Epistemology: Contexts, Values, and Disagreement ed, Christoph Jäger and Winifrid Löffler. Ontos Verlag 2012, forthcoming.

Understanding's Tethers

Catherine Z. Elgin

Abstract: It seems natural to assume that understanding, like knowledge, requires truth. But natural science affords understanding and uses models and thought experiments that are not, and do not purport to be true. To accommodate science, we need a theory of understanding that recognizes the epistemic functions of representations that are not true. I contend that models and thought experiments are felicitous falsehoods. I argue that they afford insight into the phenomena they concern by exemplifying features they share with those phenomena. In effect they show rather than say something about those phenomena.

Introduction

Understanding is holistic. Although we can know isolated facts, understanding a topic – be it the rise of the working class, the defensive strategy of the Boston Celtics, or the function of chlorophyll in photosynthesis – involves an interconnected network of commitments that suitably bear on the relevant facts. This characterization is hideously murky. Much needs to be said about the identity of and interconnections among the networked commitments, the criteria for suitably bearing on the facts, and the cognitive stance that an epistemic agent must take to both the network and the facts. Indeed, my initial claim might seem unnecessarily murky. Many take it as obvious that the strands in the network must be beliefs (perhaps justified or reliably formed beliefs) and that the tie to the facts is truth. Plenty of work would remain to be done, but at least we would have familiar semantic and epistemological resources to draw on.

However obvious it might seem, this answer is wrong for at least two reasons. One derives from Gettier cases; the other from modern science. Gettier cases disclose that justified true belief is not sufficient for knowledge because the factors that afford the justification for a

belief may have nothing to do with what makes that belief true. In such cases, it is just by luck (and too much luck) that justification and truth align. Understanding is vulnerable to the same difficulty. If the Copenhagen interpretation of quantum mechanics (C) is justified but false, and the theory that the universe is endlessly expanding (E) is unjustified but true, someone who believes C or E has a constellation of justified true beliefs but lacks understanding. It is just by luck (and too much luck) that her constellation of beliefs is true. Mature science generates a different problem. It involves idealizations, models and thought experiments which are not and are not supposed to be true, but which figure centrally and ineliminably in the understanding that science provides. The falsity of such devices does not discredit them. Plato maintained that the difference between doxa and episteme is that, unlike doxa, episteme has a tether. Gettier cases reveal that truth – even justified truth – is an inadequate tether; modern science reveals that falsehoods are not always untethered.

I will argue that exemplification supplies a tether that connects epistemologically acceptable theories or accounts to the phenomena they pertain to. Because exemplification involves instantiation, accounts that afford understanding are tied to the facts. Because exemplars need not be true, idealizations, models and thought experiments may figure ineliminably in understanding. To make my case, I need to begin by discussing understanding and exemplification.

Understanding

The term 'understanding' has a variety of uses. I am concerned with understanding as a type of epistemic success. The condition at issue in a sentence like 'The parties to the lawsuit came to an understanding' is of no interest to epistemology, for the so-called understanding is a mere matter of agreement. When Ben ventures, 'As I understand things, the committee has the

authority to decide', the word 'understand' is used as a hedge. He thinks so, but he is not sure. Such a use, although of some interest to epistemology, is far from central. Things are more complicated with a sentence like 'Lesley understands Policur cosmology', if we take it to mean that she is cognizant of, but not that she endorses, the commitments of Policur cosmology. Her cognizance is a proper topic for epistemological investigation. But it is not, under the reading I gave the sentence, a cognizance of cosmology. When we say 'Mark understands photosynthesis', we mean that he understands not, or not just, the theory, but the phenomenon. This is the sort of understanding I am primarily interested in. Typically such understanding is embodied in a theory, so the agent is apt to understand a theory as well. But my topic is the understanding of the phenomena, not merely the understanding of the theory.

It might be nice to start (or even to end) with a real definition of 'understanding', or a demarcation criterion that sharply distinguishes among understanding, misunderstanding and failing to understand. I have none to offer. But even without them, we have verdicts about what we take to be clear cases. Astronomy affords an understanding of the motions of the planets; astrology does not. Chemistry affords an understanding of the constitution of matter; alchemy does not. An adequate epistemology should largely respect these verdicts. It should count most of what we consider clear cases of understanding to be understanding, most of what we consider clear cases of failing to understand to be failing to understand, most of what we consider clear cases of misunderstanding to be misunderstanding; and it should yield principled verdicts on the cases we consider currently undecided. I am not going to venture such an explication here. For my purposes, the main point is that we can identify instances where we are pretty sure that people understand things and instances where we are pretty sure that they do not. These provide touchstones for our epistemological theorizing. One such touchstone is mature science. I take it that our best science embodies an understanding of its subject matter, and that any adequate

epistemology should reflect the fact that it does. This is more problematic than it might initially seem.

Understanding is an epistemic achievement. Ceteris paribus, our cognitive situation vis à vis a topic is better if we understand the topic than if we do not. The issue is what sort of epistemic achievement it is. Unlike knowledge, understanding is holistic. It seems possible to know isolated facts. One can know (as solvers of American crossword puzzles do) that Edo was the capital of Japan and that Ott played for the Giants, but know virtually nothing else about the history of Japan or the Giants. But understanding involves a body of mutually supportive cognitive commitments. The elements of an understanding must hang together. Moreover, the understander should grasp or appreciate how they hang together.

This might suggest that coherence is the hallmark of understanding. But delusions and conspiracy theories are often admirably coherent. The madman who thinks he is Louis XIV interprets the evidence, whatever it might be, to support his delusion; the conspiracy theorist who believes that the UN is plotting to promote vegetarianism does the same. In such cases, each belief nestles comfortably in a network of other beliefs; but many of the beliefs that form the network are unfounded. Coherence then is achieved at the price of plausibility. Seemingly obvious considerations are denied, discredited or warped to shield the belief system from objections.

Coherence theories of knowledge block such cases by imposing a truth requirement which discredits delusions and (false) conspiracy theories (Lehrer 1974). An epistemology of understanding might do the same. Then the difference between knowledge and understanding would be one of perspective. Where the epistemological status of an individual proposition *p* is at issue, the topic is knowledge. Where what is at issue is the cluster of propositions that cohere and collectively vindicate *p* and one another, or at least those that remain when the false beliefs

are excised, the topic is understanding. On such a view, coherence alone is not sufficient for knowledge or understanding.[1] Truth is required as well. Still the worry remains. Even when coherence and truth align, a coherent body of truths might float free of the facts, not be tied to them. Truth, as the Gettier cases show, is an inadequate tether.

I have argued that understanding consists of a system of cognitive commitments in reflective equilibrium (Elgin 1996). These commitments consist of beliefs, methods, standards, perspectives, and so forth. Together they constitute what van Fraassen calls a stance (2002). Some of the commitments are antecedently acceptable. We have good reason to endorse them independently of the system in question. Others are acceptable only because of their role in the system. For example, the scientific community was originally committed to the existence of positrons, not because it had any direct evidence of positrons, 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. Because the elements of the system are reasonable in light of one another, they are in equilibrium; because the system as a whole is as reasonable as any available alternative in light of the relevant antecedent commitments, its equilibrium is reflective. It is a system that the community of inquiry can, on reflection, endorse.

The commitments a community starts with need not be true (or even truth evaluable) and are not assured a place in the system it eventually endorses. Members of the community realize that their antecedent commitments are to some degree inadequate. But being the current best guesses about the matter at hand, the proper ways to investigate that matter, and the appropriate standards for evaluating results, those commitments have some claim on the community's epistemic allegiance; investigators need a reason to revise or reject them. Such reasons are often readily available, so a system under construction need not, and typically does not, incorporate all

initial, seemingly relevant commitments. But if the revised system corrects or rejects them, it should, for the most part at least, show why they seemed as reasonable as they did when they did. The resulting system is fallible and is accepted provisionally. Reflective equilibrium can always be upset by further findings. Nevertheless, a system in reflective equilibrium – one does as good a job as any available alternative at accommodating a community of inquiry's cognitive commitments about a topic, the proper methods for investigating that topic, and the standards that findings should satisfy – constitutes, I maintain, an understanding of that topic. Although there may subsequently be reason to abandon it, such a system is reasonable in the epistemic circumstances.

My account is not a pure coherence theory, since a tenable system is answerable to something outside itself – antecedent commitments, many of which are commitments about the topic. But it might seem too close to a coherence theory to be palatable. The worry is this: The tether I focused on in *Considered Judgment* seems to consist largely of beliefs or other intentional states. The community's evolving understanding must be, in a suitable sense, supported by what its members were already committed to. This may insure that they are not making up their accounts out of whole cloth, but it seems to provide no strong reason to think their conclusions are correct.

Perhaps then we should follow philosophers like John Greco (this volume), Jonathan Kvanvig (2003) and Stephen Grimm (2009) and construe understanding, like knowledge, as factive. Just as one cannot know that Ott played for the Giants unless it is true that Ott played for the Giants, one cannot understand photosynthesis unless at least most of the beliefs that comprise one's network of beliefs about photosynthesis are true. Understanding, like knowledge, then reflects the facts.

If this is correct, it is easy to see why understanding is an epistemic accomplishment. An

understanding of a topic consists of a constellation of mutually supportive truths about that topic. Since the mutual support would confer justification, this is very close to what a coherence theory of knowledge maintains (with the possible exception that in the case knowledge, the support should involve no false beliefs essentially, while understanding might tolerate a few). On a factivist account, the reason we should take a theory to afford an understanding of the rise of the working class, or the role of chlorophyll in photosynthesis, or whatever, is that it consists of a sufficiently large cluster of mutually supportive truths about the topic. Understanding then would be knowledge plus coherence.

The Challenge Posed by Science

This is an attractive picture. But it has a fatal flaw: it does not accommodate natural science (Elgin 2004). Natural science standardly and unabashedly deploys models and idealizations that are not, and are not supposed to be, true. The ideal gas law describes a gas that consists of dimensionless spherical molecules that are not subject to friction and exhibit no mutual attraction. The Hardy-Weinberg formula describes the distribution of genes in an infinite population whose members mate randomly. Not only are there in fact no such things, according to our best theories, there *could be* no such things. But the ideal gas law figures in the understanding provided by statistical mechanics and the Hardy-Weinberg formula figures in the understanding provided by population genetics. Such modeling is not treated as a temporary expedient; science considers it a good way to encapsulate, convey and advance its understanding. Although scientists expect particular models to be superseded as science progresses, they neither expect to, nor think they ought to, abandon the practice of modeling.

To deny that modern science affords an understanding of nature would be mad. So epistemology should explain how scientific models and idealizations figure in and contribute to

understanding. Doing so, I maintain, requires acknowledging that understanding is non-factive. By this I do not mean that understanding, in science or elsewhere, is wholly indifferent to facts, but that its relation to facts is more complex and circuitous than factivist epistemological theories maintain. The position I sketched earlier (and developed in *Considered Judgment*) does not privilege the literal, the descriptive, or the true. Rather, it maintains that different sentential elements of a theory or system of thought perform different functions. Some are ostensible truth-tellers. They are defective if they are not true. And their not being true undermines the contention that the system they belong to affords an understanding of the phenomena. Others are what I call felicitous falsehoods (Elgin 2004). Although they are not true, they are not defective on that account; nor is a theory or system of thought defective for containing such elements. So long as (1) the theory or system as a whole is suitably connected to the phenomena, (2) the requisite epistemic standards are met, and (3) the felicitous falsehoods perform their epistemic functions, the system affords an understanding of the domain.

Not all models are propositional. Besides equations and verbal models, science is rife with diagrams, such as the harmonic oscillator depictions in physics texts, and three dimensional models, such as tinker toy models of proteins. Non-propositional models are not, strictly speaking, false. But if interpreted as realistic representations of their targets, they are inaccurate in much the way that false descriptions of an object are inaccurate. All represent their targets to be as they are not. 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.

Actual populations are not infinite; actual gas molecules are not spherical. If a model is a felicitous falsehood, its inaccuracy stems from its target's failure to completely mirror the model. Its felicity consists in its affording genuine insight into the target. So felicitous falsehoods are

false in respects that do not interfere with, and may even facilitate, their affording such insight.

How can they do this? Suppose Jill wants to know whether Fred is in town. Fiona says, 'He is still in China'. As it happens, Fred is in Tibet. Although false, Fiona's statement is true enough in the conversational context, since if Fred is anywhere in Asia, he is not in Cambridge this afternoon. Given the interests of the interlocutors, it is a mildly felicitous falsehood. This is a trivial case, for the falsehood in question is no better than the truth. Fiona could have said, 'He is out of town' or 'He is still in Asia', or just 'No', thereby serving the immediate purpose of the exchange without running any danger of fostering false belief.

In other cases, however, felicitous falsehoods are more useful. They serve legitimate epistemic purposes more effectively than readily available truths. Thought experiments are cases of this kind. According to the gas laws, pressure, temperature, and volume are interdependent. So, we are told, as the temperature in an enclosed container of gas increases, pressure increases to infinity. In reality, that would not happen; eventually, the container would burst. A thought experiment designed to manifest the interdependence prescinds from that material inconvenience; although this is not stated, it pretends that the walls of the container are infinitely strong. In effect, the thought experiment is a fiction. In entertaining it, we suspend disbelief and play out the consequences that figure in the fictional scenario. Going along with the fiction fosters understanding. The thought experiment discloses something about the behavior of gases – namely, the interdependence of temperature, pressure and volume – which remains constant regardless of the strength of the container. This is clearly a good thing. But the worry is that this can go too far. If we start admitting fictions into the realm of science, where do we draw the line?

To make the worry vivid, consider 'The Elephant's Child', Kipling's story about how the elephant got its trunk (Kipling 1978). Once, the story says, elephants had ordinary noses. But an

insatiably curious elephant's child wanted to know what crocodiles eat for breakfast. He leaned 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 elephant's nose and tugged on it. As the elephant (with the help of a 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 to a child's understanding) of biology. Since I allow science to 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. The difference is that while the ideal gas law affords *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 Larmarckian presuppositions are false, the story seems to have nothing to do with the way the elephant got its trunk. We need to appeal to natural selection for an explanation, and that explanation will be entirely unsympathetic to any tugging-on-noses hypothesis. The challenge then is to vindicate thought experiments and models, while excluding just-so stories.

Exemplification

To meet the challenge I appeal to Goodman's notion of exemplification. I will argue that effective models and thought experiments exemplify features they share with their targets, and make reference to their targets via that exemplification; just-so stories and other infelicitous falsehoods do not.

Exemplification is the relation of a sample or example to whatever it is a sample or

example of (Goodman 1968). Soap manufacturers distribute free samples of laundry detergent to entice customers to buy; logic texts include examples to illustrate rules of implication; politicians appeal to focus groups to learn voter preferences; safety inspectors take air samples to assess air quality in mines. All such cases involve attending to something specific in order to find out something more general.

To fix ideas, let us begin with a familiar case: a sample of laundry detergent. The first thing to notice is that it is in fact laundry detergent. Unlike other advertisements, which simply inform potential customers about a product, the sample is an instance of the product. It is, however, not a *mere* instance of the product, as a cup of detergent borrowed from a neighbor might be. The sample is a telling instance. In Wittgenstein's terms, it shows rather than says something about the product (Wittgenstein, 1947). It serves to point up and make manifest some of the detergent's properties, such as its capacity to remove stains. In so doing, it refers to those properties.[2] An exemplar then is a symbol that refers to some properties it instantiates. But it does not refer to all of its properties. Exemplification is selective. An exemplar highlights some of its features by overshadowing or downplaying others. Even though the detergent sample was free, it would be a mistake to conclude that that brand of detergent is free. Price is not one of the properties that, under its standard interpretation, such a sample exemplifies. Nor does it exemplify properties like being manufactured in Cleveland or being inedible, although it also instantiates these properties.

Because commercial samples typically belong to regimented symbol systems, their standard interpretation is straightforward. Consumers normally know what properties such symbols exemplify. But not all exemplars are so regimented. The results exemplified by a public opinion poll may be difficult to interpret. It might, for example, be unclear whether a poll taken outside a farmers' market reveals dissatisfaction with food prices generally or

dissatisfaction with the price of vegetables in particular. Moreover, many exemplars are sui generis; they belong to no even roughly regimented system. Any item comes to exemplify simply by being used as a sample or example. An unassuming rock in the woods comes to exemplify 'feldspar' simply by being pointed out as an example of the mineral. A teacher can take a student's paper to provide an example of the sort of paper he wants his students to write, or the sort wants them not to write. Depending on how he uses it, it might exemplify its content, its form, or even its format. Whichever of these it exemplifies, it does so with some measure of generality. Even if the paper exemplifies its content, the teacher is not recommending that other students submit papers with the very same content. Rather, he is recommending that they do 'this sort of thing', where context, it is hoped, supplies values for the missing parameters. These are characteristics that exemplars share with indexicals; words like 'here' and 'there', 'now' and 'then' are similarly dependent on context.

To complicate matters further, the same item can exemplify different properties in different contexts. The detergent sample might exemplify 'free' in a symposium on effective advertising, even though it does not exemplify 'free' in the context of the customer's deliberations about the desirability of switching brands. In yet other contexts, the packet of soap powder does not exemplify at all. If it remains in the warehouse because the advertising campaign is scuttled, although its contents still instantiate the capacity to remove stains, the packet of detergent does not exemplify that capacity or any other feature. Interpreting exemplars then requires ascertaining the dimensions along which an item exemplifies, and the grain with which it does so. In this regard, exemplars are no different from other context sensitive symbols.

Exemplification involves a dual representational relationship. An exemplar refers to a property or cluster of properties it instantiates and thereby to the extension of that property or cluster. The properties in question might be properties that the exemplar alone instantiates. If

there are such things as haecceities, an individual who exemplified her haecceity would exemplify a feature that nothing else could share. If, as a matter of contingent fact, Trixie is the only dog who walks on two legs, she is the only individual who, as a matter of contingent fact, can exemplify the property of being a canine biped. Such exemplars represent unit classes. Others function as representatives of wider classes. By exemplifying greenness, a patch on a color chart refers greenness and thereby to green things generally.

Putting the matter this way may make the twofold reference look trivial. One green thing serves as a representative sample of green things generally, and only a green thing can do so. Sometimes exemplification is that trivial. But often it is not. For an exemplar can exemplify a cluster of properties. If the detergent sample simultaneously and without change of context exemplifies its brand, its fabric softening capacity and its capacity to remove grass stains, then it refers to the extension consisting of instances of that brand of detergent which can soften fabric and remove grass stains. If the sample is a representative sample, in relevant respects there is nothing special about it. In terms of cleaning power nothing differentiates the sample from any other equal size portions of that brand of detergent. With respect to exemplified properties then, what holds of the sample holds of that brand of detergent generally.

In determining that there is no relevant difference between the detergent sample and the soap powder it represents, we rely on background assumptions about what differences there are, and which of them would make a sample unrepresentative. If the soap in the sample has the same chemical constitution as the rest of the detergent of that brand, we are apt to conclude that it is representative. Then if the sample removes grass stains, it is reasonable to expect that other similarly sized portions of the detergent do too. In making this inference we rely on the background belief that whatever has the same chemical constitution has the same cleaning power.

The case of a focus group is slightly more complicated. A focus group is chosen on the basis of demographic features, such as age, race, gender, and income, which are supposed to be proxies for the factors investigators are interested in. With some trepidation, they take the focus group's opinions to be representative, because collectively the group constitutes a demographically representative sample. If the demographic factors are well chosen, investigators have reason to believe that what holds of the focus group holds of the class they are interested in. There is then an epistemically well grounded basis for projecting from the exemplar to that class.

They can of course be wrong. The conviction that an exemplar is representative of a particular extension is undermined if the relevant assumptions are unfounded. Some public opinion polls before the 2008 US presidential election were unrepresentative because the pollsters relied on telephone directories in deciding whom to survey. The driving idea was that to find out how people in a given region feel about the issues, you get the local phone book, call people and ask them. Home address was taken to be a reliable indicator of other demographically relevant factors. Phone numbers were thought to be keyed to home addresses. Such a polling procedure used to work. But now, unlike their elders, many young people have no land lines and no directory listings; their mobile phone numbers bear no correlation to where they live. So the polls that relied on telephone directories were skewed toward older voters. The falsity of the belief that most potential voters had land lines and were listed in local telephone directories undermined the reliability of the polling practice. The opinions gleaned were not representative of the population of interest.

To summarize: a well chosen exemplar can afford epistemic access, not only to some of its own properties, but also to a wider class of cases. Because exemplars are symbols, they require interpretation. Because their reference depends on context, interpretation is keyed to

circumstances. Although an exemplar refers to the extension consisting of items that share the exemplified properties, that extension can be described in multiple ways. So an exemplar can be informative. But because a non-trivial identification of that extension relies on fallible background assumptions, although exemplars can afford fairly direct epistemic access to the properties they exemplify, the epistemic access they provide to wider classes of cases is sensitive to the adequacy of the background assumptions.

If we focus only on commercial or pedagogical exemplars, we might conclude that the function of exemplification is primarily heuristic. Commercial samples and textbook examples are designed to display what is already known. But not all exemplars are like that. In the course of product development, the manufacturer might test a detergent sample to find out whether it is effective against grass stain. Until the test is run, no one knows whether the sample exemplifies 'eliminates grass stain' or 'fails to eliminate grass stain'. Until the public opinion poll is carried out, no one knows whether it will exemplify concern or complacency about corruption. So exemplars have the capacity to advance as well as exhibit knowledge or understanding.

Exemplification in Science

In principle any exemplar can exemplify any property it instantiates and any property that is instantiated can be exemplified. But what is feasible in principle may not be straightforward in practice. Exemplification of a particular property is not always easy to achieve, for not every instance of a property affords an effective example of it. Considerable stage setting is often needed to bring a rare or recondite property to the fore.

Exemplification is sometimes achieved simply by directing, or redirecting attention. A psychologist can bring test subjects to exemplify automaticity by redirecting attention from the contents of their responses to their reaction times. In effect, she just reframes the phenomena,

reversing figure and ground. In other cases exemplification requires isolating aspects of phenomena. To exemplify the electrical conductivity of water, a scientist distills out impurities before running her experiment. Elaborate experiments may be required to exemplify subtle, difficult to discern properties.

In yet other cases things are even more complicated. Some properties fuse in such a way that they cannot in fact be prized apart. Preferential mating, migration, natural selection and genetic drift affect the distribution of genes in a population. They are always present in nature and two of them – natural selection and genetic drift – are ineliminable, even in the lab. So biologists can neither discover in nature nor create in a lab a population that is not subject to these contingencies of fortune. They are, as we say, facts of life. But to measure genetic change, biologists need a baseline – a measure of how alleles would redistribute if no change other than redistribution was taking place. Here is where felicitous falsehoods reenter the picture. To obtain the baseline, biologists create a model – the Hardy-Weinberg model – which describes the distribution of genes in an infinite, isolated population of organisms that mate randomly and are not subject to natural selection. That the population is infinite counteracts the effects of genetic drift; that mating is random insures that neither physical proximity nor genes that give rise to attractive phenotypes have a preferential advantage; that neither selection nor migration takes place insures that no novelties are introduced into the gene pool. The distribution of genes in this imaginative scenario exemplifies the factor in genetic redistribution is not a manifestation of genetic change.

The conditions described in the model do not – indeed could not – occur in nature. Nevertheless, the model highlights a factor in what does occur in nature, and enables scientists to discern it and play out its consequences in a way that they could not do if they were restricted to accurate factual representations. Models are felicitous falsehoods that exemplify features they

share with their targets. They diverge from their targets in unexemplified features. That divergence enables them to make manifest features that are normally obscured. But because the features with respect to which they diverge are not the features with respect to which they symbolize, models refer to, hence are models *of* their targets.

As is the case with more mundane exemplars, the interpretation of models depends heavily on background assumptions that might be false. So we go out on a limb when we take the target to instantiate the properties its model exemplifies. But that limb is no shakier than the limbs we go out on when we generalize inductively from a limited body of evidence. Just as we are vulnerable to ill chosen or misleading evidence in inductive reasoning, we are vulnerable to poorly designed or misleading models. This is unfortunate. But we've long realized that Cartesian certainty is not in the offing.

This discussion reveals something about the cognitive function of models but how does it relate to understanding? Because exemplification plays a central role in systems of thought in reflective equilibrium, the role of exemplification in my account blocks the charge that my epistemology is, or is a near relative of, a coherence theory. There is no worry that a tenable theory just floats above the phenomena, that its observational and predictive success could, for all we can tell, be due to brute luck or to pre-established harmony. Exemplification requires instantiation. So any exemplar (whether or not it is a felicitous falsehood) that exemplifies a property instantiates that property. This secures a referential connection to the world.[3] Because the exemplar affords epistemic access to the property it exemplifies and a basis for believing that the target has that property, this referential connection is also an epistemological connection. Observational and predictive success are evidence that we have correctly identified the extension that the exemplar is representative of.

Moreover, felicitous falsehoods can show not only that something obtains but also why it

is significant. So they contribute to understanding by highlighting features that would otherwise be obscured, thereby enabling us to see subtle, but potentially significant patterns in the phenomena. By interpreting the behavior of actual gases as variants on the supposed behavior of ideal gas, we may be able to understand aspects of actual gas behavior that would be overshadowed by the complex interactions among differently sized, differently shaped, actual gas molecules.

The epistemological difference between felicitous falsehoods and just-so stories consists in this: because felicitous falsehoods exemplify features they share with their targets, they afford epistemic access to features of those targets. These features are typically not simple monadic properties, but intricate structural and, in many cases, dynamic properties. So felicitous falsehoods enable us to apprehend how things are beneath the welter of complexities that we typically encounter. Just-so stories do not exemplify features they share with the phenomena they purportedly concern. If they were true, such stories would be informative. This, such a story would say, is in fact how the elephant's trunk developed. But given that they are not true, they contribute nothing to the understanding of the phenomena. Felicitous falsehoods have strong tethers to the phenomena they concern; just-so stories lack such tethers.

The conviction that understanding is factive rests on the (typically tacit) assumption that truth is the only secure link between theories and the world. The driving idea is that if the terms in a theory denote real things and the claims made by the theory accurately characterize those things and (enough of) the relations among them, the theory embodies an understanding of those things. I have argued that exemplification supplies another strong and secure tether. Because exemplification requires instantiation, an exemplar is guaranteed to instantiate the properties it exemplifies. The tie to real, existent properties is thereby assured. Because exemplification requires reference to the extension of those properties, the exemplar provides an avenue of

access to other members of that extension. An exemplar, as we have seen, need not be truth evaluable. Neither the detergent sample nor the diagram of a harmonic oscillator has a truth value. And even if an exemplar, such as the ideal gas law or the Hardy-Weinberg formula, has a truth value, neither its accuracy nor its adequacy need be undermined by its falsity. So long as the properties it exemplifies are properties of the members of the extension it represents, an exemplar is accurate. So long as its serving as a representative of that extension promotes our epistemic ends, the exemplar is adequate. Questions of interpretation may arise about just which properties a given exemplar exemplifies and what (non-trivially characterized) extension it represents. But these are not obviously more troublesome than the questions of interpretation that arise for any statement of scientific fact. Still, exemplification is no epistemological panacea. The usual problems surrounding induction, misleading evidence, and so forth still arise. Nevertheless, exemplification enables us to vindicate felicitous falsehoods, and thus to explain how scientific accounts that are not and do not purport to be true figure in a genuine understanding of the way the world is.[4]

References

Elgin, C., 2004: True Enough, *Philosophical Issues* 14, 113-131.

Elgin, C., 1996: *Considered Judgment,* Princeton: Princeton University Press.

Goodman, N., 1968: *Languages of Art,* Indianapolis: Hackett.

Greco, J., 2012: Intellectual Virtues and their Place in Philosophy.

Grimm, S., 2009: Reliability and the Sense of Understanding, in *Scientific Understanding,* ed. H. de Regt, S. Leonelli, and K. Eigner, Pittsburgh: University of Pittsburgh Press, 83-99.

Kipling, R., 1978: The Elephant's Child, in *Just So Stories,* New York: Weathervane Books, 49-63..

Kvanvig, J., 2003: *The Value of Knowledge and the Pursuit of Understanding,* New York: Cambridge University Press.

Lehrer, K., 1974: *Knowledge,* Oxford: Clarendon Press.

Morrison, M. and Morgan, M., 1999: Models as Mediating Instruments in *Models as Mediators,* ed. M. Morrison and M. Morgan, Cambridge: Cambridge University Press, 10-37.

Van Fraassen, B., 2002: *The Empirical Stance,* New Haven: Yale University Press.

Wittgenstein, L., 1947: *Tractatus Logico-Philosophicus.* London: Kegan Paul, Trench, Trubner & Co.

1 Neither coherence theorists of knowledge nor coherence theorists of understanding insist that every belief in the relevant coherence cluster must be true. According to Lehrer (1974), a true belief that *p* is known just in case it coheres with the believer's doxastic system and coheres with the system that would result if all false beliefs in her doxastic system were excised. According to Kvanvig (2003), an agent who understands a topic may harbor a few, relatively peripheral false beliefs about it.

2 As I use the term, a property is whatever the members of an extension have in common.

3 The issue is slightly more complicated than this suggests, for models can be layered. A theoretical model can take as its immediate referent a phenomenological model which in turn refers to a data model which in turn refers to the facts. Each of the models exemplifies a feature it shares with its target, but only at the end of the sequence is the the target comprised of real-world phenomena. See Morrison and Morgan 1999.

4 I am grateful to Jonathan Adler and Brad Armendt for discussions of an earlier draft of this paper.