely on a fixed, regimented, vocabulary of techniques, positions, and steps that conspire to present the illusion of weightlessness, of ethereality. Classical ballet looks outward. It is about something other than dance. Still, it hardly seems to open a window on the world. It portrays fairy tales. It is populated by necromancers and their victims, by wizards and magical creatures. If there is a message here, it may be that ballet seeks to enchant, but we should be wary of being enchanted. Although ballets are about something in the sense of being comprised of representational symbols, there is evidently nothing in the world they are literally about. They are fictions. While telling stories about nutcrackers or firebirds or enchanted swans, they exemplify features like grace and delicacy, fluidity and transcendence. They make manifest how beautiful, light, and ethereal human beings can seem.

Dance as Philosophy

This cursory reverse history of dance in the West shows a paring down, a stripping away. Each genre I mentioned took its predecessors' works to contain excesses that needed to be expunged, leaving only what was essential to dance. The move away from balletic spectacle was a progression toward the exemplification of only what is essential to dance. Philosophically, this is interesting. Who would have thought that dance is a hotbed of essentialism? But this trajectory raises an important question. Dance consists of human bodies in motion. So on this essentialist account, dance should be restricted to what human bodies in motion can exemplify. What is that? We know that an item can exemplify only properties it instantiates. So the question is, what sorts of properties can human bodies in motion instantiate?

The postmodern minimalists maintain that bodies cannot instantiate properties like being weightless, or being an enchanted swan. So dances that portray them as such are in some respect violating the essence of dance. They object to ballet's pretense of defying gravity—of being lighter than one really is. But pretending is something ordinary people do, and pretending to be lighter than one is is something lots of us do. They object to the grand leaps as not the sort of thing that ordinary people do. But Michael Jordan, at the height of his career, probably made as impressively graceful leaps as Mikhail Baryshnikov. Granted, Michael Jordan is far from ordinary. But his extraordinary talent has nothing to do with dance. They object to ballet's standardized, regimented symbol system, considering it an imposition of an authoritarian dance culture rather than deriving from the way ordinary people behave. But human beings are by nature acculturated. So to find that something is a product of culture is not to show that it is not natural for beings such as ourselves. Many of our cultural practices involve subjecting ourselves and each other to rigid rules. Ordinary language is a cultural construct that imposes grammatical rules on acceptable verbal behavior of human beings. The inside—outside distinction turns out to be hard to draw.

The postmodern minimalists object to the psychological properties exemplified in both ballet and modern dance, on the grounds that they are mental, not physical. This suggests that postmodern minimalists are closet dualists. A dualist would say that a figure bowed in grief exemplifies a

posture from which one can infer that she is grieving, and that a trembling figure exemplifies motions from which one can infer that she is afraid. Still, grief and fear are mental, not physical properties. But according to materialism, mind and body are one. So the figure bowed in grief itself instantiates, and therefore is capable of exemplifying, grief. The trembling figure instantiates and is capable of exemplifying fear. No inference is needed.

This last point is critical. One way to accommodate the exemplification of emotional properties conveyed by dance is to say that they are metaphorically exemplified. One dancer droops, displaying a particular posture that metaphorically exemplifies grief. Another leaps, displaying a motion that metaphorically exemplifies joy. This may be the right thing to say. But the materialist position might be correct. In that case, postures and movements themselves can exemplify properties that we call mental. This posture just shows grief, that movement just shows joy. For this is what grief looks like and that is what joy looks like.

This brings us back to a question I left hanging earlier. If the performer is not experiencing the emotions, can the dance exemplify them? The first inclination is to say 'No', or anyway 'Not literally'. Emotions need to be experienced, and if the dancer is not experiencing them, who is? But the assumption that emotions need to be experienced to be instantiated is not as obvious as it looks. Although emotions are associated with distinctive feelings, emotions are not feelings. It is possible to have an emotion that one does not feel. This, psychiatrists tell us, is what happens when a person is in denial. Other psychological factors override or short-circuit the connection between emotion and feeling, so that an emotion that is plainly being exhibited is not felt. It is also common in what Hume calls the calm passions—emotions like an enduring grudge or the standing affection for a life-long friend, which has no distinctive feeling associated with it, but which manifests itself in a complex disposition to behave in certain ways toward that friend (see Hume, 1928, Book II). Still, one might object, even in these cases *someone* has the emotion in question. Whether it is felt or not, it resides in someone.

Could we say, then, that the emotions exemplified in *Night Journey* reside in the dancer? Even if Martha Graham does not *feel* the regret and revulsion and self-loathing the dance exemplifies, those emotions are manifest in the motions her body goes through as she dances the part of Jocasta. She has them, whether or not she feels them. If they are literally characteristics of

a human body, and are characteristics a human body can have whether it feels them or not, then they can be literally exemplified in the dance, by the dancer, regardless of what the dancer feels.

Similarly, of course, for other properties, like ethereality or weightlessness or enchantment. Real bodies cannot instantiate those properties. But they can seem to instantiate them. So they can appear to be ethereal or weightless or enchanted. And they can literally exemplify the appearance. I mention this possibility, not because I think I can demonstrate here that it is correct, but because it seems to me that dance raises interesting and important questions in the philosophy of mind. If the mind just is the body, then what dancers do with their bodies, they do with their minds. The full title of Rainer's work is *Trio A or The Mind Is a Muscle, Part 1*. I suggest that the 'is' here might be close to an "is" of identity or an 'is' of constitution—*The muscle is (at least part of) the mind.*

I suggest, then, that the postmodern minimalist contention that earlier works are somehow phony or inauthentic because they go beyond what is distinctive of bodies in motion is more problematic than it looks. It involves a tacit commitment to a limited, apparently dualistic conception of the range of things that human bodies actually do. This is not to say that the postmodern minimalists were wrong to limit their range in the way that they did. Rather, it is to suggest that their understanding of the basis of the limits they set may be wrong. One of the ways dance advances understanding is by raising philosophical questions, like 'What exactly is the relation of the body to the mind?' and 'How does the body in motion manifest intention, intelligence, emotion, and other putatively mental properties?'

Many of the features exemplified in human movement are exemplified elsewhere, as well. Patterns are abstract mathematical structures. So they might be exemplified in a dance, and instantiated not only in other human activities, but in events of different kinds. The dance of the snowflakes in *The Nutcracker*, for example, exemplifies the sort of pattern one sees when snow skitters across the ground. Patterns exemplified in dance are typically dynamic. They develop over time. So they often look like the patterns one sees in a kaleidoscope. These are complex kinematic regularities. By exemplifying such patterns, dances sensitize spectators to them, enabling those spectators to recognize them when they encounter them in other venues or to discern subtle or enigmatic aspects of them.

Dance frequently exemplifies political properties as well. Yvonne Rainer's egalitarianism is manifest in, among other things, a democracy of the performance space. There is no mandatory center of attention. Since the actions are uninflected, all the actions of all the dancers are potentially equally significant. The dance does not impel or compel the spectator to look at one spot or another. The 'costumes' of the dancers (if they are not nude) are pedestrian, thereby depriving the spectator of the sort of social cues that clothing often provides. Virtuosity is neither displayed by nor required of the dancers. All, evidently, are created equal.

The collaboration of Cunningham and John Cage was, on the face of it, odd. In works like *Points in Space*, dance and music were created independently. The dancers typically did not even hear the music until the first performance. So rather than either art accommodating itself to the demands of the other, each stands alone. Each is worthy of attention. And in performance either may call attention to or distract attention from the other. This is not an accident; nor, from the point of view of the creators, is it a defect. For the works exemplify the autonomy of the different arts. They also exemplify the capacity of autonomy to provide an occasion for serendipitous juxtapositions.

In classical ballet, as in many other dances, music and choreography work together. Each enhances and draws on the other. The interdependence thus exemplified allows for the heightened effects that neither alone could achieve. Classical ballet also invokes the 'star system' that Rainer decried. It exemplifies a hierarchy, an inegalitarianism. Politically, as well as stylistically, postmodern dance and classical ballet are antithetical.

We need not endorse its political stance to understand and appreciate a dance. If my contention is correct, what we should do is understand what the dance symbolizes, and how it reflects on other aspects of our experience. Different dances display different values. They afford access and insight into the values of egalitarianism, autonomy, inegalitarianism, and interdependence. Ordinarily, we might not think of works like *Trio A* or *The Firebird* as particularly political works. But once we realize that among the properties they exemplify are political ones, we are in a position to recognize a political dimension to other actions and institutions that we might otherwise think of only apolitically. We emerge from a performance of such a work better equipped to recognize such features in other situations where we encounter them.

Interpretation and Understanding

I have been urging that dances are symbols that exemplify features and render them epistemically accessible. But how can we understand such symbols? At a very abstract level, the answer is clear. We understand dances the same way we understand other symbols—we know how to interpret the symbol systems they belong to. I can read the menu if I understand German, for the symbols on the menu are words in German. It is plausible that ballet has something like a 'language'—a set of repeatable, conventional symbols whose interpretations are reasonably clear. But when Doris Humphrey is writhing on the floor, or Steve Paxton is carrying a mattress, or Merce Cunningham is twitching in the corner, or Jill Sigman is dancing on crutches, what are we to make of it? There does not seem anything like ballet's regimented conventions to fall back on.

Here, I think, it is worth taking seriously what writers on dance say when they speak of a choreographer's vocabulary or idiom. They take it that to understand a dance, we need to be able to interpret that vocabulary or idiom. There is no reason to think that this is easy or automatic. It may be that one needs to know quite a bit about what has been going on in contemporary dance in order to understand a new work. It may be that we need to know what has been done, what has been tried, and whether it succeeded or failed, to figure out what Sigman is up to. (But you also need to know German to read the menu.) If one has the requisite background, what is being symbolized may be perfectly clear. If not, one may wonder why, for example, the dancers are nude, or are wearing masks, or are evenly distributed across the stage, or whatever. One may wonder what the connection between the score and the dance is or why words rather than music constitute the score. There is no reason to think that the answers to these questions are obvious or are readily available to the novice spectator. We should no more expect to be able to interpret a dance in an unfamiliar idiom just by looking than we expect to be able, just by looking, to interpret a menu, much less a poem or a news report or a scientific paper, in a foreign language.

Confronted with an alien dance form, we may initially be bewildered. We venture hypotheses and test them to discover whether they make sense of what we are seeing. There are many modes of access, so we may find that we have relevant background resources to build on. We might, for example,

appeal successfully to our knowledge of other arts or of popular culture. In some cases we will formulate plausible hypotheses that assign to a work an interpretation that accounts for the features we find salient. Then we have insights to export to other aspects of experience, and to bring to the interpretation of other works of art. The interpretation we venture may or may not stand up to further scrutiny. If, for example, it yields a reading that makes a new work anomalous, when it seems continuous with the chore-ographer's previous works, we have reason to doubt its adequacy. If the insights it leads to seem banal when the work seems intriguing, then again we have reason to think our interpretation is inadequate. If, on the other hand, it makes sense of the factors we find salient and illuminates other aspects of experience, it is a prima facie plausible interpretation.

I claimed earlier that dance has the capacity to embody and convey an understanding of the wider world. I argued that this capacity is largely due to exemplification. We can now see what this involves. Dance, like other arts, exemplifies properties that are instantiated elsewhere but may fail to be noticed or properly attended to in the motley array of inputs that regularly confronts us. In exemplifying these properties, a dance draws our attention to them and stresses their significance. It thus equips us to recognize them when we see them again and intimates that we would do well to attend to them. In effect, then, dance may be a source of working hypotheses. Obviously, not all such hypotheses are sound. An insight we attempt to export from a dance may fail to illuminate anything significant about other aspects of experience. Many dances are banal. In this they do not differ from other symbols that purport to reveal things about the world. Many putatively informative conversations, and many scientific experi-ments, are banal, too. But when a dance is cognitively effective, it reveals something to us. We come to see the story of Oedipus in a new light, or come to appreciate the complex physical intelligence of ordinary move-ment, or come to recognize the precariousness of what we standardly take for granted.

In some cases we may remain bewildered. We have no idea why these people are doing those things, why anyone would consider what is going on art, or dance, or the sort of thing anyone would want to do in public. Then we can make no sense of the work. But even this may be an advance in understanding. This is Socrates's point. Knowing that you do not know is the first step toward knowledge. Appreciating that you do not understand a

work of art constitutes the first step (or at least an early step) toward figuring out how to make sense of it.

Conclusion

I have argued that dance highlights features and underscores their significance. These may be features of human behavior—mundane or extraordinary. They may be more abstract dynamic or kinematic patterns. They may embody views about weighty philosophical matters—the relation between mind and body, between autonomy and interdependence, between the ordinary and the extraordinary. Dance can deepen understanding by sensitizing us to things we tend to overlook, by undermining stereotypes, by problematizing assumptions we did not even know we were making. Since dance, like the other arts, lacks the capacity to adduce arguments, it might seem that at best, it can advance hypotheses whose merits need to be demonstrated by argument. Even if this were so, dance would have an important epistemic function in identifying hypotheses worth testing. But, I suggest, argument is not the only way to buttress epistemically tenable convictions. Examples can be equally tenable. An exemplar—even if it is a nonverbal symbol—can show that something is the case, even though it is powerless to *say* anything. A work of art, such as a dance or a symphony, being an extended, textured, multilayered symbol, can show something quite significant.

11 Casting in Bold Relief

Il faut reculer pour mieux sauter.

Earlier I said that felicitous falsehoods figure in understanding by contributing to an account's overall tenability. The question is how. It might seem that considerations that are acknowledged to be false cannot be initially tenable. That is not so obvious, since initial tenability turns on what people actually accept. That Amherst is ninety miles from Cambridge, although strictly false, is probably accepted and acceptable when considering driving from one city to the other. If so, it is initially tenable. People might well accept considerations they consider false, particularly if they consider them true enough. In any case, falsehoods that lack initial tenability can still in principle contribute to the tenability of the account they belong to. As we saw earlier, not all elements of an acceptable account need be initially tenable. Some initially nontenable or even untenable considerations earn their tenability by strengthening the tenability of a suitably tethered account. Thus, for example, early twentieth-century physics rightly accepted the positron because the alternatives—physical theories committed to the existence of negatively charged electrons with no positively charged counterparts—were less tenable than theories that postulated an as yet undetected positively charged particle. The effectiveness of such a strategy does not yet vindicate felicitous falsehoods, though. 'There exist positrons' was held to be true, although not yet independently confirmed.

How exactly can acknowledged falsehoods enhance tenability? In chapter 2, I characterized felicitous falsehoods as false in ways that do not matter. In this chapter I discuss laboratory experiments, thought experiments, and literary fictions. I argue that they all distance themselves from the facts

in order to screen out irrelevant, potentially confounding complications. This enables them to exemplify important features and patterns that would otherwise be obscured. Once we recognize and appreciate the importance of those features and patterns in artificial settings, we can test whether they afford insight into what obtains in fact. Novels and thought experiments are clearly fictions. Some are felicitous. Whether laboratory experiments are fictions is not so clear. Whatever the verdict, they are like fictions in distancing themselves from the world in order to exemplify factors that obtain but are hard to discern or whose significance is hard to appreciate in their natural settings. Although laboratory experiments, thought experiments, and fictions are unfaithful to the facts, their infidelity does not discredit them, for their contribution to understanding lies in what they exemplify, not in what they describe.

Laboratory Experiments

Every discernible item has indefinitely many discernible properties and stands in indefinitely many discernible relations. Our only hope of understanding and coping with the chaos that confronts us is to ignore most of what is there to be seen. To a considerable extent, selective disregard is automatic. In order to see anything, we overlook a lot. Where automaticity fails, we purposely block things out. We scan the forest by purposely ignoring individual trees, or focus on a tree by screening off the rest of the forest. In some cases, though, more than an act of will is required to selectively disregard what we need to. Then we may resort to experimentation. If it is unfeasible to simply pay no attention to the man behind the curtain, we may contrive a situation that excludes him.

A laboratory experiment is no mere matter of bringing nature indoors. It is a controlled manipulation of events, designed and executed to make some particular phenomenon salient. Natural entities are multifaceted. Important properties and relations are often masked by the welter of complexities that embed them. In experimenting, a scientist isolates a phenomenon from many of the forces that ordinarily impinge on it. To the extent possible, she eliminates confounding factors. She holds most ineliminable factors fixed, effectively consigning them to the cognitive background of things to be taken for granted. This enables the effect of the experimental intervention on the remaining variable to stand out. Through such a

strategy, she casts into bold relief factors that might typically be hidden from view.

Suppose a population of wild mice who were accidentally exposed to bisphenol-A later exhibited a high rate of liver cancer. To conclude that exposure to bisphenol-A caused their disease would be premature. For all we know, those mice might have been peculiarly susceptible to liver cancer or have been exposed to an unnoticed carcinogen. To glean direct, nonanec-dotal evidence of a connection between exposure to bisphenol-A and liver cancer, investigators place genetically identical mice in otherwise identi-cal environments, exposing half of them to massive doses of the chemical while leaving the rest unexposed. Because the assignment of a given mouse to one group or the other is random, the mice exposed to bisphenol-A and the mice in the control group are initially equal in expectation of develop-ing liver cancer. Their common genetic endowment and otherwise iden-tical environments neutralize a multitude of genetic and environmental factors believed to standardly influence the incidence of cancer. This blocks rival explanations that might be proposed for the elevated rate of cancer in the wild population.

The result of the experiment exemplifies the difference (if any) in the incidence of liver cancer between the two groups of mice. It not only instan-tiates the difference, it also makes that difference manifest. If the difference is statistically significant, then the result exemplifies a correlation between exposure to bisphenol-A and the incidence of liver cancer. (For simplicity, I will assume that the correlation, if there is one, indicates that exposed mice are more likely to develop liver cancer than unexposed mice. It could go the other way. Perhaps exposure to bisphenol-A protects against liver cancer.) Although correlation does not imply causation, a robust correlation often provides strong evidence of causation. In this case, the background assump-tion that moves us from a mere correlation to a causal judgment is the well-founded conviction that the experiment was so rigorously designed and executed that nothing but the exposure to bisphenol-A could have caused the difference. That being so, the result may also exemplify a causal relation.

So far we are just talking about the particular mice in the experiment. But the goal of the investigation is not primarily to discover their medical fates. It is to use their medical fates to learn something more general. Since the mice in the experiment were chosen arbitrarily from the class of mice with

a particular genome, it is straightforward to project to other mice of the same strain. The experiment then also exemplifies the increased propensity of mice of that strain to develop liver cancer when exposed to bisphenol-A. Moreover, the mice are model organisms, so there is reason to think that what holds for them also holds for the organisms they serve as models for—in this case, mammals, including humans. The experiment's result instantiates propensities to develop cancer in a variety of groups: the mice in the experiment, other mice of the same strain, mice in general, rodents, mammals. It also instantiates a vast and variegated host of other properties. But given the background assumptions that motivate and shape the inquiry, it exemplifies the propensity of mammals exposed to bisphenol-A to develop liver cancer. The inference from mice to mammals is a big jump. But if the background assumptions legitimating treating the mice as model organisms are cogent, it is reasonable to treat the experiment as exemplifying a causal connection between exposure to bisphenol-A and cancer in mammals in general.

The interpretive path I just sketched is reminiscent of Capdevila's discussion of *Rosso Galera, Rosso Guzzi.* Here the connecting links derive from medical science rather than Italian motorcycle culture. The interpretation I've given is mediated by background assumptions drawn from our current best understanding of carcinogens, of mammals, and of methods for investigating carcinogenicity. That understanding may be wrong in ways that would vitiate my interpretation. So it is fallible. But if the understanding is sufficiently accurate and adequate (even if not true in every respect), the interpretation is too. If the exposed mice show a sufficiently elevated rate of liver cancer, we are right to think that bisphenol-A is carcinogenic.

What if we are wrong? Exemplification requires instantiation. Although we have no reason to think so, suppose there is a sharp threshold. Exposure below a certain level is causally inert. Above that level, cells go wild. Then the consequence of a small animal's abrupt exposure to massive doses of bisphenol-A over a short period of time does not exemplify what happens to a large animal exposed to small doses over a long period of time. Given that humans are never exposed to levels or spikes in level comparable to those of the mice, the experiment does not exemplify a danger to humans. This could be so. An important mediating assumption might be false. In that case, the result does not exemplify anything about human vulnerability to cancer.

Still, the result exemplifies a connection that justifies our thinking that exposure to bisphenol-A increases the likelihood of developing cancer. Even if the result is misleading, it affords insight into the structure of our current understanding of the subject. The attribution of the difference in incidence of cancer to exposure to the chemical is reasonable to the extent that the scientists contrived a situation where rival explanations of the difference between the exposed mice and the control group have been blocked. The projection to other mammals is reasonable to the extent that the grounds for taking the mice to be model organisms are sound. Because the experiment takes place against a cluster of fallible background assumptions, it does not afford conclusive evidence. But because of its rigorous controls, it affords stronger and more direct evidence than a mere correlation between exposure and cancer in a wild population would.

Designing an experiment is setting a stage where events can play out. Conducting an experiment involves initiating and perhaps intervening in a course of events. Experiments are dynamic (Nersessian, 2007). They unfold as they do in part because of the scientist's actions. She instigates and perhaps interrupts, deflects, impedes, or amplifies a natural sequence. She may isolate phenomena from their normal concomitants and introduce unusual provocations. The mice in the experiment are exposed to massive doses of bisphenol-A. This allows the scientist to obtain a pronounced effect in a relatively short time. Her working assumption is that a small mammal's exposure to a large dose over a short period is equivalent to a large mammal's exposure to a lower dose over a longer period. If that assumption is true enough, she can safely extrapolate from the laboratory situation to mammals in general, including humans.

Experiments often involve creating and using items that are nowhere to be found in nature (Cartwright, 1983; Hacking, 1982). Genetically identical mice are artifacts; their genetic makeup is designed to suit the sorts of experiments in which they will be used. Pure forms of chemicals are artifacts as well, being synthesized under carefully controlled conditions to avoid contamination. This can be crucial.

The Miller—Urey experiment begins with chemicals believed to be present on Earth in prebiotic times. The experiment consists of a sequence of chemical reactions whose ultimate output consists of new organic chemicals and amino acids. It thus shows how life could have emerged from nonliving matter (Ball, 2005). For the experiment to work, the chemicals—methane,

ammonia, hydrogen, and water—had to be pure. Any hint of contamination would discredit the result. Moreover, to ensure that no organic material was accidentally introduced during the course of the experiment, the chemical processes had to be completely isolated from the environment. Although—indeed because—the experimental components and conditions were unnatural, the experiment revealed something important about the natural world. It did so not by saying that something is the case, but by showing it to be the case. It exemplified a path from four initial chemicals to amino acids.

The sequence of chemical reactions that took place in the Miller—Urey experiment may have been instantiated not only at the dawn of life, but innumerable times since. According to Goodman, any of those instantiations could exemplify the emergence of organic compounds from the four original chemicals. Just ignore all the other chemicals in the neighborhood; ignore the prevalence of oxygen and nitrogen in the atmosphere; concentrate on the reactions due entirely to hydrogen, methane, ammonia, water, and electricity. But such a natural instantiation of the sequence of reactions would be untrustworthy. The possibility of contamination would inevitably loom large. What distinguishes the experiment from other instances of the sequence is that it constitutes a context where it is manifest that nothing but the four chemicals and a bit of electricity was necessary to initiate the process. The experiment thus affords not just an instance, but a telling instance. By exemplifying that those chemicals and electricity suffice to generate new organic compounds, the experiment affords genuine insight into an aspect of nature.

An experiment unfolds over time. It has a narrative structure (Nersessian, 1993), with a well-defined beginning, middle, and end. Like a Greek tragedy, the experiment begins *in medias res*. The scientist conducts his investigation against a background of established findings and shared assumptions that frame the events and circumscribe their interpretation. The Miller—Urey experiment takes place against a cluster of background assumptions about chemical reactions. Its narrative arc starts with the enclosure of the pure chemicals in a carefully crafted, tightly sealed apparatus; it develops through the heating of the water and the emission of an occasional spark that mimics the effect of lightning; and it climaxes when the reactions are complete. The denouement consists in extracting

the resulting solution and subjecting it to chromatography to determine its chemical composition.

Experimental results do not speak for themselves. They require interpretation. Their interpretation draws on background assumptions, beliefs about instrumentation, experimental design, the course of events that constitute the experiment, and the outcome. A change in any one of these factors can prompt revisions in others. Against the background of the theory of relativity, we interpret the Michelson—Morley experiment as affording evidence of the nonexistence of luminiferous ether; prior to the advent of the theory of relativity, it afforded evidence that the ether might not exist; long after most physicists accepted the theory of relativity, Morley took the experiment to demonstrate that his interferometer was not sensitive enough to measure ether drift, which he still believed was there to be measured.

These features of experiments are well known. I mention them to highlight how distant many scientific experiments and their results are from the natural phenomena they illuminate. The items experimented upon are often artifacts constructed expressly for experimentation (see Cartwright, 1983). The circumstances in which they are placed are artificial; they are carefully contrived situations, often ones that do not naturally occur but that are designed expressly to exemplify telling features of the phenomena. For an experiment to disclose something about a range of phenomena, it must exemplify features it shares with those phenomena. But it may, and in some cases must, diverge from the phenomena in other important respects.

'Natural experiments' take advantage of natural, or anyway unarranged, distributions that mimic the laboratory scientist's random assignments to treatment and control groups. The subjects are populations that are suitably alike except for exposure to the factor under investigation. John Snow's study of the London cholera epidemic can be seen as a natural experiment in that, as far as anyone can tell, but for their source of water, the people who used the Broad Street pump were suitably similar to the denizens of the workhouse, who used a different pump, and the workers in the brewery, who did not drink water. Those two groups served as controls, enabling Snow to identify the pump as the source of the epidemic and contaminated water as the cause of cholera. Natural experiments are welcome where it is immoral or unfeasible to conduct controlled experiments. But they are

less trustworthy than laboratory experiments, for they rest on the optimistic conviction that the natural arrangement controls for all the plausible sources of variation between the two subpopulations.

For scientific results to be acceptable, they must be replicable. Clearly not every aspect of the original experiment must be reproduced. That would be impossible. Never again can an experiment instantiate *being the first time* a particular phenomenon was observed. Failing to replicate that feature is no flaw. Other features could be replicated, but there would be no epistemic value in doing so. Although the original experiment was conducted by a left-handed, squash-playing postdoc, it would be ludicrous to insist that only a left-handed, squash-playing postdoc could replicate it. Evidently, some features of an experiment are irrelevant and need not be replicated. But features that are irrelevant to one experiment may be relevant to another. Ordinarily, for example, the altitude at which an experiment is conducted is irrelevant. So many experiments originally performed in Mexico City can be replicated in New Orleans. But where altitude matters, an experiment performed in New Orleans would not qualify as a replication of one done in Mexico City.

Replication requires that subsequent experiments match the first in every respect that is pertinent to the hypothesis being tested. The second experiment must exemplify the same features as the original when both are interpreted as tests of the same hypotheses. This allows for considerable variation in the other features the various experiments instantiate. Indeed, such variation is desirable. If subsequent experiments instantiate the same unexemplified features as well as the same exemplified ones, their agreement might be misleading. Factors currently deemed irrelevant might in fact account for their common result. Instead of seeking to reproduce the original experiment as closely as possible, it is methodologically preferable to perform a variety of experiments that reproduce the features that the original exemplifies but differ from the original and from one another in those that are not exemplified (Elgin, 1988, 89).

Experiments are conducted; they do not just happen. They have a narrative structure. They are subject to interpretation and to reinterpretation if background assumptions change. They are repeatable. In short, they are close kin to dramatic enactments. This is not quite to say that experiments are works of fiction; but it is to suggest that the gulf between what we

consider fact and what we consider fiction may be narrower than is typically supposed (see Cartwright 1983).

Nevertheless, one might think, there is a crucial difference. Unlike the 'events' in a novel or thought experiment, the events constituting an experiment actually occur. They are processes that real things really undergo. The real mice exposed to bisphenol-A really exhibited an elevated rate of liver cancer. The Miller–Urey experiment produced real amino acids from real chemicals. The worry is that, unlike a standard experiment, a mere simulation—be it a computer simulation, a thought experiment, or a work of literary fiction—might omit something significant. In that case, what occurs in the simulation is not representative of what occurs in reality. But the same holds when scientists move from the field to the lab. The controls introduced in the lab might inadvertently omit something significant. If so, the result does not directly project onto the phenomena it concerns. The trustworthiness of a laboratory experiment, a natural experiment, or a thought experiment depends on the adequacy of the background assumptions that frame it and the scrupulousness with which the experiment is designed and executed.

Thought Experiments

In laboratory experiments, scientists simplify, streamline, manipulate, and omit, so that the effects of potentially confounding factors are minimized, marginalized, or canceled out. An experiment deliberately departs from nature in order to advance an understanding of nature. Rather than invalidating the experiment, this departure is what enables it to disclose barely detectable or standardly overshadowed aspects of nature. Thought experiments involve further distancing. They are not actual, and often not even possible, experiments. They are imaginative exercises designed to disclose what would happen if certain, perhaps unrealizable, conditions were met.

Their reliance on imagination may give hard-nosed epistemologists pause. The imagination is, after all, free to entertain any ideas it likes. It is not bound to respect conceptual connections, evidence, laws of nature, or the dictates of common sense. Familiar scientific thought experiments violate all of these. But as Kant (1981) emphasizes, freedom is not lawlessness. Freedom consists in being bound by laws we set for ourselves—laws we reflectively endorse as reasonable and rational. As the locus of the free

play of ideas, the imagination is not a realm in which ideas are utterly uncon-
strained, bouncing off one another like gas molecules in random motion. It is
a realm in which the play of ideas is bound by constraints the imaginer sets.
Although the constraints are self-imposed and vary from one imaginative
setting to the next, they are real. The power of thought experiments to illu-
minate the facts lies in no small measure in the flexible, variable, but none-
theless binding character of the constraints that the imagination imposes
on them. By setting such constraints and drawing out their consequences,
the imagination serves as a laboratory of the mind, a venue in which
hypotheses can be contrived, elaborated, and tested. Moreover, in scientific
thought experiments, the constraints, even if tacit, are recognized, shared,
and reflectively endorsed by a scientific community. Even so, how can a
thought experiment claim to yield any insight into the facts? Why isn't it
simply an exercise in fantasy?

Unlike fantasies, thought experiments are not essentially private; nor
are they particularly mental. Although they are imaginative exercises, they
are publicly articulated, discussed, illustrated and disputed. They consist
of verbal or pictorial representations. Their claim to be imaginative stems
from the fact that, like works of fiction, they are typically not, and in any
case need not be, representations *of* anything real. But the unreality of the
objects that ostensibly figure in them does not undermine their function.

In designing a laboratory experiment, a scientist may begin by perform-
ing something like a thought experiment. She runs through the expected
course of events in her head and perhaps describes it to her research team
before attempting to implement it in the lab. This suggests that the differ-
ence between real experiments and thought experiments lies in the fact
that thought experiments sharply truncate the experimental process. They
omit the implementation step. But scientists can't do this whenever they
please. The question is when is stopping short legitimate?

Sometimes an actual experiment of the sort envisioned *cannot* be carried
out. It is impossible or impracticable. As we saw, Einstein demonstrated the
equivalence of gravitational and inertial mass via a thought experiment
that could not be carried out. It would require sending an unconscious
subject in a windowless enclosure into a region of deep space distant from
any significant source of gravity, rousing him, and querying him about his
experiences. This is morally, practically, and physically unfeasible. Still, the
recognition that we cannot do a real experiment does not by itself legitimate

the results of stopping short. Sometimes, the infeasibility of an experiment translates into the infeasibility of finding out a particular fact. The reason Einstein's thought experiment is effective is that it takes the form of a challenge: Suppose the specified conditions were met. How could a subject tell whether he was in one situation or the other? If our best efforts to identify a way to tell the difference fail, and fail for scientifically principled reasons, we have evidence of the equivalence. Our failure indicates that, if our theories are close to correct, there is no difference to detect.

Sometimes the imaginative rehearsal reveals that an actual experiment *need not* be carried out. The mental run-through itself discloses the relevant information. Without physical implementation, Galileo's thought experiment discredits the Aristotelian contention that the rate at which bodies fall is proportional to their weight. According to Aristotle, a boulder should fall more quickly than a pebble, assuming they are composed of the same material. What, Galileo wonders, are the implications for a composite object? He imagines an object consisting of a boulder tethered to a pebble. Being composed of two rocks and some rope, the composite object is heavier than either rock alone. If Aristotle is right, it should fall more quickly than the boulder. But since, according to Aristotle, the pebble falls more slowly than the boulder, once the two are roped together, the pebble should retard the boulder's fall. It should serve as a brake. Hence the rate at which the composite object falls should be between that of the boulder and that of the pebble. The composite object cannot fall both more quickly and more slowly than the boulder, so the Aristotelian commitments are inconsistent. By exemplifying the inconsistency, Galileo's thought experiment demonstrates that the Aristotelian account cannot be correct.

One might argue that Galileo's thought experiment discredits my analysis.[1] Exemplification, I said, requires instantiation. Real chemical reactions occur in the Miller–Urey experiment. So it is plausible that by exemplifying those reactions, the experiment affords epistemic access to them, enabling us to recognize them and appreciate their significance, not only in the experimental setting but also outside of it. In a mere thought experiment, however, no physical changes actually occur. A thought experiment, not being material, cannot exemplify material properties. This is so. The sequence of ideas that constitutes Galileo's thought experiment does not instantiate material properties of falling bodies. There is, for example, no danger of being hit on the head by the falling rocks. But the *rate* at which

bodies fall and the *independence of that rate* from the weight of those bodies are abstract, mathematical properties. They can be instantiated by material and immaterial sequences alike. There is then no bar to saying that via exemplification thought experiments afford epistemic access to abstract properties that are in some cases instantiated in material objects. A thought experiment is a representation—a re-presentation—of abstract features, an imaginative re-embodiment of them. We are to imagine—that is mentally, verbally or pictorially present—a situation where the abstract features are realized. In effect, we are to investigate what would happen in a virtual reality where certain constraints are supposed to hold.[2]

Philosophers sometimes think that we resort to thought experiments only when, for one reason or another, a real experiment cannot be carried out. Perhaps Galileo could not have conducted a real experiment to conclusively demonstrate his point. Maybe he did not have sufficiently accurate timers or a high enough tower from which to run the test. Maybe he did not have the resources to eliminate the effects of air resistance, and so on. Now, however, we could conduct the experiment. Shouldn't we? Probably not. Rather than concluding that the thought experiment was a second-best strategy resorted to because of circumstances beyond the scientist's control, we should recognize that a real experiment would not have made Galileo's case any more forcefully than his thought experiment did. It would simply have muddied the waters. Once we start dropping objects from towers, we face the problem that cancer-ridden wild mice pose for biologists and the emergence of amino acids in nonisolated situations pose for chemists. How do we know that unrecognized confounding factors do not explain our finding? By deploying an austere thought experiment where the distance and duration of the fall, the presence or absence of air resistance, and a host of other possible sources of interference are simply omitted, Galileo blocks such challenges. The thought experiment demonstrates an inconsistency in the Aristotelian position—an inconsistency that would obtain regardless of the conditions under which the experiment was conducted. Just as a laboratory experiment is preferable to a 'natural experiment', the thought experiment is preferable to an actual experiment, because it is invulnerable to a host of potentially misleading challenges that an actual experiment would face.

Even in the empirical sciences, not every question can or need be answered by direct appeal to observational evidence. Thought experiments

are often appropriate where observation is not apt. This leads some to conclude that thought experiments are effective only where the issues concern conceptual or theoretical commitments. If so, thought experiments disclose something about our concepts and our theories, not something about the world. Since we can tease out commitments in the armchair, it is no surprise that mere thought experiments are effective for revealing them. But to conclude that thought experiments do not reveal anything about the way the world is would be too hasty. Galileo's thought experiment did not just exemplify an inconsistency in Aristotle's theory. It also showed that *any* theory that took the rate at which bodies fall to depend on their weight would be inconsistent. From this it follows that the rate at which bodies fall is independent of their weight. The thought experiment thus exemplifies a feature of reality, not just of theoretical or conceptual commitments. That feature is, or is a consequence of, a modal fact. The rate at which objects fall is independent of their weight because they could not fall any other way. Thought experiments, it seems, afford epistemic access at least to theoretical, conceptual, and modal matters.

A thought experiment fixes certain parameters (e.g., about the relevant laws of nature and the supposed initial conditions), provides a description of the experimental situation that sets out all and only the features considered relevant, and works out the consequences. Galileo's two rocks are assumed to be made of the same material and to be the same shape, thereby obviating the effects of material constitution and shape on the outcome. They are assumed to be falling through the same medium. Weight is assumed to be additive, and a tethered material object is assumed to be subject to the same laws of nature as untethered ones (Gendler, 2010). In effect, the thought experiment invites us to consider what would happen if certain conditions, some expressly specified and some tacitly assumed, obtain.

Like literary fictions and ordinary experiments, thought experiments have a narrative structure. We perform thought experiments by imagining a scenario in which something happens—a sequence of events with a beginning, middle, and end. Thought experiments can be construed as tightly constrained, highly focused, minimalist fictions, like some of the works of Borges. If Borges's minimalist stories are genuine fictions, there seems no reason to deny that thought experiments are too.

To understand a thought experiment requires a suspension of disbelief. We grant its (tacit and explicit) assumptions even though we know that

they do not—and in some cases cannot—obtain. Maxwell's demon provides a telling instance. According to the second law of thermodynamics, when ensembles of gases at different temperatures are brought into contact with one another and isolated from the wider environment, they will evolve to thermodynamic equilibrium. That is, the resulting ensemble will reach a stable, uniform temperature between the two original temperatures. Maxwell's question is whether this law is inviolable. To answer it, he mounts this thought experiment. Consider a vessel of air at uniform temperature. Its individual air molecules move at different velocities, even though the mean velocity of any sufficiently large arbitrary collection of those molecules is almost exactly uniform. Divide the vessel into two sections, *A* and *B*. Now imagine a demon capable of seeing and tracking individual molecules. He monitors a door between *A* and *B*, allowing only fast molecules to move from *A* to *B* and only slow molecules to move from *B* to *A*. Eventually most of the fast molecules will be in *B* and most of the slow ones will be in *A*. Simply by controlling access across the divide, the demon will raise the temperature of *B* and lower the temperature of *A*, thereby violating the second law of thermodynamics. Although we know full well that no one is, or arguably could be, endowed with the abilities ascribed to Maxwell's demon, we bracket that knowledge and see whether such a being could defy entropy. Considerable ingenuity may be required to figure out which beliefs should be suspended and which ones should be retained. This is why scientific thought experiments are embedded in theoretical discussions that fix their parameters. And it is why their implications are subject to dispute. Still, we are inclined to agree with Maxwell that the demon thought experiment shows that the second law of thermodynamics is in principle violable. It is a merely statistical law.

Like literary works and ordinary experiments, thought experiments require interpretation. Sometimes interpretations diverge. The Einstein–Podolsky–Rosen experiment is a case in point. Very roughly, the scenario is this: Two particles interact, then fly off in opposite directions. Once they are separated, the measurement of one should have no effect on the state of the other. But if we measure, say, the position of one and apply the Schrödinger equation, we can determine that the other also has a definite position. This seems to violate the uncertainty principle: being unexamined, the second particle should have no definite position. What does the thought experiment

show? The answer is not at all clear. Apparently, even the authors disagreed. Podolsky thought it demonstrated that quantum mechanics is incomplete, while Einstein thought it showed either that quantum mechanics is incomplete or that states of spatially separated objects are not independent of each other (Bokulich, 2001).

To recap: a thought experiment is an imaginative exercise designed to investigate what would happen if certain conditions were satisfied. Conducting it requires a suspension of belief, in that the conditions imagined are not realized in fact, and may be inconsistent with conditions we know to obtain in fact. It requires a suspension of disbelief, in that it asks us to entertain scenarios that we know do not and often could not obtain. It depends on background assumptions about what commitments are to be retained, what commitments are to be relaxed, and what commitments are to be abandoned in entertaining the imaginative scenario. Whether the constraints are tacit or explicit, in conducting the thought experiment the epistemic agent is bound by them. A thought experiment has a narrative structure, with a beginning, middle, and end. It is subject to interpretation, and to reinterpretation if the background assumptions change. Schrödinger's cat, originally introduced to criticize the Copenhagen interpretation, now appears in every interpretation of quantum mechanics, each offering its own account of the poor beast's state. Finally, we saw that thought experiments are valuable in investigating what is not open to direct empirical inspection—conceptual or theoretical commitments and their consequences, as well as modal properties, and, I would add, relations of cotenability on noncotenability.

Fictions

The thought experiments I have mentioned so far have been drawn from the physical sciences. But thought experiments are ubiquitous in philosophy as well. Like scientific thought experiments, those in philosophy illuminate factors that are not readily accessible to direct inspection. Because of a difference in subject matter, the range of factors illuminated by philosophical thought experiments is broader. Familiar philosophical thought experiments afford insights into normative properties (trolley problems, the experience machine), introspectively available properties (Mary, brains

in a vat), and metaphysical properties (fission and fusion in personal identity, the ship of Theseus). No more than conceptual, theoretical, and modal properties, are these open to direct empirical inspection.

Like scientific thought experiments, many philosophical thought experiments are fairly austere. But philosophical thought experiments often are not so tightly bound by theoretical constraints as scientific thought experiments tend to be. Thus there is more controversy over what we should conclude from them. (Confession: I have no idea what the Chinese Room shows.) Many are relatively autonomous. We can fruitfully entertain them against the background of multiple sets of philosophical assumptions—indeed without even being aware that we are making philosophical assumptions. The trolley problem was introduced to disclose a consequence of the doctrine of double effect. It now has a life of its own (and seemingly innumerable children and grandchildren).

If an austere thought experiment can afford epistemic access to a range of properties, and can do so in a context that is not tightly beholden to a particular theory, there seems to be no reason to deny that a more extensive thought experiment can do the same. Indeed, philosophy provides examples. Plato's *Republic* and Rousseau's *Emile* are cases in point. This opens the way to construing works of literary fiction as extended, elaborate thought experiments. They afford epistemic access to aspects of the world that are normally inaccessible—in particular, to the normative, psychological, and metaphysical aspects that philosophical thought experiments concern.

Again one might worry about the capacity of exemplification to account for this. A work of fiction, not being alive, cannot instantiate psychological or moral properties. If it cannot instantiate them, it cannot exemplify them. But it can instantiate and exemplify abstract properties that are concretized in human agents. Suppose Meg instantiates a pattern of psychological features—a network of beliefs, desires, aversions, and preferences, for example. Although the specific elements of her network are psychological, the pattern is abstract. In principle, it can be instantiated by something other than psychological elements. A fiction writer might create a scenario where that pattern is instantiated and exemplified via a sequence of descriptions. In effect, she takes a pattern that is embodied in fact, abstracts it, and re-embodies it in fiction. (Or, more likely, she abstracts individual elements instantiated in fact, finds or devises an appropriate pattern, and embodies that pattern in fiction.) Strictly, in the fictional setting it is not a pattern *of*

psychological features. But it is a pattern that is, or that may be, instantiated by psychological features. So it affords epistemic access to a pattern that we may find ourselves or our fellows instantiating.

Like an experiment, a work of fiction selects and isolates, manipulating circumstances so that particular properties, patterns, and connections, as well as disparities and irregularities, are brought to the fore. It may localize and isolate factors that underlie or are interwoven into everyday life or natural events, but that are apt to pass unnoticed because other factors typically overshadow them. This is why Jane Austen maintained that "three or four families in a country village is the very thing to work on" (1814). The relations among the three or four families are sufficiently complicated and the demands of village life sufficiently mundane that the story can exemplify something worth noting about ordinary life and the development of moral character. By restricting her attention to three or four families, Austen in effect devises a tightly controlled thought experiment. Drastically limiting the factors that affect her protagonists enables her to elaborate the consequences of the relatively few that remain.

If our interests are cognitive, though, it might seem that this detour through fiction is both unnecessary and unwise. Instead of resorting to fiction, wouldn't it be cognitively preferable to study three or four real families in a real country village? Probably not, if we want to glean the insights that Austen's novels afford. Even three or four families in a relatively isolated country village are affected by far too many factors for the social and moral trajectories that Austen's novels exemplify to be salient in their interactions. Too many forces impinge on them and too many descriptions are available for characterizing their interactions. Any such sociological study would be vulnerable to the charge that unexamined factors played a nonnegligible role in the interactions studied, that other forces were significant. Austen evades that worry. She omits such factors from her account and in effect asks: suppose we leave them out, then what would we see? Similarly, the model pendulum omits friction and air resistance, allowing the scientist in effect to ask: suppose we leave them out, then what would we see?

The question is how what we see in the fiction is supposed to inform our understanding of reality. That Elizabeth Bennet and Mr. Darcy, who do not exist, are said to behave thus and so does not demonstrate anything about how real people really behave. That an idealized pendulum, which also

does not exist, is said to behave thus and so does not demonstrate anything about how actual pendulums behave.

Recall the paint company's sample cards discussed earlier. Most people speak of them, and probably think of them, as samples *of* paint—the sort of stuff you use to paint the porch. They are not. The cards are infused with inks or dyes that are the same color as the paints whose colors they exemplify. It is a fiction that they are samples of paint. But since the sole function of such a card is to convey the paint's color, the fiction is no lie. All that is needed is something that is the same color as the paint. A fiction thus conveys the property we are interested in because, in the respect that matters, it is no different from an actual instance. The exemplars on the paint cards need not themselves be paint. Similarly in literary or scientific cases. If the sole epistemic objective is to exemplify particular properties, then in a suitable context, any symbol that exemplifies those properties will do. If a fiction exemplifies the properties more clearly, simply, or effectively than a strictly factual representation, it is to be preferred to the factual representation.

In *Nicomachean Ethics*, Aristotle (1985) suggests that we should call no man happy until he is dead. Initially this seems implausible. As is well known, part of the problem is that what Aristotle means by 'happy' is not what we mean. 'Flourishing' would be a better term. But 'call no man flourishing until he is dead', even if not quite so implausible, still seems a bit extreme. Surely, one wants to object, we can easily discern that some of our fellows are currently flourishing. Aristotle defends his idea by contending that severe enough reversals of fortune late in life would justify the conclusion that a man's life had not been a happy (or flourishing) one. Maybe so. But even if someone suffered serious misfortunes late in life, it is tempting to object, 'Well he was happy (or flourishing) up until then'. Aristotle's own example of Priam, the elderly king of Troy, invites this response. Until the Trojan War, Priam prospered. His life ended in misfortune; but throughout most of it, he apparently flourished. This is the objection students typically raise, and if one just reads Aristotle, it does not seem unreasonable.

Oedipus Rex can be read as a thought experiment that vindicates Aristotle's claim. For most of his life, Oedipus, like Priam, seemed blessed with the gifts of fortune; and as far as anyone could tell, he lived a life of Aristotelian virtue. Evidently, he flourished and deserved to flourish. When Thebes suffered a plague, which the oracle blamed on him, Oedipus discovered that

he had unwittingly killed his father and married his mother. This discovery did not just doom his future happiness, as the sack of Troy doomed Priam's. It discredited his past happiness. He had, through no fault of his own, been living a lie. We might exonerate him for the wrongs he had done, since he acted out of nonculpable ignorance. But even if he was blameless, his relations to himself, to his wife/mother, to his children/siblings, to the citizens of Thebes who suffered for his iniquities, and to his own past, were forever changed, and would henceforth be tinged with revulsion. It turns out that he had not been flourishing during the early years, even though he and everyone else thought that he was.

Oedipus Rex is a work of fiction that advances our understanding not only of Aristotle's ethics, but also of the human predicament. It underscores the limits on human knowledge and the vulnerabilities that stem from those limits; the value of knowing oneself and one's situation, and the limits on the human capacity to do so. It does not constitute a proof that Aristotle is right, but it poses a challenge: you should be wary of calling a man flourishing during his lifetime unless you are sure his situation is not like Oedipus's. It is hard to see how the challenge can be met.

Kant (1981) maintains that not only is it impossible to know whether someone else has acted morally, it is impossible to know whether you yourself have done so. For any action that accords with the categorical imperative, there is always an available self-interested maxim that might have been the real motive. So it is impossible to glean unequivocal empirical evidence for moral action. But even if we can never discern that someone has behaved morally (that is, acted on account of respect for the moral law), an author can effectively stipulate it. In *A Tale of Two Cities* Dickens portrays Sidney Carton's self-sacrifice as stemming from purely moral motives. An author can portray a situation, develop a character, and convey his thoughts, feelings, and reasons for acting, thereby blocking the explanation from self-interest. He can, that is, exemplify a pattern and demonstrate that were that pattern to be instantiated, it would be an instance of acting morally.

If Kant is right about the impossibility of finding unequivocal instances of acting morally, how do we learn what a moral action is? Even if the categorical imperative is a deliverance of reason, we need to know how to apply it. Non-Kantians face a similar problem. Everyone recognizes that it is difficult in practice to distinguish acting morally from acting out of

self-interest. If I am right, fiction can play a major role in moral education. By exemplifying genuinely moral patterns of deliberation and action, it can teach us something that is at least hard, and might be impossible, to reliably discern in real-life interactions.

People have inner lives replete with motivations, perceptions, emotions, and thoughts. Because a variety of combinations of psychological elements might yield the same outward behavior, it is impossible to uniquely determine the underlying psychological states from observations of overt behavior alone. Moreover, people often misunderstand themselves, and often have reason to mask what they think and feel. So even if we ask them and they tell us, we ought not be confident that we know how things are with them. This raises the question: how do we learn (indeed, what even leads us to suspect) that other people have inner lives that are quite unlike ours? Their feelings may not only be unlike what we actually feel about things, but may also be unlike what we would feel if we were in their place. It is one thing to be able to imagine oneself having had someone else's experiences. That is hard enough. But we can do things that require yet more imaginative dexterity. Even if Jim recognizes that had he been treated as Jane was, he would be bitter, he may also understand why she is willing to let bygones be bygones. This is an amazing cognitive achievement. I suggest that fiction plays a major role in equipping us for this kind of feat. In reading a work of fiction we take up a point of view and try it on for size. In effect, we experiment with the perspective and see how things look from there. Many works portray the world through a protagonist's eyes, conveying her experiences, feelings, and thoughts. They disclose the limitations of her perspective. Some do more. A work may afford multiple perspectives on the same series of events, disclosing the resources and limitations of each. In effect, each filters events through a different sieve.

Philosophers as well as nonphilosophers have a tendency to take the fruits of introspection at face value, or at least to grant them a higher epistemic status than outsiders' opinions about a subject's state of mind. A work like *Lolita*, written from the perspective of an utterly unreliable narrator, affords insight into the limits of introspection. A character's perspective can be so skewed or benighted that he is simply wrong about the central events of his life. Such a work can be construed as a thought experiment that undermines the conviction that a person's access to his own motives, beliefs, and other attitudes always affords better evidence than the evidence

that one's words and actions afford to others. Perhaps there is privileged access in the sense that each of us knows herself *in a way* that she knows no one else. But self-deceptive fictional characters undermine the conviction that we always know ourselves *better* than others know us.

Cavell maintains that the problem of other minds is not, or not only, the problem of ascertaining whether an entity has a mind, but the problem of figuring out what is on someone's mind. Even if I am confident that an individual has beliefs, desires, preferences, and feelings, I am woefully underequipped to identify those fine-grained mental states. The problem is not just a problem about other minds, though. I may be equally underequipped to know what my mental states are. Do I really expect I'll get the paper done, or have I managed to deceive myself into taking a hope for an expectation? Do I really want the promotion, or do I just think I want it because I know that it is the sort of thing that people in my position are supposed to want? As Cavell (1987) reads Shakespeare's tragedies, they afford evidence of the uncertainty of mental state ascriptions. For the same sorts of reasons that Lear cannot recognize Cordelia's devotion, that Othello cannot recognize Iago's malevolence or Desdemona's fidelity, that Hamlet cannot trust his judgment, we cannot be sure of the mental states that we ascribe.

A work of literature can also function as something akin to an impossibility proof—a thought experiment that exemplifies the inadequacy of its own grounding assumptions. Davenport (1983) reads *Middlemarch* as a thought experiment about marriage. Both Dorothea Brooke and Dr. Lydgate are in deeply unhappy marriages. Because in the world of the novel divorce is unthinkable (and unthought of), they are doomed to serve life sentences for their unwise choices of mates. By exemplifying the intractability of the problem they face, the novel provides reason to think that divorce, or something like it, should be an option. Davenport considers *Middlemarch* flawed because it does not allow for the possibility of divorce. I disagree. I consider it a powerful thought experiment that reveals the unacceptable consequences of institutional arrangements that do not allow for exit visas.

Metaphysical thought experiments are often science fictional. Some are so austere that in their philosophical settings we do not know what to think. Literary and cinematic fictions help us out. What should we make of Putnam's brains in a vat? *The Matrix* supplies an answer. What would a computer that passed the Turing test be like? His name is Hal. Could

beings without inner lives actually be indistinguishable from us? The way to settle such matters (even tentatively and revisably) is to design a scenario in which the consequences of such hypotheses play out. Write a story about the love lives of zombies, or about the lives of zombies incapable of love. We may find that our off-the-cuff intuitions do not stand up under elaboration or that the consequences of our assumptions are quite different from what the austere philosophical thought experiments led us to suppose.

Despite what I have said, the idea that fictions function as thought experiments that afford insight into actual human experience may seem a stretch. I do not think that it is. An example shows why. One of the mysteries of the Penn State pedophilia scandal is how Joe Paterno, the longtime football coach whom many considered a bastion of integrity, could have turned a blind eye to the actions of his assistant, Jerry Sandusky. Sportswriter Thomas Boswell ventures the following answer:

Everybody has weak spots in their character, fault lines where the right earthquake at the wrong time can lead to personal catastrophe. Most of us are fortunate that our worst experience doesn't hit us with its biggest jolt in exactly the areas where our flaws or poor judgment or vanity is most dangerously in play. It's part good luck if we don't disgrace ourselves. But when it does happen, as appears to be the case with Joe Paterno, that's when we witness personal disasters that seem so painful and, in the context of a well-lived life, so unfair that we feel a deep sadness even as we simultaneously realize that the person at the center of the storm can never avoid full accountability. . . . Forces collide, conspire, confuse, and an icon of integrity fails to act, fails to see. (2011)

If this sounds familiar to those who do not read the sports pages, it is because the passage is a precis of *Oedipus Rex* as filtered through Aristotle's *Poetics* (1973).[3] The great man, beset by hubris, does terrible things and is brought down by his tragic flaw. One can quarrel with Aristotle's reading of Sophocles or with Boswell's implicit endorsement of that reading. One can doubt that Paterno was the man of integrity he was alleged to be. One can even think that ignoring rampant pedophilia is vastly worse than inadvertent patricide and incest among consenting adults. Still, *Oedipus Rex* provides a template for understanding Paterno. Having seen the pattern in fiction, we are in a position to entertain the possibility that it affords an understanding of what happened in fact.

One objection to the thesis that fictions function epistemically is that they provide no evidence about the way the world is (Carroll, 2002).[4] Davies (2010) contends that my account is vulnerable to this objection. Whether

he is correct depends on what it is to provide evidence. According to David Lewis,

We who have lived in the world for a while have plenty of evidence, but we may not have learned as much from it as we could have done. This evidence bears on a certain proposition. If only that proposition is formulated, straightway it will be apparent that we have very good evidence for it. . . . If we are given a fiction such that the proposition is obviously true in it, we are led to ask: and is it also true *simpliciter*? And sometimes, when we have plenty of unappreciated evidence, to ask the question is to know the answer. (1983, 279)

Like Lewis and Davies, I maintain that fictions equip us to recognize and marshal information we already have. This enables us to treat that information as evidence. Whether this amounts to providing evidence or merely to providing access to evidence we already have makes no difference. Either way, they advance understanding of the actual world.

Fictions also exemplify abstract features and patterns, affording direct evidence of them, and indirect evidence that their real-world instantiation is possible (Elgin, 1996). Davies maintains that this feature makes fiction a source of hypotheses, rather than evidence of a truth. Perhaps so. But fictions do not just generate hypotheses. By exemplifying features and patterns that, if instantiated, are significant, they also provide reasons to think that the hypotheses are worth taking seriously. Some scientific simulations do the same. I see no reason to exclude symbols that play this role from the epistemic realm. If a work of fiction can afford reason to suspect that a pattern obtains in fact, or that discovering whether it obtains would be epistemically worthwhile, the fiction functions epistemically, whether or not we cash out 'reason to believe' in terms of evidence. A fiction can then render a hypothesis initially tenable.

Not every falsehood is felicitous. So the fact that we can write a fiction that exemplifies a pattern does not ensure that the pattern can be projected onto the world. "The Elephant's Child" is a fictional tale purporting to explain how the elephant got its trunk (Kipling, 1978). Once, the story goes, elephants had ordinary noses. But an insatiably curious elephant's child wanted to know what crocodiles eat for breakfast. He bent down to ask a crocodile. The crocodile, smacking his lips and thinking that young elephant would be a delicious answer to that very question, bit the youthful elephant's nose and tugged on it. As the elephant (with the help of a friendly python) pulled back, its nose elongated until it took the shape of

a trunk. That basically is the story. What is wrong with it? As a children's story, perhaps nothing. But there is surely something wrong with it as a contribution to the understanding (even a child's understanding) of biology. Since I contend that science can incorporate fictions and falsehoods, I am in no position to say, as others would, that what is wrong with the story is that it is false. So is the ideal gas law and the thought experiment about Maxwell's demon. The difference is that while the ideal gas law and Maxwell's demon afford *some* understanding of the behavior of actual gases, Kipling's story affords *no* understanding of the phylogenesis of elephant trunks. If the story, with its tacit commitment to the heritability of acquired traits, were true, it might yield some understanding of how the elephant got its trunk. But once we acknowledge that both it and its Lamarckian presuppositions are false, the story seems to have nothing to do with the phylogenesis of elephant trunks. We need to appeal to natural selection for an explanation, and that explanation will be entirely unsympathetic to any tugging-on-noses hypothesis. Kipling's story, I suggest, is neither directly nor indirectly suitably tethered to the phylogenic facts it purportedly bears on.

Experiments yield evidence, not proof. And evidence is sometimes misleading. The status of an item as a laboratory experiment or thought experiment does not hinge on its being successful in advancing understanding. I do not claim that every work of fiction succeeds. Nor do I claim that every laboratory experiment or thought experiment does. Some are muddled or confused. Some overlook real possibilities or fail to control for important variables. Some replicate what is already known or widely accepted. Some are trivial. Some thought experiments and literary works may even be the equivalent of high school science experiments, exemplifying what is already understood in an effort to show how such symbols advance understanding. Any finding must be tested by its fit with what we already have reason to believe. Galileo's thought experiment revealed an inconsistency in Aristotle's theory. It is, or is very close to, a crucial experiment. We need nothing further to show that Aristotle's theory is false. But most experiments and thought experiments, and most works of literature, work within a context of background assumptions. If the assumptions are incorrect or incomplete, an experiment or thought experiment may inherit and reinforce their inadequacies. If they are (close enough to) correct and complete, the experimental result is prima facie informative.

Still, we might resist identifying works of fiction with thought experiments. The argument from banality is the contention that the knowledge imparted by fictions amounts to little more than truisms (Carroll, 2002).[5] The guiding idea seems to be that if fiction has an epistemic function, it is to impart truths like the morals of Aesop's fables. These truths are inferred inductively or deductively from stories. Since such truths are banal, they are, for the most part at least, epistemically inert. But the patterns and features that works of fiction exemplify are far from truisms. Because exemplars display rather than merely state, they can be exceedingly fine-grained. Hitchcock's *North by Northwest* exemplifies delicate nuances in the texture of fear—a virtual continuum from trepidation to terror. There is nothing banal about learning to recognize subtle differences and project them properly onto members of the classes the exemplars typify. Nor is it likely that we will be able to capture in a pithy proposition just what such a fiction discloses.

Davies worries that I ignore a fundamental difference between literary fictions and thought experiments. Thought experiments are part of science; so to project their results onto the phenomena they pertain to is to remain within the realm of science. But to project insights drawn from fiction onto actual human relations is to move beyond literature to an extraliterary realm. This, Davies charges, is illicit, for such projection is not "intrinsic to the proper engagement of literary works *as literature*" (2010, 65). Such policing of disciplinary boundaries strikes me as ill advised. If, as I have argued, works of fiction advance understanding in much the way that thought experiments do, it makes no difference whether that advance counts as literary or extraliterary.

That a work is a rich source of insights is not a reason to doubt that it advances understanding. But it may be a reason to doubt that the work is a thought experiment. Stereotypical thought experiments tend to be austere. They symbolize along relatively few dimensions. Although they require interpretation, the interpretation of any given thought experiment is supposed to be univocal, at least until the relevant background assumptions change. But univocality is not a virtue in literary or dramatic fictions. That *Henry V* can be interpreted as pro-war and antiwar is not a defect in the play. Literary works are semantically dense and replete (Goodman, 1968). A work is semantically dense in that the characters, events, perspectives, and their alternatives are selected from a dense field of alternatives. There is no

limit to the factors, modes of description, or level of generality it could have deployed. Nor are there general rules, like the rules about statistical significance, that readily carry over from one novel to the next. In principle, every semantic difference can make a difference to what the work conveys. And a feature that makes a difference to the interpretation of one work may be irrelevant to the interpretation of another. A work is replete to the extent that it symbolizes along multiple dimensions. Syntax, vocabulary, perspective, tone, and timbre may all function symbolically. Even punctuation can matter. Any number of aspects of a fictional work can in principle contribute to its epistemic functioning. Moreover, works of fiction are apt to bear multiple correct interpretations. In several respects, then, they differ from austere, univocal thought experiments. Perhaps this is a reason to deny that works of fiction are thought experiments; perhaps it is a reason to say that only under an interpretation is a work of fiction a thought experiment; perhaps it is a reason to think that some thought experiments are more austere than others.

I favor the last alternative. Although stereotypical thought experiments are austere, there is a continuum of cases from Maxwell's demon and trolley problems through the myth of the cave and *Emile* to 'didactic fictions' like *Animal Farm* and *Uncle Tom's Cabin*, to *Middlemarch* and *Oedipus Rex*. I doubt that there is a sharp boundary between thought experiments, strictly so called, and works of fiction. But demarcating the boundary is not so important. Whether or not we call works of fiction thought experiments, I have urged that fictions, thought experiments, and laboratory experiments function in much the same way. By distancing themselves from the facts, by resorting to artifices, by bracketing a variety of things known to be true, all three exemplify features they share with the facts. Since these features may be difficult or impossible to discern in our everyday encounters with things, fictions, thought experiments, and laboratory experiments advance our understanding of the world and of ourselves.

In chapter 2 I said that despite or even because of their inaccuracy, felicitous falsehoods embody and advance understanding. Pretty clearly, thought experiments are felicitous falsehoods if any symbols are. In this chapter I have argued that fictions are, or are closely akin to, thought experiments. If so, they too can be felicitous falsehoods. I do not contend that every fiction or every thought experiment is felicitous. Some are banal. They exemplify features that are obvious and uncontroversial anyway. Some are muddled.

They exemplify nothing more than confusions that went into their design. They defy consistent interpretation and thus disclose nothing beyond their own incoherence. Some, such as racist works, mislead. They exemplify features that readers wrongly take to obtain in fact. Such falsehoods are infelicitous.

Others, however, may be felicitous in some contexts or for some audiences or for some purposes but not for others. Einstein's elevator exemplifies the equivalence of inertial and gravitational frames of reference. It operates in a context where a good deal of physical theory is presupposed. It is thus felicitous for those at home in such a context. It can advance their understanding of inertia and gravity. But it is unlikely to do the same for novices or those who consider themselves novices. They have no reason to think that *their* inability to distinguish between being at rest in a gravitational field and moving with uniform acceleration in the absence of a gravitational field is due to anything other than their ignorance. *Uncle Tom's Cabin* now strikes readers as exemplifying objectionable stereotypes and obvious moral truths. If it is not misleading, it is epistemically inert. When it was first published, however, its exemplification of the plights of the slaves was, for its intended audience—white Northern readers—revelatory. Now it is infelicitous; then it was felicitous. Even "The Elephant's Child" may deserve a closer look. Interpreted as exemplifying something about phylogeny, it is infelicitous. Interpreted as a work about the importance of having strong friends (remember the python) or about the perils of nosiness, it may be, for young children, felicitous. The felicity or infelicity of a representation turns on its function. This is so whether the representation is a free-standing one, like a work of literature, or is embedded in a scientific, philosophical, or other systematic account.[6]

12 Chekhov's Gun

If there is a gun on the wall in the first act, it must be fired by the third act.
—Anton Chekhov

Science is rife with models. Some, such as the hydraulic model of San Francisco Bay, are material. Others, such as the virial equation, are abstract. Like epistemically effective literary fictions and thought experiments, epistemically effective models are felicitous falsehoods. How, given their manifest inaccuracies, can they contribute to the understanding of their targets? A model's mode of reference is complex. Building on a paper of R. I. G. Hughes (1996), I argue that a model *represents* its target *as* having certain features. It does so by exemplifying those features and imputing them to the target. Effective models afford an understanding of their targets because their simplifications, idealizations, elaborations, and distortions make salient important features of the targets.

Representation

Science, we are told, is (or at least aspires to be) a mirror of nature, while art imitates life. If so, both disciplines produce, or hope to produce, representations that reflect the way some part of the mind-independent world is. Scientific representations are supposed to be complete, accurate, precise, and distortion-free. Although artistic representations are granted more leeway, they too are supposed to resemble their subjects. Underlying these clichés is the widespread conviction that representations are intensional surrogates for, or replicas of, their objects. If so, a representation should resemble its referent as closely as possible.

This stereotype, I have urged, is false and misleading. It engenders unnecessary problems in both the philosophy of science and the philosophy of art. It makes a mystery of the effectiveness of sketches, caricatures, scientific models, and representations with fictional subjects. Indeed, the stereotype strongly suggests that there is something intellectually suspect about such representations. Caricatures exaggerate and distort. Sketches simplify. Models may do all three. Many pictures and models flagrantly fail to match their referents. Representations with fictional subjects have no hope of matching, since they have no referents to match. The same subject, real or fictive, can be represented by multiple, seemingly incongruous representations. These would be embarrassing admissions if representations were supposed to accurately reflect the facts.

Mimetic accounts of representation fail to do justice to our representational practices. Many seemingly powerful and effective representations turn out on a mimetic account to be at best flawed, at worst unintelligible. Nor is it clear why we should want to replicate reality. As Rebecca West said, "A copy of the universe is not what is required of art; one of the damn things is ample" (1928, 131).[1] To replicate reality would simply be to reproduce the blooming, buzzing confusion that confronts us. What is the value in that? Our goal should be to make sense of things—to structure, synthesize, organize, and orient ourselves toward things in ways that serve our ends.

The problem lies in the metaphor of the mirror and the ideal of replication. Neither art nor science is, can be, or ought to be, a mirror of nature. Epistemically effective representations in both disciplines embody and convey an understanding of their subjects. Since understanding is not mirroring, failures of mirroring need not be failures of understanding. Once we recognize the ways science affords understanding, we see that the features that look like flaws under the mirroring account are actually virtues. A first step is to devise an account of scientific representations that shows how they figure in or contribute to understanding. It will turn out that an adequate account of scientific representation also affords insight into representation in the arts.

The term 'representation' is irritatingly imprecise. Pictures represent their subjects; graphs represent the data; politicians represent their constituents; representative samples represent whatever they are samples of. Thus far I have relied on context to fix the reference of different tokens of the term.

In this chapter further regimentation is called for. We can begin to regiment by restricting attention to cases where representation is a matter of denotation. Pictures, equations, graphs, charts, and maps represent their subjects by denoting them. They are representations *of* the things they denote.[2] It is in this sense that scientific models represent their target systems: they denote them. But, as Bertrand Russell (1968) insists, not all denoting symbols have denotata. A picture that portrays a griffin, a map that marks the route to Mordor, a chart that records the heights of Hobbits, and a graph that plots caloric emissions are all representations, though they do not represent anything. To be a representation, a symbol need not itself denote, but it must be the sort of symbol that denotes. Griffin pictures are representations, then, because they are animal pictures, and some animal pictures denote animals. Middle Earth maps are representations because they are maps, and some maps denote real locations. Hobbit height charts are representations because they are charts, and some charts denote magnitudes of actual entities. Caloric emission graphs are representations because they are graphs, and some graphs denote real processes. So whether a symbol is a representation is a question of what kind of symbol it is.

Following Goodman, let us distinguish between representations *of p* and *p*-representations. If *s* is a representation *of p*, then *p* exists and *s* represents *p*. But *s* may be a *p*-representation even if there is no such thing as *p* (Goodman, 1968). Thus, there are griffin-pictures even though there are no griffins to depict. There are caloric-descriptions even though there is no caloric to describe. There are also mixed cases. The class of dog-representations includes both factual and fictional representations. Factual dog-representations are representations of actual dogs; fictional dog-representations lack denotata. Literary fictions and thought experiments are *p*-representations that do not denote *p*.

Denoting symbols with null denotation may seem problematic. Occasionally it is said that in the absence of griffins, there is no basis for classifying some pictures as griffin pictures and refusing to so classify others. Such an objection supposes that the only basis for classifying representational contents is by appeal to an antecedent classification of their referents. This is just false. We readily classify pictures as landscapes without any acquaintance with the real estate—if any—that they represent. I suggest that each class of *p*-representations constitutes a small genre, a genre composed of all and only representations with a common ostensible subject

matter. There is a genre of griffin-representations and a genre of ideal-gas-representations. We learn to classify representations as belonging to such genres as we study those representations and the fields of inquiry that devise and deploy them. This is no more mysterious than learning to recognize pictures as landscapes without comparing them to the terrain they ostensibly depict.

Some representations denote their ostensible objects. Others do not. Among those that do not, some—such as caloric-representations—simply fail to denote. They purport to denote something, but there is no such thing. They are therefore defective. Others, such as ideal-gas-representations, are fictive. Since they do not purport to denote any real object, their failure to denote is no defect (see Frigg, 2010). We know perfectly well that there is no such animal as a griffin, no such person as Othello, no such gas as the ideal gas. Nonetheless, we can provide detailed representations *as if* of each of them, argue about their characteristics, be right or wrong about what we say respecting them, and, I have argued, advance understanding by means of them.

Representation-As

Symbol *x* is, or is not, a representation *of y* depending on what *x* denotes. And *x* is, or is not, a *z*-representation depending on its genre. This enables us to form a more complex mode of representation in which *x* represents *y as z*. In such a representation, symbol *x* is a *z*-representation that as such denotes *y*. Caricature is a familiar case of representation-as. Winston Churchill is represented as a bulldog; Teddy Roosevelt is represented as a bull moose.

Hughes (1996) argues that representation-as is central to the way that models function in science. A diagram represents a spring as a harmonic oscillator; the Hardy—Weinberg model represents a species as mating randomly; an ideal gas model represents molecules as displaying no mutual attraction. A model, Hughes says, is a theoretical construct that involves three components: denotation, demonstration, and interpretation. Denotation is the relation of elements of the model to elements of its target. Demonstration is the operation of the model according to its own 'internal dynamic', which enables us to use the model to derive conclusions. Interpretation is a matter of reading the conclusions derived from the model back into the target. For example, a model—such as an equation—denotes

a pendulum, an actual swinging bob. By plugging in values and performing a calculation, we demonstrate the relation between two mathematical variables. We then interpret the values of the variables as length and period, and impute them to the material pendulum.

Denotation is straightforward. It is the semantic relation of a name to the thing it names, of a picture to the thing it depicts, of a predicate to the things it is true of. But more needs to be said about demonstration and interpretation. A demonstration can be mathematical or material. "If we want to apply the wave theory of light to two-slit interference, for example, we can do one of two things. We can either model the phenomenon mathematically, or we can get out a ripple tank and model light as a wave motion in the literal sense. In either case we will find that the distance between interference fringes varies with the separation of the sources, and also with the frequency of the waves" (Hughes, 1996, S322). Mathematics does not typically strike us as dynamic. Equations sit quietly on the page. In saying that an equation has an internal dynamic, Hughes means that we reason with it, use it as an inference ticket, draw consequences from it. What is dynamic is the drawing out of consequences via inference. A major insight of his analysis is that models are not static representations. They are tools. They enable us to figure things out.

Hughes is right as far as he goes, but he stops too soon. The problem is that there are always more consequences to be drawn from any demonstration. We need to know what direction to take in our reasoning and where to stop. That is, we need to know which of the inferences that we might legitimately make should be read back onto the target. Again exemplification comes to the rescue. The demonstration, whether in the ripple tank or on paper, has irrelevant features—ones that ought to be ignored. If we continued the demonstration endlessly, the number of irrelevant features would proliferate. But being irrelevant, they are not exemplified. They should be set aside. Interpretation involves ascertaining which features the demonstration exemplifies, hence which ones we should project onto the target. This involves understanding how, in what respects, and with what degree of precision the model represents. Models are symbols. To interpret them correctly requires appreciating which of their elements represent and what those elements refer to.

Representation-of—that is, denotation—can be achieved by fiat. We simply stipulate: let x represent y and x thereby becomes a representation of y.

This is what we do in baptizing an individual or a kind. It is also what we do in ad hoc illustrations, as, for example, when I say (with appropriate accompanying gestures), 'If that chair is Widener Library, and that desk is University Hall, then that window is Emerson Hall' in helping someone to visualize the layout of Harvard Yard. Like a chair or a desk, a *p*-representation can by arbitrary stipulation denote any object. We might, for example, point to a tree-picture and stipulate that it denotes the philosophy department. But our arbitrary stipulation does not bring it about that the tree-representation represents the philosophy department as a tree.

Does representation-as then require similarity? In that case, what blocks seemingly groundless and arbitrary cases of representation-as is the need for resemblance between the representation and the referent. Only a department that resembles a tree could then be represented as a tree. But similarity does not establish a referential relationship (Goodman, 1968; Suárez, 2003). Representation is an asymmetrical relation; similarity is symmetrical. Representation is irreflexive; similarity is reflexive. Perhaps this shows only that similarity is not sufficient for representation-as. Something else determines direction. Once the direction is established, it is the similarity between symbol and referent that brings it about that the referent is represented as whatever it is represented as.

The problem is this: Via stipulation, we have seen, pretty much anything can represent pretty much anything else. So nothing beyond stipulation is required to bring it about that one thing represents another. But similarity is ubiquitous. This is the insight of nominalism. For any x and any y, x is somehow similar to y. Thus if all that is required for representation-as is denotation plus similarity, then for any x that represents y, x represents y as x. Every case of representation turns out to be a case of representation-as. In one way or another, the philosophy department is similar to a tree-picture, but it is still hard to see how that fact, combined with the stipulation that a tree-picture represents the department, could make it the case that the department is represented as a tree-picture, much less as a tree. Suppose we add that the similarity must obtain between the content of the *p*-representation and the denotation. Then for any *x-representation* and any *y*, if the *x-representation* denotes *y*, the *x-representation* represents *y* as *x*. In that case, a tree that represented the philosophy department would not represent it as a tree. But a tree-picture that represented the philosophy department would represent it as a tree.

The trouble is that contentful representations, as well as chairs and desks, can be used in ad hoc illustrations like the one I gave earlier. If the portrait of the dean on the wall represents Widener Library, and the graph on the blackboard represents University Hall, then the map in the corner represents Emerson Hall. This does not make the dean's portrait represent Widener Library as the dean. Evidently, it takes more than being represented by a tree-picture to be represented as a tree. Some philosophy departments can be represented as trees. But to bring about such representation-as is not to arbitrarily stipulate that a tree picture shall denote the department, even if we add a vague intimation that somehow or other the department is similar to a tree. The question is, what is effected by such a representation?

To explicate representation-as, Hughes (1996) discusses Sir Joshua Reynolds's painting, *Mrs. Siddons as the Tragic Muse*. The painting denotes its subject and represents her as the tragic muse. How does it do so? It establishes Mrs. Siddons as its denotation. It might represent Mrs. Siddons, a person familiar to its original audience, in a style that that audience knows how to interpret. Then, without further cues, they could recognize that the picture is a picture of her. But the painted figure need not bear any particular resemblance to Mrs. Siddons. *We* readily take her as the subject even though we have no basis for comparison. (Indeed, we even take Picasso's word about the identities of the referents of his cubist portraits, even though the figures in them do not look like anyone on earth.) Captioning the picture as a portrait of Mrs. Siddons suffices to fix the reference. So a painting can be connected to its denotation by stipulation. The painting is a tragic-muse-picture. It is not a picture *of* the tragic muse, there being no such thing as the tragic muse. But it belongs to the same restricted genre as other tragic-muse-representations. To recognize it as a tragic-muse-picture is to recognize it as an instance of that genre. Similarly in scientific cases. A spring is represented as a harmonic oscillator just in case a harmonic-oscillator-representation as such denotes the spring. The harmonic-oscillator-representation involves idealization. So it is not strictly a representation *of* a harmonic oscillator, any more than the Reynolds is a picture *of* the tragic muse.

In both cases, a representation that does not denote its ostensible subject is used to denote another subject. Since denotation can be effected by stipulation, there is no difficulty in seeing how this can be done. The difficulty comes in seeing why it is worth doing. What is gained by representing Mrs. Siddons as the tragic muse, or a spring as a harmonic oscillator, or in

general by representing an existing object as something that does not in fact exist? The quick answer is that the representation affords epistemic access to features of the object that are otherwise difficult or impossible to discern. They exemplify those features.

Nor does this occur only when a subject is represented as something fictional. Legend has it that when people objected that his portrait of Gertrude Stein did not look like her, Picasso replied, "No matter. It will" (Goodman, 1968, 33). Although this remark might be dismissed as yet another expression of Picasso's arrogance, Robert Schwartz (1985) argues that we should take it seriously. Through his portrait, Schwartz maintains, Picasso gave Stein a new look. In so doing, he not only reconfigured Stein, he changed the world. Schwartz's position can be elaborated to show how scientific models do the same. The first step is to explain how a portrait can literally give someone a new look and what follows from it.

Schwartz maintains that through his portrait, Picasso brought people to see Stein differently. It did so, I suggest, by exemplifying features of Stein that were normally overlooked. Picasso did not portray her as a dumpy, middle aged, Midwestern woman, but as magisterial, imposing, someone to be reckoned with. Through his portrait, he made these features of Stein— *being magisterial, being imposing, being someone (or something) to be reckoned with*—salient. We see Stein differently as a result. If we had previously overlooked the fact that she had those features, then not only our perception of her but also our conception of her is reconfigured once we come to recognize them.

Moreover, Schwartz insists, we gain more than just a new way of looking at and thinking of Stein. We also gain new criteria for what it takes to *look* like Stein and what it takes to *be* like Stein. These criteria afford bases for likening. People, or indeed things, that we would never previously have counted as looking like Stein are now recognized as doing so, for they share the features that her portrait exemplifies. They therefore have something significant in common not just with Stein but also with each other. They belong to an extension that we are now in a position to call the Stein-lookalikes. If, via the picture, we see in Stein herself features we had overlooked, then we can see others as like her in ways that go beyond her appearance—as *being* imposing, magisterial, someone to be reckoned with, for example. We then can recognize others as Stein-alikes on the basis of something other than their appearance.

At this point, metaphysicians might object that the extension consisting of all and only what we are now in a position to call the Stein-lookalikes and the extension of what we are now in a position to call the Stein-alikes existed all along. Picasso did not bring the extensions into existence. He did not create them. This is true. Extensions, marked or unmarked, are equally real; none is brought into existence by baptizing, exemplifying, or construing. But most extensions strike us as motley collections of apparently disparate objects. Their members have nothing notable in common. The extension consisting of the Gobi Desert, an apple pie, and a baseball in Detroit have something in common with one another: they are all members of that extension. This is true but trivial. There's little reason to care about it. From our antecedent perspective, the extensions we now call the Stein-lookalikes and the Stein-alikes seemed arbitrary collections of unrelated things. What Picasso brings out is that these particular collections are by no means arbitrary—their members have something significant in common. They share the constellation of features that the portrait exemplifies.

Still, one might balk, at least about the Stein-alikes. Arguably we can concede that the portrait instantiates, and therefore is capable of exemplifying, features like *looking magisterial*. But a portrait itself can't *be* magisterial. The claim that it exemplifies *being magisterial* is blocked from the outset, since exemplification requires instantiation. To handle this objection, we need to appeal to representation-as. The portrait represents Gertrude Stein *as* magisterial insofar as it denotes her, belongs to the genre of magisterial-figure-portrayals, and belongs to that genre because of the way it portrays her—namely as having certain discernible features that its subject purportedly shares with other members of that genre. ('Purportedly' because it is possible to misrepresent someone as magisterial.)

Modeling

Models, like fictions, can simplify, omitting confounding factors that would impede epistemic access to the properties of interest. They can abstract, paring away unnecessary and potentially confusing details. They can distort or exaggerate, highlighting significant aspects of the items they focus on. They can augment, introducing additional elements that focus attention on properties of interest. They can insulate, screening off effects that would otherwise dominate. They do these things in order to exemplify.

A familiar worry, first encountered in the discussion of Galileo's thought experiment, recurs here. The same solution holds. Many scientific models, such as equations and diagrams, are incapable of instantiating the properties they apparently impute to their targets. If they cannot instantiate a range of properties, they cannot exemplify them. Suppose, via an equation, we model a pendulum as a simple harmonic oscillator. Since exemplification requires instantiation, if the model is to represent the pendulum as having a certain mass, the model must have that mass. But, not being a material object, the model has no mass. It cannot exemplify the mass of the pendulum. Indeed, the model does not exemplify mass. Rather, it exemplifies an abstract mathematical property, the magnitude of the pendulum's mass. Where models are abstract, they exemplify abstract patterns, properties, and/or relations that may be instantiated in physical target systems. It does no harm to say that they exemplify physical magnitudes. But this is to speak loosely. Strictly speaking, they exemplify mathematical (or other abstract) properties that can be instantiated physically.

Even concrete models display the sort of divergence from their targets that we saw above. The San Francisco Bay-Delta model—a massive, hydraulic scale-model of San Francisco Bay and its environs (see Weisberg, 2013)—has no tides. So it cannot exemplify the effect of tides on the water in the bay. But it mimics the tides by gradually pumping salt water into and out of the model's basin. It thus exemplifies an effect that is instantiated by the ebb and flow of tidal waters in the bay. Speaking loosely, we might say that it exemplifies the effects of the tides. Speaking strictly, we should say that it exemplifies an effect that is instantiated both by flow of water in the model itself and by the tides.

Both literary fictions and scientific models exemplify properties and afford epistemic access to them. By omitting or downplaying the significance of confounding factors (the Napoleonic wars in *Pride and Prejudice*, intermolecular attraction in the ideal gas, friction in the model pendulum), they constitute a cognitive environment where certain aspects of their subjects stand out. They thereby facilitate recognition of those aspects and appreciation of their significance. They thus give us reason to take those aspects seriously elsewhere.

Of course, this does not justify a straightforward extrapolation to reality. From the fact that Elizabeth Bennet was wrong to distrust Mr. Darcy, we cannot reasonably infer that young women in general are wrong to

distrust their suitors, much less that any particular young woman is wrong to distrust any particular suitor. But the fiction exemplifies the grounds for distrust and the reasons those grounds may be misleading. Once we have seen them clearly in a fictional setting, we may be in a better position to recognize them in everyday situations. Nor from the fact that ideal gas molecules exhibit no mutual attraction can we reasonably infer that neither do helium molecules. But the behavior ideal gas molecules exemplify in the model may enable us to recognize such behavior amid the confounding factors that ordinarily obscure what is going on in actual gases. In using a model, we implicate that the features of the phenomena that it brackets are for current purposes negligible. Here, we suggest, we can safely ignore them.

Epistemic Access

Return to Reynolds's representation of Mrs. Siddons as the tragic muse. The tragic muse is a figure from Greek mythology who is supposed to inspire works of tragedy—works that present a sequence of events leading inexorably from a position of eminence to irrecoverable, unmitigated loss, thereby inspiring pity and terror (Aristotle, 1973). A tragic-muse-representation portrays a figure as capable of inspiring such works, one who exemplifies such features as nobility, seriousness, inevitability, and perhaps a somber dramaticality, along with a capacity to evoke or arouse pity and terror. To represent a person as the tragic muse is to represent her in such a way as to reveal or disclose such characteristics in her or to impute such characteristics to her.

The ideal gas law is an equation ostensibly relating temperature, pressure, and volume in a gas. To satisfy that equation, a gas would have to consist of perfectly elastic spherical particles of negligible volume and exhibiting no mutual attraction. The law thus defines a model that mandates specific values for size, shape, elasticity, and attraction. With these parameters fixed, the interdependence of the values of temperature, pressure, and volume is exemplified. The law and the model it defines are fictions. There is no such gas. Indeed, so far as we can tell, there could be no such gas, since such a gas would violate several well-founded laws of physics. Nevertheless, the model advances our understanding of gas dynamics. It exemplifies a relation that is important but hard to discern in the behavior of actual gases. The model

is a representation—a denoting symbol that has an ostensible subject and portrays its ostensible subject in such a way that certain features are exemplified. It represents its target (its denotatum) as exhibiting those features. So to represent helium as an ideal gas is to impute to it features that the ideal gas model exemplifies. By setting the parameters to zero, it construes the actual size, shape, inelasticity, and mutual attraction of the molecules as negligible. Strictly, of course, in helium the values of those parameters are not zero. But if they are negligible, they can safely be ignored. In that case, the imputation allows for a representation that discloses regularities in the behavior of helium that a more faithful representation would obscure. The model then foregrounds the interdependence of temperature, pressure, and volume, making it and its consequences manifest.

Representing a philosophy department as a tree might exemplify the ways the commitments of the various members branch out from a common, solid, rooted tradition, and the ways the work of the graduate students further branch out from the work of their professors. It might intimate that some branches are flourishing while others are stunted growths. It might even suggest the presence of a certain amount of dead wood. Representing the department as a tree, then, affords resources for thinking about it, its members and students, and their relation to the discipline in ways that we otherwise would not.

I said earlier that when x represents y as z, x is a z-representation that *as such* denotes y. We are now in a position to cash out the 'as such'. It is *because* x is a z-representation that x denotes y as it does. x does not merely happen to denote y and happen to be a z-representation. Rather, in being a z-representation, x exemplifies certain properties and imputes those properties or related ones to y. 'Or related ones' is crucial. A caricature that exaggerates the size of its subject's nose need not impute an enormous nose to its subject. By exemplifying the size of the nose, it focuses attention, thereby orienting its audience to the way the subject's nose dominates his face or, through a chain of reference, the way his nosiness dominates his character. The properties exemplified in the z-representation thus serve as a bridge that connects x to y. This enables x to provide an orientation to its target that affords epistemic access to the properties in question.

There is no guarantee that the target has the features the model exemplifies, any more than there is any guarantee that a subject represented as the tragic muse has the features that a painting representing her as the

tragic muse exemplifies. This is a question of fit. A model may fit its target perfectly or loosely or not at all. Like any other case of representation-as, the target may have the features the model exemplifies. Then the function of the model is to display those features and make their significance manifest. We may see the target system in a new and fruitful way by focusing on the features that the model draws attention to. In other cases, the fit is looser. The model does not exactly fit the target. A target that does not instantiate the specific properties its model exemplifies may instantiate more generic properties that subsume the exemplified properties. If gas molecules are roughly spherical, reasonably elastic, and far enough apart, then we may gain insight into their behavior by representing them as perfectly elastic spheres with no mutual attraction. Perhaps we will subsequently have to introduce correction factors to accommodate the divergence from the model; perhaps not. It depends on what degree of precision we want or need.

Sometimes, although the target does not quite instantiate the features exemplified in the model, it is not off by much. Where their divergence is negligible, the models, although not strictly true of the phenomena they denote, are true enough of them. This may be because the models are approximately true, or because they diverge from truth in irrelevant respects, or because the range of cases for which they are not true is a range of cases we do not care about, as for example when the model 'blows up' at the limit. Where a model is true enough, we do not go wrong if we think of the phenomena as displaying the features that the model exemplifies. Whether a representation is true enough is a contextual question. A representation that is true enough for some purposes or in some respects is not true enough for or in others. This is no surprise. No one doubts that the accuracy of models is limited. In yet other cases, the model simply does not fit. Then it affords little or no understanding of its target. Not everyone can be informatively represented as the tragic muse. Nor can every object be informatively represented as a perfectly elastic sphere.

Earlier I dismissed resemblance as the vehicle of representation. I argued that exemplification is required instead. But for x to exemplify a property of y, x must share that property with y. So x and y must be alike in respect of that property. It might seem, then, that resemblance in particular respects is what is required to connect a representation with its referent (Giere, 1999). There is a grain of truth here. If exemplification is the vehicle for

representation-as, the representation and its object resemble one another in respect of the exemplified properties. But resemblance, even resemblance in a particular, relevant respect, is not enough, as the following tragic example shows.

On January 28, 1986, the space shuttle *Challenger* exploded because, owing to cold weather, its O-rings failed to effect a seal. The previous day, engineers involved in designing the shuttle had warned NASA about that very danger and faxed data to NASA to support their concern. The printouts contained complex representations conveying vast amounts of information about previous shuttle flights. They included measurements of launch temperatures for previous flights and measurements of six types of O-ring degradation after each flight. Had loss of elasticity been plotted against temperature, the danger would have been clear. The evidence that the O-rings were vulnerable in cold weather was contained in the data. But it was obscured by a mélange of other information that was also included (Tufte, 1997). So although the requisite resemblance between the highly complex, multipurpose model and its target obtained, it was overshadowed in the way that a subtle irregularity in an elaborate tapestry might be. As it was presented, the data instantiated but did not exemplify the correlation between temperature and loss of elasticity. They did not represent the O-rings as increasingly inelastic as the temperature dropped. Because the correlation between O-ring degradation and temperature was not perspicuous, the NASA decision makers did not see it. The launch took place; the shuttle exploded; the astronauts died. When the goal of a representation is to afford understanding, its merely resembling the target in relevant respects is not sufficient. The representation must make the resemblance manifest.

Drawing on Hughes's account, I have argued that models exemplify features they share with their targets and impute those features to their targets. Where they are successful, they afford epistemic access to aspects of their targets that we might otherwise miss. When an economic model represents a transaction as occurring in a situation where there is perfect information, it is not purporting or pretending that actual financial transactions occur in such contexts. Rather, it factors out what might be seen as interference. By exhibiting what would occur in such a context, it equips us to understand why and to what extent differences in information available to different parties to a transaction matter.

Models are selective. They highlight some features of their targets by marginalizing or downplaying the significance of others. The utility of any particular model is therefore limited. The Hardy—Weinberg model represents a population as consisting of infinitely many members of a species who mate randomly, where that population is not subject to mutation, migration, or genetic drift. It exemplifies the pattern in the redistribution of alleles in the absence of evolutionary pressures. Inasmuch as evolutionary pressures are always present, the model cannot, nor does it pretend to, account for allele distribution generally. It is, however, very useful for some purposes. If population geneticists want to understand how significant an evolutionary factor such as migration is, they need a base rate. They need, that is, to know how alleles would redistribute in its absence. Features that can be neglected in some contexts or for some purposes cannot always be neglected.

By exemplifying a feature, a model affords epistemic access to it, and provides reason to suspect that it is significant. It thereby equips us to see the target differently than we otherwise might. Instead of thinking of a feature as just one of innumerable features of the target, we have reason to take it seriously. Familiar gas models represent gas molecules as lacking attractive force. If the results of our calculations are confirmed when we read them back into the target, we have reason to think that intermolecular forces do not play a significant role in gas dynamics. Knowing as we do that every material object attracts every other one, we do not conclude that there is no attraction. Rather, we conclude that for the sort of understanding we currently seek, at the level of precision we are interested in, intermolecular attraction is negligible. This suggests that we think of the target in terms of the features exemplified in the model. Roughly, it invites us to think of actual gases as ideal gases with distortions, springs as harmonic oscillators with confounding features, and so forth.

Likening

Models liken. If effective, they exemplify features that the members of their target class share. They thereby have the capacity to bring to light significant affinities among members of the class. Let's look at a couple of cases. Strictly, the equation $pV = nRT$ would hold only of spherical molecules. In effect, the model defines a fictional gas whose molecules are spherical.

However, the model is supposed to illuminate the behavior of real gases, whose molecules are lumpy. Moreover, molecules of different gases have different configurations of lumps. On the face of it, the shapes of oxygen molecules and argon molecules do not strike us as particularly similar. But the model intimates that the actual differences in shape can be ignored. Where the model is appropriate, they are negligible. And where the differences are negligible, gas molecules are relevantly alike. By representing actual gas molecules as spheres with distortions, and making the case that the distortions have negligible effects, the model likens actual gas molecules to one another.

A more interesting case is biology's evolutionary game-theoretic model. It applies to, among other things, predator—prey interactions, mating behavior, fertility and fecundity rates, and altruism. It is a dynamic model, displaying change over time. Even if one is completely convinced that natural selection underwrites all these phenomena, my list seems to consist of wildly disparate items. It is far from obvious that, at any level of abstraction, predator—prey interactions and altruism are the same sort of thing. Off the cuff, they seem virtually antithetical. Although we readily grant that in one way or another, natural selection accounts for all of the items on my list, there seems no a priori reason to suspect that there is a single way that natural selection accounts for all of them. It would not be surprising if each had a separate evolutionary explanation. What the game-theoretic model does is show that at a certain level of abstraction, these phenomena are all alike. They are instances of the same sort of process. The model thus unifies the domain and effects a likening among seemingly disparate phenomena. We can now appreciate why, from an evolutionary perspective, altruistic behavior and predatory behavior are alike.

If we think of the target in terms of the features the model exemplifies, we may recognize and appreciate the significance of affinities we would otherwise overlook. Newton's law of gravity enables us to recognize that the moon's staying in orbit, the ebb and flow of the tides, and the apple's falling from the tree are fundamentally the same sort of thing. The glaring differences between them are to be ignored. In important physical respects, the motions are alike—they satisfy the same equation, and thereby exemplify the same sort of behavior. Perhaps surprisingly, a good way to understand why satellites remain in orbit is to represent them as falling bodies.

Models do not, of course, only omit. If they did, there would be no reason to consider them inaccurate. They could be construed as schematic or partial representations of the phenomena they bear on. But models also distend, distort, exaggerate, and even introduce elements that answer to nothing in the target. These distentions, distortions, exaggerations, and interpolations scaffold representation-as. A distended, distorted, or exaggerated model can exemplify features that obtain in the target but are difficult to discern there. For example, according to Kepler's laws, the orbit of the Earth is elliptical, with one focus at the sun. The major axis and the minor axis are almost equally long. The shape of the ellipse is very close to that of a circle. But to highlight the fact that the orbit is not a circle, models of Kepler's laws typically represent the major axis as considerably longer than the minor axis. This would be objectionable if we took such models to exemplify the actual shape of the orbit. But since they exemplify only the orbit's being elliptical rather than circular and not the precise shape of the ellipse, there is no difficulty. Maxwell's representation of electromagnetic radiation as involving rotating vortices separated by 'idle wheels' denotes that radiation and, through the introduction of fictional elements, makes manifest how the relational structure of the production and transmission of electric and magnetic forces in an electromagnetic medium is like the relational structure of the production and transmission of mechanical forces in a mechanical medium (Nersessian, 2008, 29–48). Through this model, Maxwell effected a likening of electromagnetic and general dynamical processes.

To properly understand a model or other representation-as we must interpret it correctly. Some features of the representation are inert. It makes no difference, for example, what typeface the statement of Kepler's laws is in or whether the hydraulic model of San Francisco Bay is properly oriented toward the magnetic North Pole. The laws represent the orbit of the Earth, regardless of font, and the model would simulate the effects of the tides equally well if it were rotated 30 degrees. Some features function as scaffolding. They do not themselves exemplify features of the target. But they enable the model to exemplify the features it imputes to the target. The idle wheels in Maxwell's model and the canine features in Churchill's caricature scaffold in that they enable the representations to highlight the features they exemplify. They are thus not irrelevant to *how* the symbol represents, although they do not figure in *what* it represents. They are not to be read

back into the target. Finally, there are the features that the model exemplifies: these are the ones it imputes to its target.

Thus far, I have assumed that a model represents something that is, in Williams's phrase, there anyway. The target is as it is regardless of how or whether it is represented. If so, the role of the model is to bring to light unanticipated or unappreciated aspects of the phenomena it models. Sometimes this is so. The planets are there anyway and are unaffected by their susceptibility to being modeled as point masses. If all models fit this mold, it would not be unreasonable to balk at the claim that anything properly called 'worldmaking' is going on (see Goodman, 1978). But there are cases where the relation between model and target is more dynamic and interactive than this picture suggests. The demarcation of the target is the product of a negotiation between an antecedent partition of the domain and the resources a model can bring to bear.

Hasok Chang's (2004) account of the tandem development of the thermometer and temperature brings this out. One might have thought that the inventors of the thermometer began with an antecedently recognized magnitude—temperature. Their challenge was to come up with a way to measure it. That is: given temperature, invent the thermometer. But, Chang argues, this is an oversimplification. At the outset, there were phenomenologically distinct experiences as of hotter or colder. Being grounded entirely in sensory experience, judgments about felt differences were subject to unwelcome fluctuations. Two perceivers of the same phenomenon would give different verdicts. The same phenomenon, when judged against different backgrounds, would be felt to be hotter or colder. "Put one hand in a bucket of hot water and the other one in cold water; after a while take them out and put them in a bucket of lukewarm water; one hand feels that the water is cool, the other feels it is warm" (Chang, 2004, 43). Hint that it is a hot coal, and an ice cube dropped down someone's back will be felt as a burn. Scientists couldn't rest satisfied with these verdicts. Too much seemed to depend on the conditions of perception and sensibilities of the perceivers. So they sought to create a device that would measure the 'real' differences, but bracket those that depended on the varying sensibilities and conditions of perceivers. They couldn't quite succeed. The best devices they could come up with did not quite reflect the differences in felt hotness or coldness even when obvious biases and idiosyncrasies were controlled for. Initially, of course, this might just be evidence that their devices were

inadequate. But eventually, they came to trust the devices. Rather than concluding that thermometers were not really measuring heat, scientists concluded that what perceivers take to be experiences of heat are actually experiences of heat with confounding factors, such as barometric pressure and humidity. They redefined the target of their measurement to legitimize the measuring device.

The thermometer is not just a measuring device, it is a model. It represents changes in the height of a column of mercury as changes in temperature. That is, the thermometer exemplifies magnitudes or changes magnitude and imputes them to the target. So Chang's discussion shows that models can figure in the constitution of the phenomena they model. The process of inventing the thermometer was in part a process of constituting a specific magnitude to be measured.

Multiple Models

This account enables us to see why different models of the same target serve different purposes. For some purposes, representing light as a wave is valuable; for others, it is preferable to represent light as a particle. For some purposes, representing a customer as an ideally rational agent with full information is appropriate; for others, representing her as an individual with idiosyncratic and mutually inconsistent desires and preferences and with limited information is preferable. Each model exemplifies different features and affords epistemic access to different aspects of the target.

Another question looms. Why, when we refine a model, does its more idealized predecessor endure? The ideal gas $pV = nRT$ was exported from thermodynamics to statistical mechanics. The model has undergone numerous refinements which successively de-idealize its assumptions. The van der Waals equation, for example, accommodates intermolecular forces which the earlier model scrupulously ignores. Further refinements in the equations of state yield increasingly realistic representations. Eventually we get the virial equation,

$$PV/NkT = 1 + B/V + C/V^2 + D/V^3 + E/V^4 + \cdots,$$

"which can be rendered arbitrarily precise by extending the equation indefinitely, with each added term being derivable from increasingly detailed

assumptions about the intermolecular forces" (Doyle et al., 2015, 5). Armed with the virial equation, why don't we jettison the predecessors or consign them to the dustbin of the history of science? Boyle, Charles, van der Waal, et al. seem to have been supplanted.

One possible answer is intellectual inertia. We don't clean our conceptual closets all that often, so remnants of outgrown commitments remain. This, perhaps, is why we remember the old models. But it doesn't explain why we continue to use them, particularly if we consider them outdated. And it certainly does nothing to justify their continued employment.

Another possibility is that the older models are easier to use. They require fewer and simpler calculations. But the virial equation is not particularly difficult. The math is straightforward. So there is little practical advantage in using the earlier models. And even if there were a practical advantage, it is far from obvious that availing ourselves of that advantage would be epistemically legitimate.

I suggest that it is a mistake to think that the virial equation is to be preferred simply because it can be made arbitrarily precise, and can accommodate increasingly detailed and accurate assumptions. I am not denying that the equation is an impressive accomplishment. Nor am I denying that it is the right model to use in certain circumstances. What I am denying is that what we want of a model is as precise and detailed a representation as we can get. The aim of modeling is to come up with a representation that affords an understanding of the phenomena, not one that replicates the phenomena. That is why we can be entirely justified in deploying an idealized model when a more realistic one is available. The highly idealized $pV = nRT$ exemplifies a relation between pressure, temperature, and volume, and in imputing it to its target implicates that further factors, such as intermolecular forces, are negligible. By omitting the negligible, we put ourselves in a position to appreciate the relations that matter. In contexts where those forces are negligible, we do well to set them aside. If, for example, we want to understand how a pressure cooker works, or how a football might deflate in cold weather, we need no more than $pV = nRT$. In other contexts, intermolecular forces are nonnegligible. Then we should use a different model—one that accommodates them *to the extent that* they are nonnegligible. Maybe this means using the van der Waals equation. Maybe it means going all the way to the virial equation.

If I want to make the world safe for the ideal gas model, however, there is a worry that has to be faced. I said earlier that not every feature of a model exemplifies. So a seemingly efficient way of accommodating the cases where intermolecular forces are negligible and the cases where they are only slightly significant would be to use the virial equation and interpret it so that only some of the terms exemplify. That would amount to neglecting the terms whose contribution is negligible. Rather than taking different models to exemplify different features of the target, we would assign different interpretations (in Hughes's sense of 'interpretation') to a single model. This could be done. It might even be a good idea. Still, there is the pesky ideal gas law rattling around in our brains. Why?

According to Chekhov, if a gun is on the wall in the first act of a play, it must be fired by the third act. His edict has an analogue in science. Suppose we are in a context where $pV = nRT$ would be appropriate. Instead, we might use the virial equation,

$$PV/NkT = 1 + B/V + C/V^2 + D/V^3 + E/V^4 + \cdots,$$

and ignore the bits we don't need. But if we don't need them, why are we given them? What is the point of all the additional bells and whistles? (What is that gun doing on the wall if it's never going to be fired?) To include multiple terms is to implicate that they have a bearing on the issue. This follows from Grice's second maxim of quantity: "Do not make your contribution more informative than is required" (1989, 26). In including unnecessary details, we court misunderstanding; we invite the audience to conclude that the details are in fact necessary. Somehow, even in this simple case, audience members may think, we are supposed to recognize the significance of E/V^4.

But all models have features that do not represent. Kepler's laws are in some font or other; the diagram of the harmonic oscillator is some height or other. The hydraulic model of San Francisco Bay is some distance or other from the magnetic North Pole. To treat these features as significant is to misinterpret the model. Similarly, even if Chekhov could have left the gun off the wall, the wall would be some color or other. He could not insist that if the color of the wall performs no function, the wall should have no color, for a transparent wall would itself be deemed significant. The

standard accommodation to the inevitable is to make the ineliminable fea-
tures as unobtrusive as possible. So the wall would be a neutral shade, the
font would be a standard font, the diagram would be designed simply to fit
neatly on the page, and the model of San Francisco Bay would have no fea-
tures that even hinted that its distance from the magnetic North Pole mat-
tered. Arguably, then, we could supplant the other models with the virial
equation and simply interpret it so that the irrelevant features are, like the
color of the wall or the font of the equation, sidelined. This could be done.
But it would make the model ambiguous. Sometimes the equation would
exemplify the value of E/V^4, sometimes it would not.

Science places a premium on intersubjective agreement. Because scien-
tists build on one another's findings, it should be clear exactly what those
findings are and what justifies them. So other things equal, science seeks
univocal symbols. Ambiguity is thus, as far as possible, to be avoided. This,
I suggest, is why, rather than replacing the more idealized equations of state
with the virial equation, statistical mechanics retains them. One way or
another, context has to be accommodated. The choice is between admitting
multiple models, each suited to a separate range of contexts, and admitting
a single model whose interpretation varies with context.

Because models depend on exemplification, they are selective. A model
makes some features of its target manifest by overshadowing or ignoring
others. So different models of the same target may make different features
manifest. Where models are thought of as mirrors, this seems problematic.
It is hard to see how the nucleus of an atom could be mirrored without
distortion as a liquid drop and as a shell structure.[3] Since a single material
object cannot be both fluid and rigid, there might seem to be something
wrong with our understanding of the domain if both models are admissi-
ble. But if what one model highlights is that in some significant respects the
nucleus behaves like a liquid drop, and another model highlights that in
some other significant respects it behaves as though it has a shell structure,
there is in principle no problem. There is no reason why the same thing
should not share some significant properties with liquid drops and other
significant properties with rigid shells. It may be surprising that the same
thing could have both sets of features, but there is no logical or concep-
tual difficulty. The models afford different perspectives on the same reality.
And it is no surprise that different perspectives reveal different aspects of
that reality. There is no perfect model (Teller, 2001) for the same reason that

there is no perfect perspective. Every perspective, in revealing some things, inevitably obscures others (van Fraassen, 2008).

Conclusion

In chapter 2, I said that models like the ideal gas and the harmonic oscillator are felicitous falsehoods. They afford an understanding of their targets because their inaccuracies do not interfere with, and may even enhance, their epistemic function. We see now how this can be so. Effective models exemplify features they share with their targets and represent their targets as having those features. This makes them epistemically felicitous. But not all models are, even in my extended use of the term, falsehoods. A model house on a building site is a real house. You could live in it. Eventually someone probably will. A model organism is a real, live organism—a mouse or frog or fruit fly. Like models that are felicitous falsehoods, such models exemplify features they share with their targets and represent their targets as having those features. And like all exemplars, they symbolize selectively. The model house shares a vast number of features with houses it serves as a model of. It does not, however, exemplify all of them. Not being made of marshmallow is a feature that the model and its target share; but it is not exemplified by the model or imputed to the target. *Drosophila melanogaster* is a model organism that exemplifies general genetic features bearing on heritability of traits, susceptibility to mutation and the like; and it represents other organisms as having those features. It does not, however, in its standard uses, exemplify features that are peculiar to invertebrates, flying animals, or insects. The model house exemplifies its floor plan, the placement of windows, the height of the roof, and the like. It typically does not exemplify the colors of the walls in the different rooms, the way it is furnished, and so forth. Features that are not exemplified are representationally inert. Those features, whether shared with the target or not, are negligible. Understanding is not impeded by neglecting them.

13 Borders and Crossings

Disciplinary Demands

There are often multiple ways to understand a given topic, for the relevant information may be integrable into diverse accounts. A geneticist, a nutritionist, and an epidemiologist are apt to understand the increasing incidence of diabetes differently. Even if they draw on the same empirical data, each interprets those data in terms of different models and appeals to different background theories for support. Each highlights features that the others marginalize. Different histories of the same episode show similar diversity. One may focus on political factors, another on economic considerations, a third on the effects on domestic life. This enables them to discern different patterns in the data, and draw different inferences from them.

Can works of art afford historical understanding? Or are the commitments of art and history too disparate for a single work to satisfy the demands of both? To examine the issue, I focus in this chapter on Maya Lin's Vietnam Memorial. As a monument, it denotes and commemorates something deemed to be historically significant. It makes reference to the past. That alone does not indicate that it affords any understanding of the events it refers to. The Washington Monument affords no understanding of the first US president. Nor, if it does afford some understanding, should we automatically conclude that it affords a *historical* understanding. The question is *how* the Vietnam Memorial refers to the episode it commemorates. Does it do so in ways that comport with the expectations of the discipline of history?

A discipline is constituted in large measure by methodological and normative framing commitments that are reflectively endorsed by the

legislating members of a particular realm of epistemic ends. Natural science is committed to accepting only findings that are intersubjectively accessible, backed by empirical evidence, preferably testable, and defeasible. Although romantic poetry affords an understanding of the natural world, it does not afford a *scientific* understanding, for it does not respect science's framing commitments. Mathematics demands proof. Although there is plenty of evidence for Goldbach's conjecture, that evidence does not figure in a *mathematical* understanding, since instantial induction—regardless of the number of instances—does not amount to proof. History, like natural science, requires that evidence be publicly accessible, publicly assessable, and defeasible. The evidence must concern past events. A historical understanding of those events typically treats them as sequential and shows why, in the circumstances, they were to be expected (Hempel, 1965). This is far from a detailed characterization of what disciplinary understanding consists in. My point is to emphasize that disciplines are more than ways of marking off turf. They carry commitments about how a particular subject matter is to be properly understood.

One critical question is how strong a discipline's framing commitments are. Should they be construed as inviolable requirements, as rules of thumb, or as something in between? Disciplines evolve. So presumably framing commitments cannot be utterly rigid. But how flexible they are or should be, and whether this varies from one discipline to the next, remain open questions. In this chapter I am concerned primarily with history. I shall treat its framing commitments as touchstones that establish strong presumptions but not rigid requirements. Some works that are treated as uncontroversially providing historical understanding seem not to fully conform to any plausible rigid rubric.

The question I address here is whether it is plausible to contend that works of art, functioning as such, can embody and convey the sort of understanding of a period that the discipline of history demands. Obviously works of art can serve as historical evidence. Like other artifacts, they are products of an era and can be used to shed light on what was going on during that era. Poe's "The Raven" and Mussorgsky's *Night on Bald Mountain* can, for example, serve as evidence of the nineteenth-century fascination with the macabre. My question is whether works of art can wholly or largely constitute an understanding of an episode, where an episode is a constellation of related events that take place over a limited span of time.

I will suggest that the answer is yes. In particular, I will urge that some monuments, and perhaps some other works of art, are jointly artistic and historical symbols. Nor is their performing both roles a coincidence. They embody and convey a historical understanding of the episode they concern, and they do so because of their aesthetic functions.

Must Understanding Be Discursive?

Grimm (2008), Khalifa (2012), and others contend that understanding is keyed to explanation. This at least strongly suggests that those who understand a phenomenon can give or accept a verbal account of it. If so, all understanding is discursive. In that case, my contention that dance embodies understanding is incorrect. They are surely right that formulating and articulating an explanation of a phenomenon is a manifestation of understanding it. But Lipton (2009) makes a compelling case that the ability to explain is not necessary. Someone could, he maintains, display her understanding of the retrograde motion of the planets by properly manipulating an orrery.[1] Moreover, I would add, someone could teach why the planets display retrograde motion by giving a nonverbal demonstration using an orrery. With the orrery, we can just see why, from the perspective of an observer on Earth, Mars appears to reverse its direction as it orbits. If we cannot correctly express our understanding in words, that seems at best a minor defect. Still, our inability to explain might just be put down to inarticulateness. We apprehend and appreciate what is going on, but words fail us. Lipton's second example goes deeper. Galileo demonstrated that gravitational acceleration is independent of weight by showing that the alternative leads to a contradiction. We understand that this must be so—that it could not be otherwise—even if we have no idea why gravitational acceleration is independent of weight, or what an explanation of that independence would look like. In general, it is very hard to explain why a is independent of b. A third example, this one mine, is Richard Feynman's dropping an O-ring into a glass of ice water to demonstrate that O-rings lose elasticity in the cold. This enabled scientific novices to understand why the O-ring's failure caused the space shuttle *Challenger* to explode. Feynman's point could be captured in words. Indeed, expert witnesses testifying before a congressional committee had put it into words. But the congressmen did not understand the verbal explanations. The demonstration did the trick

(Feynman, 2001). These examples show that understanding need not be discursive. It is possible to understand something without being able to put one's understanding into words, and possible to understand something without understanding a lucid verbal explanation of it in a language one speaks fluently.

Lipton does not convince everyone. Khalifa (2012) contends that the understanding Lipton's examples provide is somehow deficient. The best understanding of the phenomena would be explanatory. He gives no reason to think so. What could supply a better understanding of the independence of an object's weight from its rate of fall than Galileo's demonstration that it could not be otherwise? Granted, that understanding might prompt us to wonder why it could not be otherwise. And to understand that would likely require a good deal of theorizing. So embedding the demonstration in a tenable theory would be worthwhile. But that does not show that the unembedded understanding that the demonstration provides is in any way second best.

Grimm (2008) advocates expanding our conception of explanation so that Lipton's demonstrations count as explanations. Then to demonstrate that *p* is to explain that *p*. This seems ad hoc. Explanations are discursive; they say why something is so. Demonstrations exemplify; they show why something is so. We currently lack a workable conception of explanation that comprehends demonstration.

Whether or not demonstrations afford the best sort of understanding, whether or not they are (perhaps in an extended sense) explanations, there is little doubt that they afford *some* understanding of the phenomena they concern. As we have seen, understanding admits of degrees. Children, novices, and cutting-edge inquirers often have a level of understanding that falls short of our (realizable or unrealizable) ideal. That is all that we need. For our purposes here, the issues raised by Lipton's critics are irrelevant.

Still, the examples come from science. Our question is whether the point extends to history. To answer that, we need some characterization of what is required for the distinctive sort of understanding that works of history provide. I cannot give a criterion for being a work of history, or for constituting a distinctively historical understanding. But some features stand out. A work that embodies and conveys a historical understanding is a selective account of the past. Selectivity is mandatory because a mere amalgamation of everything that is known about an episode of any

complexity—the Reign of Terror, the moon landing, the ratification of the Nineteenth Amendment—would be such a motley chronicle of events that it would make no sense. Selectivity fosters coherence. The various elements that comprise a work of history should collectively constitute a coherent account. The account should be as of events structured sequentially in time. Williams (2002, 243) maintains that the account itself should be structured sequentially. If so, it will have a narrative form, with a beginning, middle, and end (White, 1965).

It is not obvious that narrative form is strictly required. Garry Wills's *Lincoln at Gettysburg* (1992) is largely thematic. It weaves the Gettysburg Address into several historical narratives: a political narrative, a rhetorical narrative, and a narrative about the nineteenth-century cult of death. Each enriches and deepens the understanding that the others provide. Arguably, all three are needed to establish Wills's thesis that the Gettysburg Address fundamentally changed America's conception of itself as a nation. This suggests that a narrative form may be just one, albeit fairly standard, way to structure events in time so as to make historical sense of the episode they belong to. A history should be accurate (Williams, 2002). Just how accurate, and in what respects it should be accurate, is controversial. Thucydides composed the speeches he purportedly quoted in *The History of the Peloponnesian War*, including Pericles's famous funeral oration. No one knows what, if anything, Pericles or the other orators actually said. Contemporary historians often use hedging verbs. They say that their protagonists *would have* done this, or *must have* known that, or *were bound to* have accepted the other. These locutions signal that although the historians have no specific evidence to back their claims about the protagonists, such actions and attitudes were characteristic of people relevantly like the protagonist in the period under discussion. Moreover, like the sciences, histories often make use of schematized models, such as 'triangular trade', 'coup d'état', and 'oligarchy'. These, like scientific models, idealize, simplify, and thereby depart from truth.[2] When they contribute to historical understanding, they are felicitous falsehoods. Rather than saying that historical accounts must be true, it is preferable to say that they must be true enough. Moreover, a work of history must be backed by evidence, and the chain of evidence should bottom out in primary source material—firsthand testimony, contemporary documents, artifacts of the period under investigation. To be sure, historians can draw on the work of other historians. Appeal to secondary sources

is not ruled out so long as the chain of evidence, however long, terminates in acceptable primary source material. Histories, being empirical, must be defeasible. Information about the past must be capable of discrediting a historical account. I will call anything that embodies and conveys a historical understanding a work of history.

Nonnarrativity

If what is suitably accurate is true enough, documentaries such as *The Sorrow and the Pity* qualify as works of history. Although they are not primarily verbal, they are nonfictional cinematic representations that provide a historical understanding of the episodes they concern. I will go further and urge that static works can also provide historical understanding. Before I do so, however, I want to return to the scientific cases discussed above.

The manipulation of the orrery shows why Mars seems to change direction. It highlights consequences for terrestrial observers of the fact that the Earth travels around the sun more quickly than Mars does. Galileo's thought experiment is a reductio. It makes manifest the inconsistency inherent in the assumption that gravitational attraction and weight are linked; that is, it shows that the assumption cannot be correct. Feynman's demonstration shows why the O-rings could not create a seal: the material, having lost elasticity, could not expand to seal the gaps that the O-rings were supposed to seal.

In all three cases, the primary mode of reference is exemplification. The orbs on the orrery exemplify motions; the postsubmersion O-ring exemplifies inelasticity. They instantiate and thereby refer to some of their properties, making them salient, and enabling interpreters to appreciate them and their significance. As we recognized earlier, in Galileo's thought experiment nothing actually falls. Since nothing has the property of falling, nothing exemplifies that property. Rather, object-representations (either mental representations, or publicly available descriptions or depictions) exemplify an abstract mathematical property that they share with actual falling bodies. The thought experiment exemplifies the independence of the abstract mathematical property that in material objects is a measure of weight from the abstract mathematical property that in material objects is a measure of rate of fall. It thereby exemplifies the independence of weight from gravitational acceleration. In each case, the symbols that embody and convey an

understanding of a phenomenon are not primarily discursive or representational. They afford understanding via exemplification.

I have argued that in the arts, nondenoting symbols embody and convey an understanding of a topic (Elgin, 2002). It would not be surprising, then, that works of art can afford an understanding of the past. This involves more than conveying discrete bits of knowledge about the past. It is easy to see that a Gainsborough portrait conveys the knowledge that eighteenth-century British merchants wore ruffles, and that a Heda still life conveys the knowledge that in the seventeenth century the Dutch had access to lemons. These are isolated facts. They provide evidence about the past. A historical *understanding*, on the other hand, is an integrated account of a past episode that involves a sequence of events. So to claim that works of art embody *an understanding* requires saying more than that they convey separate factoids about the past or even that they convey factoids that can serve as evidence about the past. And to say that they embody a *historical* understanding requires saying that they do so in terms of a selective sequence of events. A documentary plainly satisfies these requirements. I suggest that a static work of art can do so as well.

Monuments are works of public art, expressly designed to commemorate what are taken to be historically important episodes. If any works of art embody and convey historical understanding, monuments are apt to do so. I will argue that Maya Lin's Vietnam Memorial (henceforth the Memorial) does exactly that.

The Memorial is located on the Mall in Washington, DC, a long parkland dotted with monuments that collectively constitute a patriotic gloss on US history.[3] The Mall is not flat. At the Memorial's location there is a hill. But the Memorial is not on the hill. Rather, it is gouged out of the hill. It is a gash in the land, an open wound. That wound in the physical land is a metaphor for a wound in the country, the political land. The Memorial is made of black stone. It consists of two conjoined triangles that share an edge and open out toward the Mall in a V shape with a 125° angle. One triangle points toward the Washington Monument; the other, toward the Lincoln Memorial. The Vietnam Memorial thus locates the episode it commemorates with respect to the fundamental ideals on which America was founded. It alludes to those ideals.

Etched in the stone are the names of the 58,272 US military personnel who were killed or declared missing in action during the Vietnam War. The

names are all the same size and are in the same font. Rank is not included. Every one deserves to be remembered. Each is represented as being as important as any other. The recording of the names thus exemplifies the ideal that all are created equal, an ideal articulated in the Declaration of Independence, operationalized in the Constitution, and reinforced in Lincoln's Gettysburg Address. The stone is highly polished; viewers see their own reflections as they look at the names. This invites the living to identify with the dead and to reflect on their own attitudes toward the Memorial and what it commemorates. The order of the names is the order in which their bearers died or went missing. The shape of the Memorial displays the course of the war, starting small in 1959, ascending to the apex in mid-1968, then descending to 1975—from a trickle to a flood to a trickle. The configuration thus constitutes a narrative arc, displaying the trajectory of the war in terms of its American casualties. Collectively the names represent the war's cost, not in money, but the irrecoverable cost in lives.

The Memorial also symbolizes American ambivalence about the war. Among the ideals that were ensconced at the founding of the country are freedom of speech, freedom of assembly, and the right to petition the government for redress of grievances. The Memorial's pointing across the Mall to the Washington Monument and the Lincoln Memorial can be read as alluding to the patriotism of those who exercised these rights by assembling on the Mall to protest against the war. Then there is the wound. Is it simply a mark of the loss of lives? Is it an honorably acquired war wound? Or is it a wound on the honor of the country that professes ideals that it does not live up to? Did the war instantiate the ideals or betray them? The legacy of the war is controversial. The ambiguity of the gash expresses the nation's ambivalence about the war.

It seems obvious that the Memorial conveys an understanding of the war. But does it convey a historical understanding? Let's look back at the touchstones mentioned above. The Memorial is structured sequentially in time. The names are listed, not in alphabetical order, or by rank or unit or age or state, but in the order in which their bearers died or were declared missing in action. So you can read the course of the war by reading the names in order. The names could have been put in columns of the same height. Instead, there is a quite literal narrative arc, peaking at the apex of the triangles, displaying the escalation and de-escalation of the war. The work is grounded in primary sources—the Department of Defense casualty lists and

documents pertaining to those missing in action. The work is defeasible. It would be discredited if, for example, the list was irresponsibly augmented or truncated ('irresponsibly', because a few erroneous inclusions or exclusions would presumably not discredit the work) (Kvanvig, 2003).

The work is plainly selective. Much is omitted. There is no recognition of, for example, the geopolitical context of the conflict, the Machiavellian machinations of politicians, the perspective of the Vietnamese people. But all works of history are incomplete. There is simply too much information about any historical episode to incorporate it all into a coherent, comprehensible account. The historian has to solve complex problems of selection—what to include and what to omit, at what level of precision to present the material, from what orientation to address the issue, what themes to emphasize. She must, that is, decide what features of the episode her account will exemplify. The Memorial omits any explicit recognition of the reasons for the war. It tells its story by counting the cost. And it measures the cost in lives according to the metric 'all are created equal'. The Memorial does not and does not purport to convey a comprehensive understanding of the war. No single work of history could do that. Still, you will enrich your understanding of this period in US history, if you properly interpret it and incorporate what it conveys.

We might doubt that an expert on the war could gain an understanding from the Memorial. He would presumably already grasp the insights and orientation that the Memorial presents. This is not so obvious. What I have called the narrative arc of the piece might make manifest aspects of the escalation and de-escalation that he had overlooked. If so, the Memorial would function for the historian in much the way that a work written by another historian would. It would serve as a secondary source. Still, it is perhaps unlikely that the understanding of a scholar who specialized in the war would be much enhanced by the work. If the Memorial is a work of history, it is a work of popular history, functioning not primarily to deepen the understanding of experts, but rather to deepen the understanding of ordinary people.[4]

Arguably, then, the Memorial affords historical understanding. It bears at least one interpretation that satisfies plausible requirements for being a work of history. A separate question, though, is whether it does so by functioning as a work of art. Or does it simply play two distinct, mutually independent roles?

Aesthetic Functions

The history of aesthetics is littered with failed attempts to articulate necessary and sufficient conditions for being, or functioning as, a work of art. At best we have clear cases, and what Goodman called symptoms of the aesthetic—that is, defeasible but presumptive indications that an item is functioning aesthetically. The Memorial displays Goodman's symptoms. Its doing so underwrites its capacity to provide historical understanding. That being so, then at least by Goodmanian standards, it is functioning as a work of history by functioning as a work of art. The symptoms are these:

(1) Syntactic density, "where the finest differences in certain respects constitute a difference between the symbols" (Goodman, 1978, 67–68). Had the stone been less reflective, or the orientation toward the Mall been different, or the Memorial been placed atop the hill, it would have been a different symbol.

(2) Semantic density, "where symbols are provided for things distinguished by the finest differences in certain respects" (Goodman, 1978, 68). The names could have been given with ranks, units, and so on. Fonts could have been different for military personnel of different ranks. That is hardly unprecedented. It is what is done in the war memorials in Canterbury Cathedral. The choice to omit such markers was a choice from among a dense range of alternatives.

(3) Relative repleteness, "where comparatively many aspects of a symbol are significant" (Goodman, 1978, 68). As we have seen, a vast number of features of the Memorial symbolize: its position, orientation, shape, configuration, the order of names, the font, the polish of the stone, and so forth.

(4) Exemplification, "where a symbol, whether or not it denotes, symbolizes by serving as a sample of properties it literally or metaphorically possesses" (Goodman, 1978, 68). The number of casualties, the rate of casualties, the names of the casualties, the relation of the Memorial to other monuments, and the shape of the Memorial are literally exemplified; the ideals, realized or betrayed, the wound in the body politic, and the ambivalence about the war are metaphorically exemplified.

(5) Complex and indirect reference, "where a symbol performs several integrated and interacting referential functions, some direct, some mediated through other symbols" (Goodman, 1978, 68). The triangles pointing toward the Lincoln Memorial and the Washington Monument make direct reference to those monuments. Via that reference, the Memorial makes indirect reference to the ideals that the two monuments stand for. It indirectly refers to their metaphorical exemplification of 'all are created equal', and in turn refers back to the exemplification of that ideal in the

names on the Memorial itself. It refers to the exercise of the constitutional rights by the antiwar protesters and both makes reference to and calls into question the nature of the wound symbolized by its shape and position. And so forth. This makes it plausible that it is by functioning as a complex aesthetic symbol that the Memorial conveys an understanding of history.

One might, however, wonder whether the Vietnam Memorial is a special case. In some respects the answer may be 'yes'. The obvious narrative arc is distinctive. But as I suggested earlier, narrativity may be optional. Then the relevant requirement is that a work of history be "a selective account of the past, structured sequentially in time, that is designed to make sense of it" (Williams, 2002). Regardless of what we say about how representative the Vietnam Memorial is, I do not mean to suggest that every memorial or monument functions as a work of history. A work can commemorate an event without affording any historical understanding of it. But the criteria for a history can be realized at different levels of generality or abstractness. So a suitably accurate memorial that commemorates wars or battles, or the reigns of kings or the tenures of poets laureate in a temporal sequence, might function as a work of history, even though it lacks the specificity of Lin's Memorial.

Moreover, a work of art that is not expressly designed to commemorate a *particular* historical episode might plausibly function as a work of history. Consider, for example, *The Disasters of War*, a series of prints Goya made in response to the Peninsular War (1807–1814) and its aftermath. They focus on the horrors of war: humiliation, degradation, rape, torture, mutilation, desecration, destitution, death. They concern not the causes of war, but the effects—in particular, the effects on the innocent: women, children, non-combatants, priests. The prints open out into an indefinite space, suggesting that the mayhem flowing from war is random. It could occur anywhere to anyone. There is no differentiation between friend and foe. Atrocities are perpetrated by both sides. Nothing in the prints points to hope, charity, or even rationality. War is monstrous. It strips people of reason and empathy and turns them into savages. Many of the prints can be correlated with specific events in the Peninsular War. Interpreted as a commentary on *that* war, they function in much the way the Vietnam Memorial does. But the series might also be interpreted more generally. It is a template or schema into which you can insert any war you please: the Trojan War, the Anglo-Boer War, the Chaco War, the Vietnam War. It makes no difference. The atrocities Goya depicts will be found there too.

Does *The Disasters of War*, interpreted as a general comment on war, convey a historical understanding? Maybe. If it is true that wars generally instantiate the depicted atrocities, then the series is a schema. Each instance affords a historical understanding of a particular war. And if that is so, then it is reasonable to say that the schema figures in a historical understanding of war as such. According to Hempel (1965), lawlike generalizations about human conduct figure in or are tacitly presupposed by explanations in history. Treated as a schema, *The Disasters of War* is akin to those generalizations. If Hempel's generalizations figure in an understanding of history, it is plausible that Goya's prints do too.

I have been assuming that a work of history can be purely descriptive. It can present a cogent account of what happened. It need not provide a causal explanation of why it happened. If this is right, then Hempel's generalizations need be no more than statements of robust Humean regularities. They may afford a basis for prediction, but need not contribute to explanations of how it came to pass that events played out in the way they did. So although Lin's Memorial provides no explanation of the course of the Vietnam War, and Goya's prints provide no explanation of why wars lead to the commission of atrocities, that is not by itself an argument against treating them as affording a historical understanding. Arguably they function like Lipton's orrery or Feynman's O-ring.

Maybe Not

Even so, we might want to resist the claim that a work of art, functioning as such, affords genuinely historical understanding. It is not implausible to think that aesthetic standards are too subjective to underwrite the epistemic standing history demands. Earlier I argued that aesthetic standards are not wholly subjective. As we have seen, in discussions about art, reasons are adduced and disputed. Still, the range of reasons aesthetic judgments can appeal to may seem too open ended to satisfy the disciplinary demands of history. Thus far, my discussion has appealed to loosely characterized touchstones. Perhaps we should tighten them, excluding symbols whose function turns on their displaying objectionable symptoms of the aesthetic.

One worry is that works of art often admit of multiple acceptable interpretations. Written histories strive for and often achieve univocality. Different histories may provide different interpretations of a given event. But each

individual work is apparently supposed to provide only one. We can accommodate this worry by saying that it is only under an interpretation that a work of art functions as a work of history. This is a trivial concession, since it is only under an interpretation that a paper in the *Journal of American History* functions as a work of history. Only interpreted symbols are about something.

Perhaps the point is that unlike the journal article, the work of art lacks a standard interpretation. Every competent historian knows how to read the journal article, and they all pretty much agree about what it says. The same cannot be said for a work of art. Connoisseurs, such as Greenberg and Fry, may permanently disagree. If this is the worry, we could insist that only under a relevant, accessible, nonidiosyncratic interpretation can a symbol function as a work of history. Except perhaps for the bit about the antiwar protests, my interpretation of the Vietnam Memorial satisfies this requirement. It is a standard, relevant, accessible, and nonidiosyncratic interpretation of the monument (see Lin, 2000). It is intersubjectively accessible and assessable. Reasons can be given for it and for the interpretation of the war that it provides. So the restriction would not exclude works of art that have standard interpretations from functioning as works of history under those interpretations. In any case, the restriction is ill advised. An account of the siege of Stalingrad written in an obscure code would be nonstandard, but assuming it was duly researched it would be a work of history. The issue is not whether a work has a standard interpretation, but whether it has an intersubjectively accessible and assessable interpretation. Insofar as the code could be learned, an otherwise historically creditable account of a past episode written in code would qualify as a work of history.

Maybe the worry is not semantic, but syntactic. Should we perhaps deny that works of art afford historical understanding because they are not verbal? There is, as far as I can see, no reason to privilege the verbal over other modes of symbolization. And if we do, we exclude not just prints, monuments, and documentary films, but also diagrams, charts, and maps. To restrict historical understanding to that which is captured in a language would be costly. We would have to deny, for example, that Minard's map affords an historical understanding of the fate of Napoleon's army in the Russian invasion. Semantic density seems both innocuous and unavoidable. Syntactic density is sometimes welcome.

I suggested earlier that the historian's solution to the problem of selection is a matter of deciding what aspects of an episode are to be exemplified. He

chooses events and event descriptions that highlight the features of the episode that he considers central to the sort of understanding he seeks to provide. This determines the items he includes, the vocabulary he uses to describe them, and the orientation he takes toward them. Describing the construction of the first transcontinental railroad, Howard Zinn says this: The Central Pacific Railroad

spent $200,000 on bribes in Washington to get 9 million acres of free land and $24 million in bonds, and paid $79 million, an over payment of $36 million, to a construction company which was really its own. The construction was done by three thousand Irish and ten thousand Chinese, over a period of four years, working for one or two dollars a day. (Zinn, 1980, 254)

This description configures the facts in such a way that they exemplify the plutocrats' venality, corruption, collusion, and callous indifference both to workers, who were members of despised minorities, and to ordinary citizens, who might reasonably object to their legislators being bribed, their land being given away, and bonds being sold to them at artificially inflated prices. It is hardly laudatory, but that does not discredit Zinn's description. The passage exemplifies a pattern to be found in behavior of the plutocrats behind the construction of the first transcontinental railroad. This is not the only pattern. A different account might highlight the technological achievements, the logistical difficulties, the economic benefits the railroad supplied, or even cast the protagonists in a less unfavorable light. Then different features of the construction of the railroad would be exemplified.

Some histories describe seemingly insignificant or irrelevant items in order to exemplify something of importance. Early in *Lincoln at Gettysburg*, Wills describes the layout of the Gettysburg cemetery. The Union dead are buried according to the states they came from, with no recognition of differences in importance either of the state or of the soldiers. "By considerable ingenuity in grading the cemetery's incline, and by arranging the graves in great curving ranks, [William Saunders, the cemetery's architect] avoided preferential treatment of states or inequality in the ranks of the fallen" (1992, 22). The book's central theme is that the Gettysburg Address changed America's self-conception from a union of states to a union of free and equal citizens. The cemetery design and Lincoln's speech, Wills suggests (1992, 30), are mutually reinforcing because they exemplify the same values.

Exemplification is thus critical to historical understanding. The historian juxtaposes events, agents, actions, and standard conditions so that

collectively they exemplify a pattern or trajectory in the course of events. An understanding of an episode must make sense of it. To do that, it must highlight some features and ignore or downplay others.

Written works of history typically deploy a variety of modes of reference. They denote the events and agents they concern. They exemplify features and patterns in the play of events. They refer indirectly to events, actions, and authorities that are not explicitly mentioned. They cite their sources and tacitly allude to whatever makes those sources credible. They may also refer indirectly to other histories that support or clash with the position they are taking. This can be done by a variety of means, including the selection of events deemed worthy of mention, the order of exposition, even word choice. Some indirect reference is unavoidable. A history of the Bogside Massacre must say where it took place. If it locates the event in Derry, it thereby implicitly slants toward the Irish; if in Londonderry, toward the British. Either way, it indirectly refers to a history of animosity. Each name is freighted with significance, making partisan reference to Irish history. No neutral name is available. Complex and indirect reference too seem unavoidable.

Goodman's remaining symptom is relative repleteness. A symbol is *relatively replete* to the extent that it symbolizes along multiple dimensions. Goodman illustrates the difference between repleteness and attenuation by comparing a Hokusai drawing of Mount Fuji with an electrocardiogram (EKG). Even if we imagine the same configuration of ink on paper playing both roles, there is an enormous difference in the ways they symbolize. Only two features matter in the EKG—the shape of the wave and the frequency with which it repeats. In the drawing, any of a vast number of features might matter: the size, shape, color, and intensity of the line; the weave, texture, and material of the paper; differences, however subtle, in shading, from one part of the picture to the next; and so on. The picture is inexhaustible. There is always more to be found.

One might think the same holds for the EKG. In principle, there may also be subtle, medically significant differences in wave patterns that cardiologists have previously overlooked. So it too bears repeated inspection; but only up to a point. The number of dimensions along which it symbolizes is fixed in advance. The cardiologist does and should overlook the fact that the line gets paler as the test continues. That is an indication that the printer's toner is running low. It is not her problem. Moreover, the machine has

an antecedently determined level of precision. Past a certain point, further wobbles in the curve are insignificant. They might be manifestations of background vibrations or noise in the system or fine-grained irregularities in the paper the graph is printed on. They should be disregarded. Perhaps one could glean a bit more information than we currently do from a given EKG, but cardiologists rightly think that the information the printout provides is exhaustible. If they want to learn more, they move on to different tests.

What should we say about works of history? Written histories are, I suspect, more replete than EKGs. But they are apt to be less replete than works of art. Although they symbolize along several dimensions, arguably at least, the number of dimensions is limited and intersubjectively agreed upon in advance. That a history includes or omits the names of the founding members of the Royal Society may be significant. Even the order in which the names are listed may be significant. But it is doubtful that the omission of maiden names of the founding members' maternal grandmothers is significant. If we are loath to consider works of art to be works of history, perhaps the reason lies here. When interpretation is too open-ended, intersubjective agreement is sacrificed. Scholars become permanently at odds as to what a particular work conveys about the past.

I have argued that a fundamental difference between scientific and aesthetic understanding lies in the fact that science puts a premium on intersubjective agreement; aesthetics does not. To attain such agreement, science limits precision and constrains repleteness. If only a few antecedently recognized dimensions of a symbol matter, and only up to an antecedently recognized point, agreement among the knowledgeable is readily achieved. Scientists build on one another's findings. It is therefore important that they can tell exactly what those findings are and why and to what extent they are to be trusted. They sacrifice repleteness and precision to achieve agreement. Because artists do not build on one another's works in the same way, they have no incentive to make such sacrifices. That connoisseurs never agree about whether the weave in the paper of the Hokusai drawing or the direction of *Le Compotier*'s brush strokes contributes to its aesthetic function makes no difference. Artists who are influenced by Hokusai or Cézanne can decide for themselves whether to take that feature of the work to be significant. Historians fall somewhere in between. They do not build on their predecessors in the way that empirical scientists do. So agreement

about what exactly a given historian accomplished is not mandatory. But it is not irrelevant either. What has been done sets the stage for what comes next. So perhaps a reluctance to treat works like the Vietnam Memorial or *The Disasters of War* as works of history stems from a well-founded desire to achieve a measure of accord. In that case, we should say that although such works embody and convey something close to a historical understanding, that understanding remains beyond the boundaries of history proper.

The question is where to draw the line. We could decide that because of their potentially unbounded repleteness and precision, works of art do not afford genuinely historical understanding. Or we could decide that repleteness and precision need not inhibit and may even enhance their capacity to afford historical understanding. The decision turns not on the sort of information the items convey or on the accuracy or reliability of that information. It turns on how historians want to configure their discipline— what framing commitments they want to endorse, what restrictions they will place on the accounts it sanctions, and so on. And that turns on how they want to use those accounts. If historians want to be able to build on previous histories in the way that scientists build on previous science, they should limit precision and repleteness. If the growth of historical understanding takes a different path, then perhaps not. The decision depends on the relation of scholars to one another, not on their relation to their object of study.

14   Fallibility's Payoff

Skeptical worries often emerge from the recognition that the evidence for a contention does not ensure its truth.[1] It might therefore seem that, having scuttled the truth requirement on epistemic acceptability, I can evade such worries. Sadly, I cannot. For the evidence for a contention typically does not ensure that it is true enough either. In this respect, veritists and nonveritists are in the same boat. The way out might seem obvious. Acknowledging our epistemic vulnerability, we should endorse fallibilism, thereby evading skepticism.

I begin this chapter by investigating fallibilism with respect to knowledge. On most construals, it seems to be a concession to epistemic frailty. Our evidence is almost inevitably too sparse to provide the level of justification we would like. We need to live with this. I suggest that such a position is less stable and more problematic than it might first appear. I go on to consider fallibilism with respect to understanding. I argue that, far from being a concession to frailty, it is a source of strength. It both prompts and enables us to deepen our understanding in ways that an infallibilist position would not. I urge that the capacity to make mistakes is an epistemic achievement rather than a failing. I then discuss how actual errors sometimes mark epistemic advances. Finally, I argue that a fallibilist theory of understanding is epistemically richer than an infallibilist one would be. Weaving in mechanisms for detecting, correcting, and compensating for errors strengthens the fabric of understanding.

Fallibilism

Although many philosophers claim to be fallibilists, it is not entirely clear exactly what fallibilism amounts to. It teeters between skepticism and

dogmatism. To acknowledge that despite the evidence I could still be wrong seems an entirely appropriate confession of intellectual humility. It recognizes a gap between the support my reasons supply and what it would take to make me infallible. But an epistemic agent's claim that she could be wrong is more complicated and less obvious than it first appears.

The contention that we are infallible seems unacceptably arrogant. But epistemologists who accept 'When you know, you cannot be wrong' and maintain that epistemic agents sometimes know are committed to the view that those agents are infallible with respect to what they know. What does it take to underwrite infallibility? Two candidates are plausible: either the premises entail the conclusion or the premises give the conclusion a probability of 1. Both are simultaneously too weak and too strong. Entailment alone cannot secure infallibility. What is required is entailment from known premises. Unless the premises are known, they convey no right to be sure; for all an epistemic agent can tell, one of her premises might be false. Entailment from a false premise surely conveys no right to be sure. If the agent has to ensure that the premises are known, a regress looms. Something similar occurs if we seek to ground infallibility in probability. It is not enough that the evidence give the conclusion a probability of 1; that evidence must itself be epistemically secure. What justifies treating an item as evidence? Does each bit of evidence also have a probability of 1? Here too a regress looms.

Luckily, we need not attempt to block the regresses. For the standards in question are too demanding in any case. To avoid skepticism about the external world, we must settle for less. Evidentialists maintain that evidence must probabilify a conclusion (Adler, 2002); but there is a threshold—perhaps a movable threshold (S. Cohen, 2014)—with a probability of less than 1. Reliabilists maintain that the belief that constitutes knowledge must be produced or sustained by a reliable process—just how reliable may vary with context (DeRose, 2011). But they admit that even an epistemically reliable process sometimes fails. There is always a gap between the justifiers and the conclusion they support.

A critical question concerns how to interpret the 'might' in 'I still might be wrong'. The inability to secure an entailment shows that there is a logical or conceptual possibility of error. Regardless of the strength of my evidence or the reliability of my method, it is still logically or conceptually possible that I am wrong. This no doubt is true, but what should we make of it?

Perhaps whenever there is the remotest possibility of error, I should suspend judgment. If this is what fallibilism contends, it amounts to (or comes perilously close to) skepticism. If, on the other hand, fallibilism maintains that assuming the gap is not too broad, it makes no difference that I might be wrong, it is hard to distinguish fallibilism from dogmatism. The fallibilist and the dogmatist then treat the very same contentions as knowledge, and for the very same reasons. Evidently, the only difference is that the fallibilist displays a sort of mock modesty, while still claiming to know.

If the fallibilist sides with the skeptic, her predicament may be worse. However reliable her methods and however solid her evidence, everyone concedes that some chance of error remains. Ought the agent therefore suspend judgment? To accept a contention is to be willing and able to use it as a premise in cognitively serious assertoric reasoning or as a basis for cognitively serious action (L. J. Cohen, 1992). Suspension of judgment deprives the epistemic agent of that resource. Sometimes such deprivation is entirely reasonable; to refuse to use a contention as a premise in assertoric reasoning or as a basis for action is a mark of epistemic responsibility if the contention's epistemic standing is doubtful. The detection of what appeared to be a distinctive sort of gravitational wave was taken in 2014 to support the hypothesis that the universe began with a Big Bang ripple of cosmic inflation. But astrophysicists recognized that there was a slight chance that the observed wave pattern was due to cosmic dust. (This turned out to be true.) Because they thought the region where the wave was detected was relatively dust free, the dust alternative was considered improbable. Still, it was probable enough that an epistemically responsible astrophysicist would have to suspended judgment. She could, were she so inclined, use the cosmic inflation hypothesis in hypothetical reasoning, but until the controversy was resolved (should we say until the dust settled?), she should use neither hypothesis assertorically. Suspending judgment in a case like this seems entirely reasonable. But widespread suspension of judgment is more difficult and costly than it first appears. Action requires assuming that things are one way or another. Suspending judgment leaves a person paralyzed. She has no reason to favor any particular course of action. She ceases to be an agent.

Many epistemologists urge that a mere logical, conceptual, or metaphysical possibility does not put a contention in epistemic jeopardy (Adler, 2002). So the mere logical, conceptual, or metaphysical possibility that we

are wrong about *p* gives us no reason to suspend judgment about whether *p*. The possibility that concerns us, they maintain, is epistemic. And some logical, conceptual, and metaphysical possibilities are not epistemic possibilities. The question fallibilism poses, then, is whether it is possible to know that *p*, even though it is *epistemically* possible that ~*p*. Some, such as David Lewis (1999b) and Fred Dretske (2000), argue that we need not and ought not concede this. Our evidence, they contend, regularly supplies conclusive reasons for knowledge claims. When it does, there is no epistemic possibility that we are wrong. In such cases, we are infallible. When we are infallible, we ought to be dogmatic. For there is no real (that is, epistemic) possibility of error.

The difficulty is that we are seldom if ever in a position to rule out all potential defeaters to a knowledge claim. As Lewis and Dretske recognize, to arrive at the position they advocate requires circumscribing the range of potential defeaters that we need to accommodate. We need only, they maintain, consider what occurs in 'nearby' possible worlds or is among the relevant alternatives. Remote possibilities and irrelevant alternatives can be ignored. But a little reflection on the fates of our peers—similarly situated epistemic agents who turned out to be wrong—undermines confidence in this conclusion (Adler, 1981).[2] The most familiar examples are lottery cases. The holder of the winning ticket had as good reason to think that she would lose as the losers did. Her winning the lottery was surely a remote possibility.[3] But she was wrong. Since she did not know she would lose, and the other ticket holders were in the same epistemic position, it is hard to see how they could know that they would lose. Less fortunate cases include Vogel's (1990) car theft victims. They had as good reason as their peers to think that they knew where their cars were parked, but, their cars having been stolen, they were wrong. Those who were wrong were as justified as those who were right. That's the lesson of the gap. We move too quickly when we dismiss some metaphysical, logical, or conceptual possibilities as too remote to qualify as epistemically possible, or as sufficiently epistemically possible to matter.

Dogmatism is hard to square with epistemic misfortune. Fallibilism recognizes that epistemic misfortune is always possible. This suggests that we construe fallibilism as contending that an epistemic agent can know that *p*, even though her justification is fallible, where for justification to be fallible is for it to be the case either that her justification for *p* is not conclusive, or that the justification for the evidence that *p* is not conclusive. According to

fallibilism, Leite says, "it is appropriate to attribute knowledge to someone just in case you take that person to have decisive, specific evidence against those possibilities of error which you take to have some reason in their favor" (2004, 247).

Leite here only characterizes what is required to appropriately attribute knowledge. More is needed. For the fact that an epistemically responsible ascriber has no reason to think that a possibility of error obtains does not guarantee that no such possibility obtains. Nor is there any guarantee that two epistemically responsible ascribers, surveying the same situation, would agree about what possibilities of error have some reason in their favor. So it is not straightforward to infer from the appropriateness of ascribing knowledge to a subject to the fact that the subject actually knows (see Stroud, 1984).

Still, Leite's characterization can be plausibly elaborated. We might say that an epistemic agent has knowledge that *p* just in case she in fact has decisive, specific evidence against the possibilities of error that have some reason in their favor. To know, then, does not require eliminating *all* possibilities of error; it only requires eliminating those possibilities that there is some reason to believe obtain. Possibilities that there is no reason to believe, although real, can simply be ignored.

The difficulty is that sometimes such possibilities are realized. Even if there is just one chance in a million that *p* is false, when that chance is realized, *p* is false. So the agent's claim to knowledge is defeated despite the fact that she had just as good evidence for her claim, had decisive, specific evidence for the possibilities of error that she takes to have some reason in their favor, and used just as reliable methods as her epistemically luckier peers. Rather than conclude that the existence of such situations casts doubt on the adequacy of their positions, many epistemologists simply consign them to the realm of epistemic misfortune. If they don't happen too often, such epistemologists maintain, there's no problem. Knowledge involves justification and truth. Where we have adequate justification and our belief is true, we know. Where we have adequate justification and our belief is false, we do not. It is regrettable that adequate justification does not ensure truth. But we can live with an element of luck in our epistemology. In the absence of bad luck, we still know (see Pritchard, 2005).

Do the dogmatist and the fallibilist differ in anything other than self-presentation? Both set the level of justification required for knowledge as

less than a probability of 1. Both are vulnerable to epistemic misfortune, but seem not to be daunted by that vulnerability. Still, there are differences. Fallibilism, unlike dogmatism, is prey to a variant of Moore's paradox (Moore, 1991); dogmatism, unlike fallibilism, is prey to Kripke's dogmatism paradox (Kripke, 2011a).

Moore's paradox is an assertion of the form

p but I do not believe that *p*

or

I believe that *p*, but ~*p*.

There is a good deal of controversy concerning exactly what kind of infelicity a Moore's paradoxical assertion involves (see Green & Williams, 2007). It is, however, widely acknowledged that for Moore's paradox to occur, the sentence must be asserted, not merely entertained or supposed; and the attitude the agent self-ascribes or disavows is belief.

The schema that concerns us,

I know that *p*, but it is possible that ~*p*,

is a variant on 'I believe that *p* but ~*p*'. Let us call this the knowledge variant. The first conjunct of the knowledge variant is stronger than its counterpart in the canonical Moore's paradox, since in asserting it, its utterer ascribes to himself truth and justification, as well as belief. Still, belief is a central component of what he self-ascribes. The second conjunct—'it is possible that ~*p*'—is weaker than its counterpart in the canonical form of the paradox. It does not contend that *p* is false, only that it might be. But like the canonical version of Moore's paradox, the knowledge variant gives epistemic assurance in one breath and takes it back in the next.

Canonical Moore's paradoxes are first personal. There is something deeply untoward about 'I believe that *p*, but ~*p*'. There is nothing untoward about 'Meg believes that *p*, but ~*p*'. The knowledge variant is different. Because knowledge is factive, 'Meg knows that *p*, but it is possible that ~*p*' is as problematic as its first-personal counterpart. Such statements are concessive knowledge attributions (see Dougherty & Rysiew, 2009). The first

conjunct attributes knowledge; the second concedes the possibility of error. The divergence from canonical cases may be reason to deny that concessive knowledge attributions are strictly instances of Moore's paradox, but they are at least close cousins. Let us call Meg's epistemic predicament quasi-Moore's paradoxical.

To appreciate the difficulty, consider what an epistemic agent should do in a quasi-Moore's paradoxical situation. If Meg knows that p, she can responsibly assert that p, give her full assurance to others that p, and use p as a premise in any inference or as a basis for any action where it is relevant, regardless of the stakes. If she knows that p, she has a right to be sure that p. But if there is some chance that she is wrong in thinking that p, her situation is different. If the chance is small enough, perhaps she need not suspend judgment, but minimally she should hedge her bets, refrain from relying on p in inferences or actions when the stakes are very high, and convey her degree of uncertainty about p to her interlocutors. The problem is that she cannot do both. There seems to be no coherent course of thought or action open to the subject in the knowledge variant of Moore's paradoxical situations. If this is what fallibilism commits us to, we are in a bind.

It might seem, then, that we should revert to dogmatism. Maybe we should set a satisfiable threshold on the justification required for knowledge, and declare that when that threshold is reached, our evidence is conclusive. We should simply deny that there is any real (epistemic) possibility that we still might be wrong. But dogmatism faces a (putative) paradox of its own (Kripke, 2011a).

If S knows that p, then p is true.
If p is true, then any evidence against p is misleading.
So to avoid being misled and to preserve her knowledge, S should disregard any evidence against p.
Therefore, S should be dogmatic about p.

By itself, this is not strictly a paradox. It is merely an unwelcome conclusion. It becomes a paradox when we connect knowledge with rationality. Since rationality requires responsiveness to evidence, on the recommended approach, knowledge and rationality diverge. Dogmatism requires closing our minds to new evidence, lest knowledge will be lost. Rationality requires keeping an open mind.

Neither fallibilism nor dogmatism with respect to knowledge seems wholly satisfactory. Dogmatism is intellectually arrogant. Fallibilism's alleged

intellectual humility is hard to get a handle on. The admission that, despite the evidence, an epistemic agent still could be wrong, seems correct. The problem is to see how to construe it as anything other than either an invitation to skeptical paralysis or an expression of mock modesty. Here is where the move from a focus on knowledge to a focus on understanding pays dividends.

Intellectual Humility

Despite the 'ism', I suggest, fallibilism is not primarily a doctrine, but a stance—a stance of intellectual humility. The stance is not a passive concession in the face of epistemic frailty. Rather, it is an active orientation toward a domain of inquiry and our prospects of understanding it. Unlike the dogmatist, the fallibilist does not simply dismiss wholly justified or reliably generated false beliefs as bad luck. To do so is to assume that there is no insight to be gleaned from the failures. That is shortsighted.

Suppose a medication for multiple sclerosis is considerably more effective than any available alternative, but it has serious side effects in 2 percent of the cases. A physician would reasonably and responsibly prescribe the medication. She might even say that she knows that it is safe and effective. But she would not stop there. She would also warn her patients about the potential side effects, how to recognize them, and what to do if they occur. That would improve their understanding of the treatment. This is so not only for the patients who actually experience the side effects. All the patients understand their medical situation better if they are aware that the effective medication exposes them to risks and learn the ways to recognize deleterious side effects for what they are. Moreover, if the side effects are severe, researchers would presumably attempt to figure out their cause, to identify the subset of MS patients whose members are vulnerable to them, and to see whether they can find a way to eliminate them or moderate their severity. They too understand the medication better if they recognize the risk. No one would simply say 'Bad luck!' or 'Better luck next time!' to members of the 2 percent and move on. Such a response would not only be heartless, it would sacrifice useful information. The recognition that they do not know why or to whom the side effects will occur is epistemically valuable.

It might seem that the entire payoff for exercising intellectual humility is that by doing so we either uncover previously undetected errors, close previously open gaps in our understanding, or provide additional assurance that our conclusions were correct. This is surely one sort of payoff. But, I will argue, there is another important epistemic benefit. In taking the possibility of error seriously, we treat it as itself worthy of attention. We put ourselves in a position to identify the potential fault lines in our currently accepted account. If current understanding of the phenomena is wrong, where is the error apt to be located? What are the weakest links in the argument? If there is an error, what would show it? If there is no error, what accounts for the gap? By addressing such issues, we learn more about the phenomena, our assumptions about the phenomena, and our methods for investigating such things. One might think that this is simply changing the subject. Rather than asking about the epistemic status of *p*, we are now asking something else. This is incorrect. To ascertain *p*'s epistemic status, we need to determine how it figures in a network of epistemic commitments. That depends on how, where, and why it is vulnerable. Before looking at the epistemic benefits of the possibility of error, let us consider the epistemic benefits of actual error.

Error as a Mark of Success

According to Davidson (2001), the identity of a belief derives from its place in a rich, textured array of relevant, true beliefs. A person who lacks the requisite background cannot have a particular belief. So, for example, unless Paul understands a good deal of physics, he cannot rightly or wrongly believe that electrons lack mass. He would not know what an electron is or why the question of its mass even arises. It might seem that he could garner the relevant knowledge via testimony. But being ignorant of physics, although Paul can hear the words 'electrons have mass' and recognize the utterance as information-conveying testimony, the utterance is to him like testimony in a language he does not speak. He cannot draw nontrivial inferences, incorporate it into his picture of the world, or use it as a basis for action. At best, he can parrot the statement and perhaps draw trivial logical inferences from it.

Ordinarily, few if any specific beliefs are required to back a given contention. One can approach an issue from different directions. So apart from

very general beliefs, such as 'an electron is some sort of subatomic particle', no specific beliefs are required to equip the agent to form beliefs about a topic. Moreover, there are bound to be undecidable cases. It is not clear how sparse a complement of relevant beliefs a person can have and still, rightly or wrongly, believe that *p*. Indeed, the answer might depend on what sort of belief is in question. If little is known or reasonably believed about the topic, a relatively sparse constellation of relevant beliefs might suffice. If plenty is known, then the agent's constellation of beliefs might have to be considerably denser. Nevertheless, a fairly wide and dense cluster of relevant beliefs is, Davidson maintains, required to anchor each particular belief.

Davidson's point turns on the fact that most of a person's beliefs about a topic are trivial. He recognizes that there can be profound disagreements about such matters as whether electrons have sharply demarcated boundaries, with no resolution in sight. The assurance that the parties to the dispute are all talking about the same thing (and that that thing is the electron) is based on their believing a lot of the same electron-relevant trivia. They believe, for example, that electrons are subatomic particles, that they are small, that somehow they are constituents of material objects, that they are more plentiful than artichokes, that they are not cats, and so forth. The number of a person's trivial true beliefs about a topic swamps the number of her controversial beliefs about it.

Davidson insists that most of the beliefs that constitute the background must be true. I disagree. Rather, sufficiently many of the agent's relevant beliefs must be at least true enough and reasonably accepted. If we insist that for Fred to have beliefs about quarks, most of his beliefs about quantum mechanics must be true, we ask too much. Perhaps most (or, at least, enough) of what is currently accepted in quantum mechanics is true enough. But it would be unsurprising if many of the contentions of current theory turn out to be not exactly true.

I suggest that truth per se is not required. One reason is that at the cutting edge of inquiry, and often well inside the cutting edge, an epistemic agent's opinions may be at best true enough—true in the ways that matter and to the extent that they matter for their effective functioning in her cluster of epistemic commitments. Another reason is that truth is the wrong currency for assessing the opinions of babies and nonhuman animals, as well as for assessing adult opinions whose modes of representation are not truth apt.

Still, the idea that the identity conditions of beliefs and opinions depend on their place in a cluster of reasonably accurate cognitive commitments seems right. If Paul is ignorant of physics, he has no idea what an electron is or what it takes for a particle to have or lack mass. Too great an ignorance of a topic leaves a person out of touch with its subject matter. He can have no views about it. If Paul cannot tell an electron from a quark, then nothing in his cognitive system equips him to have views about the one but not the other.

It follows that to be in a position to make a mistake marks a significant epistemic achievement. Only someone who has some understanding of a topic has the resources to have mistaken beliefs about it. Only someone who understands a good deal about it has the resources to make a significant mistake. Only because Fred is cognizant about quantum mechanics can he harbor the erroneous opinion that quarks and antiquarks exchange charms. (They exchange muons.) Being in a position to have erroneous beliefs about a topic requires a significant measure of understanding of that topic.

Learning from Our Mistakes

From early childhood, we're advised to learn from our mistakes. Perhaps this means that we should learn not to make a particular mistake again. Once you've discovered that you get burned when you touch a hot stove, you should permanently refrain from touching hot stoves. Karl Popper (2002) considers this the key to science. Science, he urges, can never prove anything. It can only disprove. Thus, he concludes, scientific progress consists in formulating theories and attempting to disprove them. Each time we generate a disproof, we make progress. We know that that particular theory is false. If we do not make that mistake again, we get closer to the truth. What we learn from our mistake, then, is that things are not the way the mistaken theory says they are.

Where the mistake concerns an isolated fact, with few and weak connections to the agent's other epistemic commitments, that the proposition is false may be all that is available to learn. Joe's mistaken opinion that King Philip I of Spain was King Philip II of Portugal is a mistake of this kind. (Philip II of Spain was Philip I of Portugal.) Similarly when mistakes are wild guesses. When a student discovers that he is mistaken in thinking that the value of Avogadro's number is 17, he is not much closer to knowing the

actual value of Avogadro's number. Still, learning from mistakes through a process of elimination is exceedingly labor-intensive, and unlikely to get very far. There are indefinitely many ways the world might be. So the idea that we will get to the truth by sequentially eliminating hypotheses that have shown themselves to be mistaken is not promising. Neither individually nor as a species will we survive long enough to eliminate all the false alternatives.

This characterization of the issue treats every error as a stab in the dark. When we understand little about a topic, we are not in a position to learn much from a mistake. But our situation is rarely so bleak as this suggests. Usually, we understand enough that the range of plausible alternatives is restricted. If Maria realizes there are only three configurations in which a particular protein might fold, eliminating one of them yields considerable information about the protein—it folds in one of the remaining two ways. If Bill is aware that only five candidates are running for office, the news that one has withdrawn again yields considerably more information than if he thought that any one of the nearly 7 billion people on earth might be elected.

But even if we grant that our cognitive situation is frequently better than that of the student who is wildly guessing values for Avogadro's number, it might seem that our prospects of learning much from a mistake are bleak. The problem is a consequence of holism. Quine argues, correctly I believe, that "statements about the external world face the tribunal of experience not individually, but only as a corporate body" (1961, 41). This means, as Walden says, "We never properly get evidence for a proposition; we get evidence for a proposition relative to methodological prescriptions about what counts as evidence for what. . . . These methodological prescriptions are also part of our theory, and thus . . . they are actively confirmed or denied along with the rest of the theory" (2007). The entire cluster of factual and methodological commitments is confirmed or disconfirmed together. It follows that what we learn when we discover that we have made a mistake is that something in a wide constellation of commitments is wrong. But the discovery of the error alone does not target any particular member of the constellation.

What this bleak assessment overlooks is that typically not all of the elements of the constellation are equally vulnerable. Since constellations of epistemic commitments overlap, some elements of the constellation that

contains a mistake may be independently supported by their roles in other constellations that have been confirmed. Then we have prima facie reason to believe that they are not the locus of error. The more solidly supported our background assumptions, the more precise the focus a mistake can provide.

At the opposite extreme from random guesses are cases where we are quite confident that we know what will happen, but turn out to be wrong. In such cases, our mistaken conviction is a telling error. An experiment that to our surprise fails to confirm a seemingly well-established theory is such a case. Let me briefly describe such an experiment (Suzuki et al., 2007). A plasmid is a circular DNA molecule found in bacteria. Inserting mutations into plasmids is a common technique in genetic engineering. Bacteria containing those plasmids then transfer the mutation to other bacteria through conjugation, a process by which bacteria exchange nuclear material. That is the established, well-confirmed background against which the following experiment took place. A bacteriologist in Marcin Filutowicz's laboratory attempted to insert two different mutations into a single bacterial plasmid. Given the previous successes and the then current understanding of bacteriology, he had good reason to believe he would succeed. Introducing mutations into plasmids is a boringly routine task. But his belief was mistaken. No matter how often he tried, and no matter how carefully he proceeded, he could not produce a live bacterium containing the two mutations. This constituted a surprising experimental failure. In light of the then current understanding of bacteriology, the procedure ought to have worked.

A Popperian might conclude that what we learn from this mistake is that one cannot introduce the two mutations into the same plasmid and obtain a live bacterium. A Quinean might conclude that there is a mistake somewhere in the cluster of assumptions that led bacteriologists to believe that two mutations could be introduced at once. Both would be right. We do learn these things. But, I suggest, we learn something more.

Introducing either mutation alone did not produce the untoward result. The background assumptions and methodology were the same whether one or two mutations were introduced. So rather than simply concluding that something must be wrong somewhere in the vast constellation of substantive and methodological assumptions, Filutowicz was in a position to zero in on the error—to recognize that it arose from thinking that if you can

successfully introduce each of the mutations, you can successfully introduce both. That is, given the depth and breadth of the well-established background commitments, he could use the mistake to probe the current understanding of plasmids.

The mistake put him in a position to consider more carefully what occurs within a plasmid when mutations are introduced. Having evidence that the two mutations together did something that neither did alone, he hypothesized that when the two mutations were introduced together, they caused the plasmid to overreplicate and destroy the containing bacterium. He confirmed the hypothesis and went on to devise a bacterium that suppressed the overreplication, thereby creating a 'Trojan Horse'. The new bacterium contains the normally overreplicating plasmid. In conjugation, it passes on the propensity to overreplicate to bacteria that cannot survive the process. This yields an antibacterial agent that evidently does not trigger antibiotic resistance.

Because so much was understood about bacteriology and about the process of introducing mutations into plasmids, once the error was discovered, the mistaken assumption was extremely informative. Filutowicz and his associates did not just learn that you cannot introduce the two mutations at once, or that there is something wrong somewhere in the cluster of assumptions that led them to think that they could do so. Because the mistake focused attention on the replicating behavior of plasmids, it opened the way to new and fruitful insights. If it turns out that pathogenic bacteria are incapable of or less prone to developing resistance to the 'Trojan Horse', then the fruits of the scientist's mistaken belief will be far more valuable than the insights that would have been gained had his original expectation about the experimental outcome been true.

One other case worth mentioning is the Michelson—Morley experiment. At the end of the nineteenth century, physicists recognized that light consists of waves. They assumed that waves require a medium of propagation. Since light waves travel through space, they posited that space was filled with a medium, the luminiferous ether. The objective of the experiment was to measure the flow of ether across the Earth as the Earth travels around the sun. Although the experiment was designed and conducted with exquisite care, the result was null. No ether drift was detected. Over a period of years, the experiment was redesigned and ever more sensitive inferometers

were used, to no avail. The belief that light consists of waves in an ethereal medium is mistaken.

This familiar story may seem to have the same moral as the previous one. But the payoff is different. The cognitive consequences of the mistaken belief that Filutowicz and his colleagues shared were local and limited. Because the tightly woven web of commitments they relied on consisted in large part of independently confirmed strands, they could reasonably quickly identify the erroneous belief and see what routes of inquiry it opened up. The rest of the commitments largely held firm when the mistake was corrected. The Michelson—Morley experiment was different. There too, scientists had a tightly woven fabric of established cognitive commitments that led them to believe that the experiment would succeed. Their failure eventually convinced them that luminiferous ether does not exist. But this did not, and could not, lead to a local and limited revision of their understanding. It tore the fabric of cognitive commitments apart. For if light waves do not require a medium of propagation, light is radically different from what science supposed. And if that is so, then many of their other assumptions about matter and energy had to be revised as well. I do not claim that the Michelson—Morley experiment was a 'crucial experiment' that by itself falsified Newtonian physics. Rather, its effect was Socratic. It made manifest to the scientific community the extent to which they did not understand what they thought they understood.

If an opinion is supported by a tightly woven tapestry of reasons, and the opinion turns out to be erroneous, more than that particular opinion must be revised or rejected. The question arises: How could it be wrong? What have we been missing, or overlooking, or underestimating, or misconstruing? The realization that this is *not* the way things are in a particular area can often afford avenues of insight into the way things are. At the very least, it enables us to focus attention on particular aspects of our system of commitments. It not only motivates us to ask questions we would not otherwise ask, it often also provides the resources for answering them.

The human propensity for error is typically regarded as a regrettable weakness. Certainly the propensity to make careless mistakes is a weakness. So is the propensity to jump rashly to erroneous conclusions. But, I have suggested, the propensity to make thoughtful, educated mistakes may be a strength. This is not to say that every error is felicitous. Some, like wild

guesses, provide virtually no information beyond the fact that things are not the way we supposed. Some, although well grounded, are simply off the mark. Some are cognitively (and often morally) culpable. They fail to take account of available information and therefore lead us in the wrong direction. But sometimes, we get things wrong in epistemically fruitful ways. Once discovered, such errors provide not just incentives but resources for serious, focused, effective inquiry. By revealing not only that but also where we have got things wrong, they point us in the direction of advancing our understanding. We are lucky we are disposed to make such mistakes.

The Possibility of Error

We have seen that the capacity to be mistaken about something is an epistemic achievement and that actual errors can embody and engender insights that investigators would otherwise miss. But the issue for fallibilism is whether there is any epistemic value to recognizing the *possibility* of error. Let us grant that little is to be gleaned from the unfocused possibility of error that global skepticism presents. That Fred would be wrong in a demon world shows only that the content of his belief is not an obviously necessary truth. He probably recognizes that anyway. But the capacity to identify local fault lines—to see where, even if skepticism is false, errors might creep in—is a valuable epistemic asset.

Suppose that rather than simply taking a test result at face value, investigators seriously entertain the possibility that it is a false positive. They may already know the probability of its being a false positive. Let's say that it is 3 percent. The issue before them is whether the case before them is one of the 3 percent. How might they tell? In focusing on this question, they put themselves in a position to learn more about what the test detects and how exactly it does so, what range of conditions it is sensitive to, what the probabilities of the various conditions are, whether there are circumstances in which the test is particularly likely to yield false positives or subjects for whom it is particularly likely to give misleading results. Answering such questions does not always require amassing additional evidence. Sometimes the answers can be found simply by reexamining the data at hand. By, for example, using sophisticated regression techniques to focus on what the data discloses about the 3 percent, they may be able to extract additional

information. The idea that the only thing the evidence for p discloses is whether p is the case is shortsighted. It blinds epistemic agents to additional information embedded in the evidence.

It is fashionable to interpret modal claims as truths about other possible worlds. So to contend that a hypothesis could be false is to say that there exists a possible world where it is false. For our purposes, this interpretation is unhelpful. Rather, we should recognize that the possibility of error is a property the hypothesis has in the actual world. That property—a vulnerability—derives from the nature and strength of the support the hypothesis gets from the account and the evidence for it. What it portends depends on the weave of the fabric of commitments that constitute the account and on the contribution of the hypothesis to the strength of the fabric. If an erroneous contention is tightly woven into a fabric of epistemic commitments, an epistemic agent could be wrong about it only by being wrong about many other things. The contention that species evolved is supported not just by biological evidence (from genetics, anatomy, physiology, ecology, etc.) but also by the findings of other sciences. If it were false, geology would have to be badly wrong about how fossils are formed and embedded in sedimentary rock, chemistry would have to be badly wrong about organic compounds, and physics would have to be badly wrong about carbon dating. Not easily could science be so wrong as to discredit the claim that species evolved. If, on the other hand, a contention is a mere tassel, loosely dangling from a fabric of understanding, it could easily be wrong. Urban legends are a case in point. It is said that alligators, having been brought from Florida as babies and flushed down the toilets when they grew too big, now roam the New York City sewers. We would not have to be wrong by much to be wrong about that. Our only reason for believing it is that we had it from a putatively reliable source. On learning that it is wrong, we readily give it up and perhaps downgrade our assessment of the reliability of our source.

The value of recognizing where we might be wrong is not only, or even mainly, prudential. It is not just that an agent attuned to the possibilities of error can buttress her defenses. For possibilities of error disclose aspects of the topography of the space of reasons. We appreciate how the network of commitments hangs together when we recognize how it might fall apart, how easily it might fall apart, and what the consequences of its doing so would be.

Appreciating the possibility of error is not just ruefully admitting, 'Well, despite my evidence, I could still be wrong'. It involves recognizing the ramifications of potential errors—the damage they would do to a network of epistemic commitments. That recognition provides incentives to double check results, shore up findings, hedge bets. It also provides incentives to develop methods for doing such things. A first stab at double checking is to rerun the same test, redo the same calculation, reexamine at the same evidence. This yields a measure of assurance, for it functions as a stay against carelessness. But to run the same test twice does not protect against underlying misconceptions. So investigators do well to address their conclusion from different perspectives: do different tests, perform different calculations, run different regressions, scrutinize the data from different perspectives. Granted, a result that withstands such reconsideration might still be incorrect. But if investigators demonstrate that a finding stands up to diverse forms of testing, they ensure that they would have to be wrong about a lot to be wrong about it. Because they have woven it into their fabric of commitments, such a result is stronger than the result they would obtain by simply amassing more evidence using the original method.

If we recognize that intellectual humility is an epistemic asset, don't we still face a variant of Moore's paradox? The understanding variant is:

My understanding of φ embeds my accepting that p, but it is possible that ~p.

Is this a problem? Let us temporarily set aside the worry that in uttering the understanding variant I give my assurance with one breath and take it back with the next. Still there is the question of what I should do. We saw that the subject in the quasi-Moorean predicament brought on by the knowledge variant was in a bind. There seemed to be no coherent course of reasoning or action available to her. The subject in the situation described by the understanding variant has no such problem. Since she accepts that p, she can use p in inference and action when her ends are cognitive. But because her understanding of φ embeds an appreciation of her epistemic vulnerability in accepting that p, it also gives her reason to double check, to hedge her bets, to be sensitive to intimations of error, and to convey her level of epistemic insecurity to her interlocutors. To say that she can use p for inference and action does not entail that she can or should do so blindly.

A consequence of fallibilism is that concessive knowledge attributions are often true. When I know that p, there remains some possibility that ~p. When my understanding of φ embeds a commitment to p, there remains

some possibility that ~*p*. If we disregard the possibility of error, we lose valuable information. So why are concessive knowledge attributions untoward? I suggest that the reason is Gricean. The second maxim of quantity is "Do not make your contribution more informative than is required" (Grice, 1989, 26; see also Dougherty & Rysiew, 2009). If everyone knows, and everyone knows that everyone knows that empirical claims are fallible, then the fallibility of an empirical claim goes without saying. To mention that I might be wrong is to highlight the possibility of error—to make it salient. I ought only *say* that I might be wrong in circumstances where I consider *p*'s epistemic standing especially vulnerable. But in such a case, I ought not claim to know that *p*. What goes without saying ought not be said.

I have argued that fallibilism is not a higher-order stance toward our understanding of a topic. It is, or should be, woven into the fabric of understanding. By recognizing the possibility of error, we can exploit our epistemic vulnerability and gain insight into our understanding of a topic. Rather than being a weakness, our vulnerability to error is a strength.

Notes

2 Distancing from Truth

1. Although I harbor Quinean qualms about propositions, I set them aside here to characterize mainstream epistemology in its own terms.

2. An 'account' is what in *Considered Judgment* I called a 'system of thought'. I use the bland term 'account' to avoid getting involved in disputes about the relation between theories, evidence, and models.

3. I do not have strong intuitions about this case, but I do not think it is clearly wrong to say that the cat has such a belief.

4. As we will see, the issue is not so straightforward as Williams thinks.

5. I will argue later that some falsehoods that purport to be true are nonetheless successful because they are felicitous falsehoods. For now, however, the focus is on sentences that manifestly play a non-truth-telling role.

6. I call them falsehoods out of respect for classical bivalence. One might equally argue that they are neither true nor false. All that I require is that that they are epistemically felicitous although not true.

7. Whether so-called ceteris paribus laws are really laws is a subject of controversy. See, for example, *Erkenntnis* 57 (2002), for a range of papers on the issue. Although for brevity I speak of them as laws, for my purposes, nothing hangs on whether the generalizations in questions really are laws, at least insofar as this is an ontological question. I am interested in the role such generalizations play in ongoing science. Whether or not they can (in some sense of 'can') be replaced by generalizations where all the caveats and restrictions are spelled out, in practice scientists typically make no effort to do so. Nor, often, do they know (or care) how to do so.

8. i and r are the angles made by the incident beam to the normal and n_1 and n_2 are the refractive indices of the two media.

9. "The profit rate is the level of profits in the economy relative to the value of capital stock" (Bannock, Baxter, & Davis, 1998, 397).

10. The veritist might attempt to soften the blow by arguing that we should be expressivists about science. But it is implausible in the extreme to maintain that natural science is just an expression of our attitudes rather than a discipline whose success depends on the way things actually are.

11. Conceivably, of course, the equations in question will be superseded by some other understanding of the subject, but the fact that the equation we consider true does not have an analytical solution provides no reason to think so. Nor does it provide reason to think that the considerations that supersede it will be mathematically more tractable, much less that in the long run science will be free of all such irksome equations.

3   From Knowledge to Understanding

1. The parenthetical caveats are included to allow for the possibility that some knowledge—e.g., perceptual knowledge—is direct knowledge and not in need of justification. I do not believe that this is so, but at this point in the argument it is advisable to leave this possibility open.

2. This maneuver is modeled on the 'corrected doxastic system' in Lehrer's (1974) epistemology.

3. Although it is undeniably a cognitive accomplishment, I also bracket understanding a language and related feats like understanding a concept or a passage of prose. Understanding a language is a matter of being capable of interpreting and, perhaps, speaking, reading, and/or writing it correctly. It is integral to understanding in general. But on the face of it, it is not the same as understanding a topic, so for now I set it aside.

4. 'Hardly more secure' rather than 'no more secure', because a measure of additional support is supplied by the fact that military strategists agree that such a strategy might work.

5. The restriction to cases where her ends are cognitive is crucial. Depending on the topic and her attitude toward the topic, she may be unwilling all things considered to use the information in reasoning or in action.

6. I owe this example to Kent Bach.

4   The Fabric of Understanding

1. One might wonder whether the contents of a novel qualify as deliverances. Arguably, they do not present themselves as candidates for epistemic acceptance. If this

is right, the objection does not get off the ground. In chapter 11 I discuss the epistemic functions of fiction. Here I am willing to treat the contents of a novel as deliverances, and let the nineteenth-century novel serve as a challenge to theories that take coherence to be conducive of tenability.

2. I say 'direct claims' because I believe that novels play a significant role in the advancement of understanding. Just how they do so will emerge in chapter 11. See Elgin (1996), 183–200.

3. I thank Fred Feldman for this objection.

4. Which I think is probably an unintelligible claim.

5 Epistemic Normativity

1. This chapter was made possible through the support of a grant from the Intellectual Humility Project at St. Louis University, the Varieties of Understanding Project at Fordham University, and the John Templeton Foundation. The opinions expressed here are those of the author and do not necessarily reflect the views of the Intellectual Humility Project, the Varieties of Understanding Project, or the John Templeton Foundation.

2. Agents are responsible for actions they perform when drunk or under the influence of recreational drugs. As Aristotle (1985) argues, even if they are out of control when they perform the particular actions in question, because they were in control when they imbibed to the point where they lost control, they are responsible for their drunken behavior. And in general, if x ought to φ, then x is ceteris paribus obliged to refrain from putting herself in a position where she cannot φ. The bearing of this point on my argument is this: if someone irresponsibly neglects to acquire or exercise an epistemic virtue, he can be held responsible for the epistemic defects that result.

3. Such heteronomy is character of what Sosa (2007) calls animal knowledge. It is plausible that animals cannot reflectively endorse or criticize their attitudes, and are subjectively vulnerable in just the way I suggest. On the other hand, they presumably do not go in for, and do not need to go in for, self-reflection. So the vulnerability is not a subjective problem for them.

4. I think that epistemic agents can hone their perceptual, representational, and imaginative capacities. So they are not entirely passive even with respect to inputs. But I will not argue that here.

5. I am grateful to an anonymous reviewer for this example.

6. I am grateful to Michael Ashooh for this example.

7. I am grateful to Jonathan Adler for this point.

8. I am grateful to Alvin Goldman and Sanford Goldberg for pressing me to clarify this point.

9. In the cases Kitcher is concerned with, rival theories might be approximately equal in predictive success, and/or each might best the other across some part of the domain. Then their predictive track records compare favorably with one another. Moreover, since scientific acceptability involves multiple desiderata, there are trade-offs. A measure of predictive success might be sacrificed for an increase in precision or scope. In such cases, the practices as wholes compare favorably with one another.

10. I am grateful to Thomas Kelly for raising this issue.

6   Intellectual Integrity

1. This chapter was made possible in part through the support of a grant from the Spencer Foundation and in part through a grant from the Intellectual Humility Project of St. Louis University and the John Templeton Foundation. The opinions expressed here are those of the author and do not necessarily reflect the views of the Intellectual Humility Project, the John Templeton Foundation, or the Spencer Foundation.

2. As noted, the upper-case letters indicate that as Williams uses them 'Sincerity' and 'Competence' are quasi-technical terms that diverge slightly from their standard meanings. I follow his lead. The main deviation of 'Well-Intentioned' as I use the term is that it is role specific. An agent can be a Well-Intentioned auto mechanic or informant, while being Ill-Intentioned when it comes to paying all the taxes he owes.

3. Bernard Williams distinguishes between the ethical and the moral. Ethics is a broader category concerned with the question how should we live. Morality is concerned specifically with issues pertaining to rights, duties, and obligations. If we follow Williams, what the act calls 'the ethical conduct of research' is more properly called the 'moral conduct of research'. For my purposes, it does not matter which term we use.

4. Kuhn (1970) denies that science is cumulative. In scientific revolutions, much that has been accepted is called into question. This does not undermine my contention that scientists build on one another's findings. For even a revolutionary scientist draws on evidence, methods, and instruments that others contrived.

5. Although I doubt that this is possible, my point here is weaker. Science as we know it could not be so generated.

6. There are gray areas here. It may not be obvious whether a failed investigation rises to the level of significance to make it worth publishing. As a result, journals may be more reluctant to publish failed studies than successful ones. (I thank

Jonathan Adler for this point.) But there are also cases where scientists and the institutions they work for suppress failed studies because they have a vested interest in a research program that is undermined by the failure, or because they consider the failure a fluke, or both.

7. It may of course do more. When a pharmaceutical company suppresses results that show a drug to be dangerous or ineffective, it illicitly conveys assurance that the drug is safe and effective or at least not known to be unsafe or ineffective. In emphasizing the epistemological consequences for the scientific community, I do not mean to deny that there are other serious consequences, not only for the scientific community but also for the public at large.

8. This is not just an 'in principle' worry. One study showed that pharmaceutical-industry-funded trials were far more likely to report positive outcomes (85.4%) than either government-funded trials (50%) or nonprofit-funded trials (71%). This does not demonstrate that the industry-funded trials were biased. Possibly pharmaceutical companies do not run clinical trials unless they are pretty sure that the drugs being tested are effective. But in light of the industry's financial interest in the outcome, these statistics give one pause. See Bourgeois, Murthy, and Mandl (2010).

9. It might seem that at least the methodology section of an acceptable research paper needs to be true. Even if scientists cannot guarantee the truth of their results, they can, and therefore should, provide a true description of how the result was arrived at. But a true history of the path to the conclusion is not what the methodology section is supposed to provide. The history of investigation leading to a particular result is typically a circuitous path, with numerous false starts, missteps, and dead ends. How they actually arrived at the result—the precise meandering path they took—is of no interest. The question the methodology section is supposed to answer is not 'How did we come by our result?' but 'Why should you accept our result?' Much of the actual history of the investigation is irrelevant to the answer to the latter question. So scientists provide a sanitized, streamlined description of a method that is grounded in and justified by the investigation they carried out and that, if followed, would yield their result. Rather than a true description of what they did, it is a rational reconstruction. To vindicate the open invitation requires supplying assurance not that the conclusion is true but that, in light of current standards, the evidence warrants it. The methodology section is written to show how the evidence conveys warrant.

10. Competence is field based. The idea is that anyone with the appropriate knowledge and training in a specific branch of science could have done the work. It is not that a competent geneticist could have performed or interpreted an experiment in plasma physics, or even that the bench scientist could do the statistical analysis provided by another member of his team.

11. I thank Jonathan Adler for this point.

12. For example, Jan Schön's misconduct was discovered when it was noticed that supposedly distinct experiments were reported to all have identical background noise. (See Goodstein, 2010, 98–100.)

7   Objectivity: A Bulwark against Bias

1. This chapter was made possible through the support of a grant from the Varieties of Understanding Project at Fordham University and the John Templeton Foundation. The opinions expressed here are those of the author and do not necessarily reflect the views of the Varieties of Understanding Project or the John Templeton Foundation.

2. I say 'purportedly' and 'putatively' because some of the pictures have religious subjects. I am not remotely qualified to say whether heaven, if it exists, is three dimensional, or whether angels, if they exist, are visible. But it is plain that the artists represented heaven as a three-dimensional physical space, and angels as visible in such a space.

3. Initially, I suppose, the several judges negotiated until their assessments converged.

8   Irreconcilable Differences

1. There are exceptions. If a speaker knows that her interlocutor holds a position that she does not, she can responsibly say, 'Given your commitments, *p* is a reason for you. Since I do not share those commitments, it is not a reason I share. But you should take it seriously'.

2. Here I use the term 'paradigmatically factual' for cases that are uncontroversially factual because I want, at least for now, to leave open the question of whether aesthetic judgments are or are not judgments of fact.

3. My focus on well-known critics may suggest that I am committed to something like Dickie's (2007) contention that the art world determines what counts as art and what counts as good art. This is not so. I am concerned with the ways that and reasons why people who engage seriously with works of art provide reasons for their claims. I appeal to a dispute between well-known critics because they are sufficiently focused and articulate that the points they raise readily strike us as reasons. The same considerations arise in discussions on blogs devoted to heavy metal music or *Game of Thrones*.

4. Caveat: although Greenberg is clearly responding to Fry's discussion of *Le Compotier*, he casts his discussion as about Cézanne in general. He does not specifically mention *Le Compotier*. Both because of the context of his discussion and because

of the utility of the assumption for my purposes, I take him, like Fry, to be discussing *Le Compotier.*

5. There is nothing about the gustatory realm that excludes it from aesthetic judgment. There are food critics, just as there are critics of painting, music, and dance. Food critics, like other art critics, give reasons for their judgments. They point to factors, like the slightly metallic aftertaste or a soupçon of lemon, which others may not have noticed, to back up their claims. In this they differ from folks like Jim and Kim who simply do not find the same food tasty.

9   Exemplification in Understanding

1. This chapter was made possible through the support of a grant from the Varieties of Understanding Project at Fordham University and the John Templeton Foundation. The opinions expressed here are those of the author and do not necessarily reflect the views of the Varieties of Understanding Project or the John Templeton Foundation.

2. See McGowan (2003) on metaphysical privileging.

3. I am grateful to Sean Parker for this example.

4. This is not to say that any account that exemplifies some deleterious feature of lying is an adequate understanding. If important features are omitted or underrated, or if the relations among the features are confused, the understanding may be skewed.

5. Here I draw heavily on Jullien (2006, 158–164), both for the example and the discussion of the role of exemplification in mathematics. She takes the example from G. H. Hardy.

11   Casting in Bold Relief

1. I am grateful to Georg Brun and Christoph Baumberger for raising this objection.

2. This is consonant with Platonism, but does not require it. Perhaps abstract properties and patterns exist only if instantiated, but instantiations, whether material or virtual, can be created or can naturally emerge.

3. I originally found reference to Boswell's column in Snyder (2012), 55.

4. Carroll discusses but does not endorse this objection.

5. Carroll discusses but does not accept this argument.

6. I am grateful to Jonathan Adler, Geordie McComb, Amélie Rorty, and Michael Stuart for helpful comments on earlier drafts of this chapter.

12   Chekhov's Gun

1. A variant of this quote is regularly misattributed to Virginia Woolf. I am grateful to John Kulvicki for giving me the correct wording and the correct reference.

2. This use of 'denote' is slightly tendentious, both because denotation is usually restricted to language and because even within language it is usually distinguished from predication. As I use the term, predicates and generic nonverbal representations denote the members of their extensions. See Elgin (1988), 19–35.

3. I am grateful to Roman Frigg for this example.

13   Borders and Crossings

1. Both Grimm (2008) and Lipton (2009) require that the agent who understands is not acting by rote. Her capacity to perform correctly must be stable across a range of counterfactual circumstances.

2. I am grateful to Kenneth Walden for this point.

3. I am indebted to Ajume Wingo for this interpretation of the Mall.

4. This is not to say that all works of art that afford historical understanding do so only or primarily for the inexpert. Some works—for example, those with religious themes—convey understanding only to those with considerable expertise.

14   Fallibility's Payoff

1. This chapter was made possible through the support of a grant from the Intellectual Humility Project of St. Louis University and the John Templeton Foundation. The opinions expressed here are those of the author and do not necessarily reflect the views of the Intellectual Humility Project or the John Templeton Foundation.

2. Adler's position evolved. So the skepticism he evinces in Adler (1981) is not present in his later theory.

3. Pritchard disagrees. But to get the result that the probability of winning a fair lottery with long odds is not remote, he relies on a very strong safety principle. "*Safety III*: For all agents . . . if an agent knows a contingent proposition φ, then in nearly all (if not all) nearby possible worlds in which she forms her belief about φ in the same way as she forms the belief in the actual world, that agent only believes that φ when φ is true" (2005,163). If knowledge requires *Safety III*, we have very little knowledge. Moreover, it seems at least odd to say that although winning the lottery is very improbable, it is also very possible. To make such a claim seems to unduly downplay the force of the evidence.

References

Abel, G. (2012). Knowing-how: Indispensable but inscrutable. In S. Tolksdorf (Ed.), *Conceptions of knowledge* (pp. 245–267). Berlin: De Gruyter.

Adler, J. (1981). Skepticism and universalizability. *Journal of Philosophy*, *78*, 143–156.

Adler, J. (2002). *Belief's own ethics*. Cambridge, MA: MIT Press.

Adler, J. (2009). Why fallibility has not mattered and how it could. In H. Siegel (Ed.), *Oxford handbook of the philosophy of education* (pp. 83–103). Oxford: Oxford University Press.

Alberti, L. B. (2011). *On painting*. Cambridge: Cambridge University Press.

Alston, William (1988). The deontological conception of epistemic justification. In J. Tomberlin (Ed.), *Philosophical Perspectives* (Vol. 2): *Epistemology* (pp. 257–299). Atascadero, CA: Ridgeview.

Ankeny, R. (2014). The overlooked role of cases in causal attribution in medicine. *Philosophy of Science*, *81*, 999–1011.

Aristotle. (1973). *Poetics*. In *Introduction to Aristotle*. Ed. Richard McKeon. Chicago: Chicago University Press.

Aristotle. (1985). *Nicomachean ethics*. Indianapolis: Hackett.

Austen, J. (1814). Letter to her niece, Anna Austin Lefroy, September 9, 1814. In *Letters of Jane Austen*. Bradbourne Edition, http://www.pemberley.com/janeinfo/brablt 16.html.

Ball, P. (2005). *Elegant solutions*. Cambridge: Royal Society of Chemistry.

Banes, S. (2003). Gulliver's hamburger: Defamiliarization and the ordinary in the 1960s avant garde. In S. Banes (Ed.), *Reinventing dance in the 1960s* (pp. 3–5). Madison: University of Wisconsin Press.

Bannock, G., Baxter, R., & Davis, E. (1998). *Dictionary of economics*. New York: John Wiley.

Bealer, G. (1998). Propositions. *Mind*, *107*, 1–32.

Bennett, J. (1974). The conscience of Huckleberry Finn. *Philosophy*, *69*, 123–134.

Berkeley, G. (1957). *A treatise concerning the principles of human knowledge*. Indianapolis: Bobbs Merrill.

Berker, S. (2013). The rejection of epistemic consequentialism. *Philosophical Issues*, *23*, 363–387.

Bokulich, A. (2001). Rethinking thought experiments. *Perspectives on Science*, *9*, 285–307.

BonJour, L. (1985). *The structure of empirical knowledge*. Cambridge, MA: Harvard University Press.

Boswell, T. (2011). Penn State Coach Joe Paterno reaches a sad conclusion. *Washington Post*, Nov. 9.

Bourgeois, F. T., Murthy, S., & Mandl, K. (2010). Outcome reporting among drug trials registered in ClinicalTrials.gov. *Annals of Internal Medicine*, *153*(3) (Aug. 3): 158–166.

Brewer, S. (1998). Scientific expert testimony and intellectual due process. *Yale Law Journal*, *107*, 1535–1682.

Brogaard, B. (2005). I know, therefore I understand. Unpublished MS.

Burge, T. (1993). Content preservation. *Philosophical Review*, *102*, 457–488.

Capdevila, R. (2009). *Construing architecture, constructing philosophy*. Dissertation presented to the Deparatment de Filosofia, Universitat Autònoma de Barcelona.

Carr, D. (1979). The logic of knowing how and ability. *Mind*, *88*, 394–409.

Carroll, N. (2002). The wheel of virtue: Art, literature, and moral knowledge. *Journal of Aesthetics and Art Criticism*, *60*, 3–26.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Oxford University Press.

Cartwright, N. (1999). Aristotelian natures and modern experimental method. In *The dappled world* (pp. 77–104). Cambridge: Cambridge University Press.

Cavell, S. (1987). *Disowning knowledge in six plays of Shakespeare*. Cambridge: Cambridge University Press.

Chang, H. (2004). *Inventing temperature*. Oxford: Oxford University Press.

Chisholm, R. (1973). *The problem of the criterion*. Milwaukee, WI: Marquette University Press.

Cohen, L. J. (1992). *An essay on belief and acceptance*. Oxford: Clarendon Press.

Cohen, S. (2014). Contextualism defended. In M. Steup, J. Turri, & E. Sosa (Eds.), *Contemporary debates in epistemology* (pp. 69–75). Malden, MA: Wiley Blackwell.

Comfort, N. C. (2001). *The tangled field: Barbara McClintock's search for the patterns of genetic control*. Cambridge, MA: Harvard University Press.

Cunningham, Merce. (1986). Interview. *Points in space* (CD).

Davenport, E. (1983). Literature as thought experiment (on aiding and abetting the muse). *Philosophy of the Social Sciences*, *13*, 279–306.

David, M. (2005). Truth as the primary epistemic goal: A working hypothesis. In M. Steup & E. Sosa (Eds.), *Contemporary debates in epistemology* (pp. 296–312). Malden: Blackwell.

Davidson, D. (2001). Rational animals. In *Subjective, intersubjective, objective* (pp. 95–106). Oxford: Oxford University Press.

Davies, D. (2010). Learning through fictional narratives in art and science. In R. Frigg & M. C. Hunter (Eds.), *Beyond mimesis and convention* (pp. 51–70). Dordrecht: Springer.

DeRose, K. (2011). *The case for contextualism*. Oxford: Oxford University Press.

Dewey, J. (1916). *Democracy and education*. New York: Free Press.

Dickie, G. (2007). What is art? An institutional analysis. In J. Adler & C. Elgin (Eds.), *Philosophical inquiry* (pp. 852–862). Indianapolis: Hackett.

Dougherty, T., & Rysiew, P. (2009). Fallibilism, epistemic possibility, and concessive knowledge attributions. *Philosophy and Phenomenological Research*, *78*, 123–132.

Douglas, H. (2004). The irreducible complexity of objectivity. *Synthese*, *138*, 453–473.

Doyle, Y., Egan, S., Graham, N., & Khalifa, K. (2015). Non-factive understanding: A statement and a defense. Unpublished MS.

Dretske, F. (2000). Conclusive reasons. In *Perception, knowledge, and belief*. Cambridge: Cambridge University Press.

Elgin, C. Z. (1988). *With reference to reference*. Indianapolis: Hackett.

Elgin, C. Z. (1996). *Considered judgment*. Princeton, NJ: Princeton University Press.

Elgin, C. Z. (2002). Art in the advancement of understanding. *American Philosophical Quarterly*, *39*, 1–12.

Elgin, C. Z. (2007a). The laboratory of the mind. In J. Gibson, W. Huemer, & L. Pocci (Eds.), *A sense of the world: Essays on fiction, narrative, and knowledge* (pp. 43–54). London: Routledge.

Elgin, C. Z. (2007b). Emotion and understanding. In G. Brun, U. Dogouglu, & D. Kunzle (Eds.), *Epistemology and emotions* (pp. 33–50). London: Ashgate.

Elgin, C. Z. (2011). Making manifest: Exemplification in the sciences and the arts. *Principia*, *15*, 399–413.

Federal Register. 2009. Aug. 20 (Vol. 74, No. 160), http://edocket.access.gpo.gov /2009/E9-19930.htm.

Feldman, R. (2008). Modest deontologism in epistemology. *Synthese*, *177*, 41–66.

Feynman, R. (2001). *What do you care what other people think?* New York: W. W. Norton.

Fine, A. (1998). The viewpoint of no-one in particular. *Proceedings and Addresses of the American Philosophical Association*, *72*, 9–20.

Firth, R. (1981). Epistemic merit: Intrinsic and instrumental. *Proceedings and Addresses of the American Philosophical Association*, *55*, 5–23.

Folkman, J. (1996). Fighting cancer by attacking its blood supply. *Scientific American*, *275*(3), 50–54.

Fricker, E. (1994). Against gullibility. In B. K. Matilal & A. Chakrabarti (Eds.), *Knowing from words* (pp. 125–161). Amsterdam: Kluwer.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing.* New York: Oxford University Press.

Frigg, R. (2010). Fiction and scientific representation. In R. Frigg and M. C. Hunter (Eds.), *Beyond mimesis and convention* (pp. 97–138). Dordrecht: Springer.

Fry, R. (1952). *Cézanne.* London: Hogarth.

Fumerton, R. (2010). Foundationalist theories of epistemic justification. In E. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2010 Ed.), http://plato.stanford .edu/archives/sum2010/entries/justep.foundational.

Gendler, T. (2010). Galileo and the indispensability of scientific thought experiment. In *Intuition, imagination, and scientific methodology* (pp. 21–41). Oxford: Oxford University Press.

Giere, R. (1999). *Science without laws.* Chicago: University of Chicago Press.

Ginsborg, H. (2010). Primitive normativity and skepticism about rules. *Journal of Philosophy*, *108*, 227–254.

Goldman, A. (1986). *Epistemology and cognition.* Cambridge, MA: Harvard University Press.

Goldman, A. (1999). *Knowledge in a social world.* Oxford: Clarendon.

Goodman, N. (1968). *Languages of art*. Indianapolis: Hackett.

Goodman, N. (1977). *The structure of appearance*. Dordrecht: Reidel.

Goodman, N. (1978). *Ways of worldmaking*. Indianapolis: Hackett.

Goodman, N. (1984). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Goodstein, D. (2010). *On fact and fraud*. Princeton, NJ: Princeton University Press.

Greco, J. (2014). Episteme: Knowledge and understanding. In K. Timpe & C. Boyd (Eds.), *Virtues and their vices* (pp. 285–302). Oxford: Oxford University Press.

Green, M., & Williams, J. N. (Eds.). (2007). *Moore's paradox*. Oxford: Clarendon Press.

Greenberg, C. (1961). *Art and culture*. Boston: Beacon.

Greenberg, C. (1966). Modernist painting. In G. Battcock (Ed.), *The new art* (pp. 100–110). New York: Dutton.

Grice, P. (1989). Logic and conversation. In *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Grimm, S. (2006). Is understanding a species of knowledge? *British Journal for the Philosophy of Science*, *57*, 515–535.

Grimm, S. (2008). Explanatory inquiry and the need for explanation. *British Journal for the Philosophy of Science*, *59*, 481–497.

Grimm, S. (2014). Understanding as knowledge of causes. In Abrol Fairweather (Ed.), *Virtue epistemology naturalized: Bridges between virtue epistemology and philosophy of science*. New York: Springer.

Habermas, J. (2001). Truth and society: The discursive redemption of factual claims to validity. In *On the pragmatics of social interaction* (pp. 85–104). Cambridge, MA: MIT Press.

Hacking, I. (1982). Experimentation and scientific realism. *Philosophical Topics*, *13*, 71–88.

Hacking, I. (1983). *Representing and intervening*. Cambridge: Cambridge University Press.

Hardwig, J. (1991). The role of trust in knowledge. *Journal of Philosophy*, *88*, 693–708.

Hempel, C. G. (1965). Studies in the logic of explanation. In *Aspects of scientific explanation* (pp. 245–295). New York: Free Press.

Herodotus. (1985). *The histories*. New York: Penguin.

Hetherington, S. (2011). *How to know*. Malden, MA: Wiley-Blackwell.

Hills, A. (2015). Understanding why. *Noûs*, *50*, 661–688.

Hughes, R. I. G. (1996). Models and representation. In *PSA 1996* (Vol. 2, pp. S325–336). Philosophy of Science Association.

Hume, D. (1928). *A treatise of human nature*. L. A. Selby-Bigge (Ed.). Oxford: Clarendon Press.

Hume, D. (1977). *An enquiry concerning human understanding*. Indianapolis: Hackett.

Hume, D. (2007). On the standard of taste. In J. Adler & C. Elgin (Eds.), *Philosophical inquiry* (pp. 815–824). Indianapolis: Hackett.

James, W. (1951). The will to believe. In A. Castelli (Ed.), *Essays on pragmatism* (pp. 88–109). New York: Harner.

James, W. (1983). *Principles of psychology*. Cambridge, MA: Harvard University Press.

Jullien, C. (2006). *Esthetique et modes Operatoires en Mathématiques*. PhD dissertation, Nancy-Université.

Kant, I. (1981). *Grounding for the metaphysics of morals*. Indianapolis: Hackett.

Kant, I. (1987). *Critique of judgment*. Indianapolis: Hackett.

Keller, E. F. (1983). *A feeling for the organism*. New York: Henry Holt.

Khalifa, K. (2012). Inaugurating understanding or repackaging explanation? *Philosophy of Science*, *79*, 15–37.

Kipling, R. (1978). The elephant's child. In *Just so stories* (pp. 49–63). New York: Weathervane Books.

Kitcher, P. (1990). The division of cognitive labor. *Journal of Philosophy*, *87*, 5–22.

Klein, P. (1999). Human knowledge and the infinite regress of reasons. *Philosophical Perspectives*, *13*, 297–325.

Korsgaard, C. (1996a). *The sources of normativity*. Cambridge: Cambridge University Press.

Korsgaard, C. (1996b). The reasons we can share. In *Creating the kingdom of ends* (pp. 249–274). Cambridge: Cambridge University Press.

Kripke, S. (1982). *Wittgenstein on rules and private language*. Cambridge, MA: Harvard University Press.

Kripke, S. (2011a). On two paradoxes of knowledge. In *Philosophical troubles* (pp. 27–49). Oxford: Oxford University Press.

Kripke, S. (2011b). A puzzle about belief. In *Philosophical troubles* (pp. 125–161). Oxford: Oxford University Press.

Kuhn, T. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kvanvig, J. (2003). *The value of knowledge and the pursuit of understanding*. Cambridge: Cambridge University Press.

Kyburg, H. (1997). The rule of adjunction and reasonable inference. *Journal of Philosophy*, *94*, 109–125.

Lehrer, K. (1974). *Knowledge*. Oxford: Oxford University Press.

Lehrer, K. (1986). The coherence theory of knowledge. *Philosophical Topics*, *14*, 5–25.

Leite, A. (2004). Is fallibility an epistemological shortcoming? *Philosophical Quarterly*, *54*, 232–251.

Lewis, C. I. (1946). *An analysis of knowledge and valuation*. La Salle, IL: Open Court.

Lewis, D. (1983). Truth in fiction: Postscript. In *Philosophical papers* (Vol. *1*, pp. 276–280). Oxford: Oxford University Press.

Lewis, D. (1999a). New work for a theory of universals. In *Papers in metaphysics and epistemology* (pp. 8–55). Cambridge: Cambridge University Press.

Lewis, D. (1999b). Elusive knowledge. In *Papers in metaphysics and epistemology* (pp. 418–445). Cambridge: Cambridge University Press.

Lin, M. (2000). Making the memorial. *New York Review of Books*, Nov. 2.

Lipton, P. (2004). *Inference to the best explanation*. London: Routledge.

Lipton, P. (2009). Understanding without explanation. In H. de Regt, S. Leonelli, & K. Eiger (Eds.), *Scientific understanding: Philosophical perspectives* (pp. 43–63). Pittsburgh: University of Pittsburgh Press.

Lo, Bernard, & Fields, Marilyn J. (Eds.). (2009). Principles for identifying and assessing conflicts of interest. In *Conflict of interest in medical research, education, and practice* (pp. 44–61). Washington, DC: National Academies Press.

Locke, J. (1894). *An essay concerning human understanding*. Oxford: Clarendon Press.

Longino, H. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton, NJ: Princeton University Press.

Lopes, D. (1996). *Understanding pictures*. Oxford: Oxford University Press.

Mazur, J. (2014). *Enlightening symbols: A short history of mathematical notation and its hidden powers*. Princeton, NJ: Princeton University Press.

McClintock, B. (1973). Letter to J. R. S. Finchman, May 16, 1973. http://profiles.nlm.nih.gov/LL/Views/Exhibit/documents/biographical.html.

McGowan, M. K. (2003). Realism, reference, and grue. *American Philosophical Quarterly*, *40*, 47–57.

McGowan, M. K. (2009). On silencing and sexual refusal. *Journal of Political Philosophy*, *17*, 487–494.

Metropolis, N. (1987). The beginning of the Monte Carlo Method. *Los Alamos Science*, *15* (Special Issue dedicated to Stanislaw Ulam), 125–130.

Mill, J. S. (1978). *On liberty*. Indianapolis: Hackett.

Moore, G. E. (1991). Russell's theory of descriptions. In P. A. Schilpp (Ed.), *The philosophy of Bertrand Russell* (pp. 177–225). La Salle, IL: Open Court.

Moore, G. E. (1993). A defence of common sense. In *G. E. Moore: Selected Writings*. London: Routledge.

Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. S. Morgan & M. Morrison (Eds.), *Models as mediators* (pp. 10–38). Cambridge: Cambridge University Press.

Morrison, M. (1999). Models as autonomous agents. M. S. Morgan & M. Morrison (Eds.), *Models as mediators* (pp. 38–65). Cambridge: Cambridge University Press.

Nagel, T. (1986). *The view from nowhere*. Oxford: Oxford University Press.

Nersessian, Nancy (1993). In the theoretician's laboratory: Thought experimenting as mental modeling. In *PSA 1992* (Vol. 2, pp. 291–301). Philosophy of Science Association.

Nersessian, N. (2007). Thought experimenting as mental modeling: Empiricism without logic. *Croatian Journal of Philosophy*, *7*(20), 125–161.

Nersessian, N. J. (2008). *Creating scientific concepts*. Cambridge, MA: MIT Press.

Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Harvard University Press.

Oxford dictionary of science. (1999). Oxford: Oxford University Press.

Pascal, B. (2007). The wager. In J. E. Adler & C. Z. Elgin (Eds.), *Philosophical inquiry* (pp. 784–786). Indianapolis: Hackett.

Plato. (1997). *Plato: The complete works*. J. Cooper (Ed.). Indianapolis: Hackett.

Popper, K. (2002). *The logic of scientific discovery*. London: Routledge.

Porter, T. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Pritchard, D. (2005). *Epistemic luck*. Oxford: Oxford University Press.

Pritchard, D. (2010). Knowledge and understanding. In D. Pritchard, A. Millar, & A. Haddock, *The nature and value of knowledge: Three investigations* (pp. 1–88). Oxford: Oxford University Press.

Quine, W. V. (1961). Two dogmas of empiricism. In *From a logical point of view* (pp. 20–46). New York: Harper & Row.

Rainer, Y. (1995). Some retrospective notes on a dance for 10 people and 12 mattresses called parts of some sextets. In M. R. Sandford (Ed.), *Happenings and other acts*. London: Routledge.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

Riggs, W. (2003). Balancing our epistemic goals. *Noûs*, *37*, 343–352.

Rowling, J. K. (1999). *Harry Potter and the prisoner of Azkaban*. New York: Scholastic.

Russell, B. (1968). On denoting. In *Logic and knowledge*. New York: Capricorn.

Russell, W. (1783). *The tragic muse: A poem addressed to Mrs. Siddons*. London: Printed for G. Kearsley.

Ryle, G. (1949/2009). *The concept of mind*. New York: Routledge.

Sarkar, S. (2014). Ecology. In E. Zalta (Ed.), *Stanford encyclopedia of philosophy* (Spring 2014 Ed.), http://plato.stanford.edu/archives/spr2014/entries/ecology/.

Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.

Schwartz, R. (1985). The power of pictures. *Journal of Philosophy*, *82*, 189–198.

Sellars, W. (1963). *Science, perception, and reality*. London: Routledge & Kegan Paul.

Sellars, W. (1968). *Science and metaphysics*. London: Routledge & Kegan Paul.

Shah, N. (2002). Clearing space for doxastic voluntarism. *Monist*, *85*, 436–445.

Sigman, J. (1998). Ordinary people: *Trio A* and how dances signify. In *Bodies, souls, and ordinary people: Three essays on art and interpretation*. Dissertation presented to the Department of Philosophy, Princeton University.

Sklar, L. (1999) Philosophy of science. In R. Audi (Ed.), *Cambridge dictionary of philosophy* (pp. 700–704). Cambridge: Cambridge University Press.

Snyder, M. (2012). Teaching Jo Pa. *Academe*, *98*, 55.

Sosa, E. (2007). *A virtue epistemology*. Oxford: Oxford University Press.

Stadler, I. (1982). The idea of art and of its criticism. In T. Cohen and P. Guyer (Eds.), *Essays in Kant's aesthetics* (pp. 195–220). Chicago: University of Chicago Press.

Strevens, M. (2008). *Depth: An account of scientific explanation.* Cambridge, MA: Harvard University Press.

Strevens, M. (2010). Varieties of understanding. Unpublished manuscript.

Strevens, M. (2012). Ceteris paribus hedges: Causal voodoo that works. *Journal of Philosophy*, *109*, 652–675.

Strevens, M. (2016). How idealizations provide understanding. In S. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding: New essays in epistemology and the philosophy of science* (pp. 37–49). New York: Routledge.

Stroud, B. (1984). *The significance of philosophical scepticism.* Oxford: Clarendon Press.

Suárez, M. (2003). Scientific representation: Against similarity and isomorphism. *International Studies in the Philosophy of Science*, *17*, 225–243.

Suzuki, H., et al. (2007). Novel strategy to treat multi-drug-resistant gram-negative pathogens using bacterial conjugation. *Plasmid*, *57*, 237–238.

Szabó, Z. G. (2001). Fictionalism and Moore's paradox. *Canadian Journal of Philosophy*, *31*, 299–300.

Teller, P. (2001). The twilight of the perfect model model. *Erkenntnis*, *55*, 393–415.

Treanor, N. (2013). The measure of knowledge. *Noûs*, *47*, 577–601.

Tufte, E. R. (1997). *Visual and statistical thinking: Displays of evidence for decision making.* Cheshire, CT: Graphics Press.

USA Diving. (2013). *Official technical rules 2015–2017.* http://usadiving.org.

Van Cleve, J. (2005). Why coherence is not enough: A defense of moderate foundationalism. In M. Steup & E. Sosa (Eds.), *Contemporary debates in epistemology* (pp. 168–180). Malden: Blackwell.

Van Fraassen, B. (2008). *Scientific representation.* Oxford: Oxford University Press.

Venter, J. C., et al. (2001). The sequence of the human genome. *Science*, *291*(5007), 1304–1351.

Vogel, J. (1990). Are there counterexamples to the closure principle? In M. D. Roth & G. Ross (Eds.), *Doubting* (pp. 13–27). Dordrecht: Kluwer.

Walden, K. (2007). A practico priori. Unpublished MS.

Weatherson, B. (2008). Deontology and Descartes' demon. *Journal of Philosophy*, *105*, 540–569.

Wedgwood, R. (2002). The aim of belief. *Noûs*, *16* (supplement), 267–297.

Weisberg, M. (2013). *Simulation and similarity.* Oxford: Oxford University Press.

West, R. (1928). *The strange necessity*. Garden City: Doubleday.

White, M. (1965). *Foundations of historical knowledge*. New York: Harper & Row.

Williams, B. A. O. (1973). Deciding to believe. In *Problems of the self* (pp. 136–151). Cambridge: Cambridge University Press.

Williams, B. A. O. (1978). *Descartes: The project of inquiry*. New York: Penguin.

Williams, B. A. O. (1985). *Ethics and the limits of philosophy*. Cambridge, MA: Harvard University Press.

Williams, B. A. O. (2002). *Truth and truthfulness*. Princeton, NJ: Princeton University Press.

Williamson, T. (2000). *Knowledge and its limits*. New York: Oxford University Press.

Wills, G. (1992). *Lincoln at Gettysburg: Words that remade America*. New York: Simon & Schuster.

Wilson, M. (2006). *Wandering significance*. Oxford: Clarendon Press.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Oxford University Press.

Wittgenstein, L. (1961). *Tractatus logico-philosophicus*. London: Routledge & Kegan Paul.

Zimba, J. (2009). On the possibility of trigonometric proofs of the Pythagorean theorem. *Forum Geometricorum*, *9*, 275–278.

Zinn, H. (1980). *A people's history of the United States*. New York: Harper & Row.

Index