

---

# Why (Study) the Humanities?

THE VIEW FROM SCIENCE

*Jenann Ismael*

## 9.1. Introduction

Questions about the value of the humanities and the relationship between the sciences and humanities have been very much in the news recently.<sup>1</sup> Just a brief review in the public press shows scientists and humanists weighing in and responding to one another. Public opinion is shifting in favor of science and technological education. There are two related challenges that have been leveled about the value of the humanities.

1. There is a threat to the perceived value of the humanities in the culture at large. This manifests itself in two ways: reduced public support for humanities research, and students being steered away from studying the humanities in university.<sup>2</sup> One often hears complaints of the form, “Why should we as a culture invest resources in humanities education? Why should we spend good money for our children to study French literature, or why should the state subsidize degrees in philosophy? Science and engineering degrees are effective ways of getting jobs and we (as a country) need

---

1. Stanley Fish, Steven Pinker, Philip Kitcher, Daniel Dennett, and Martha Nussbaum have all weighed in on the public discussion. A new report commissioned by a bipartisan quartet of lawmakers looks at the role of the humanities and social sciences in public education: <http://www.humanitiescommission.org/>.

2. This is the threat addressed here: <http://today.duke.edu/2013/06/humanitiesreport>.

more scientists and engineers, but what is a degree in cultural anthropology worth?”

2. And then there is the threat to the humanities mounted by those who claim that in a scientific worldview the human sciences will ultimately be absorbed into (or replaced by) the hard sciences, and there will be no place for the humanities as a source of knowledge.<sup>3</sup> Those who defend this position base it on the claim that we are bits of matter, alongside other bits of matter, governed by material laws, and so understanding ourselves is not different in principle from understanding celery or cells. The great complexity of the human has fostered the illusion that human behavior is different in kind, but that position is indefensible from the point of view of science.

On the face of it, these are two quite different challenges. In response to the second, the position that some have retreated to is to relinquish the claim that the humanities provide a source of knowledge about the human being and hold that it “makes our lives better” in other ways. In response to the first, some have argued that contrary to appearances, a humanities education will make you a better lawyer, businessman, get you a better job, or make you better at public relations (see Rand 1999). Martha Nussbaum has argued that it makes you a better citizen and so is crucial to the success of the polis.<sup>4</sup> A more dismissive answer is provided by Stanley Fish, who rejects corporate or economic values as the sole or ultimate arbiter of value. He demands to know why the humanities should have to justify themselves by those standards any more than corporate values have to justify themselves by the standards of the humanities. One might make more money as an engineer, but one would be culturally illiterate, historically ignorant, and un insightful. Rhetorically, Fish is right. There is no reason that the humanities should have to justify themselves by the standards of the sciences. His response, however, makes it sound as though the humanities are a pleasurable diversion to be enjoyed by those with the privilege of leisure, and that is not the kind of defense that

---

3. Alex Rosenberg is the most outspoken advocate of the position, though one finds glimmers of it in the opinions of many scientists and philosophers, and it is a challenge that any self-styled naturalist has to address in their own thinking.

4. Nussbaum (2010). Fareed Zakaria, in an interestingly related recent piece in the *Atlantic*, has argued that education in the humanities, as opposed to STEM education, makes one more creative and entrepreneurial and that is what has allowed America to flourish economically despite lagging behind in STEM education.

will persuade a parent to support a child's decision to study art history rather than (say) physics. A deeper defense would say what kind of knowledge the humanities provide, how it differs from that of the sciences, and why it is indispensable in a well-lived life. It would address the relationship between the scientific vision of the human being and the humanistic one explicitly, and say whether the humanistic vision is undermined by what science is teaching us about ourselves. This way of putting it brings it close to the second problem, so I want to begin by addressing that one.

## 9.2. The Physics of Open Systems

Every physical system falls under the scope of general microscopic laws that govern the universe as a whole. Those laws are exact and exceptionless. At the level of human behavior those laws are also local and deterministic.<sup>5</sup> Consider any subsystem of the universe and an enclosing sphere of any diameter around that system.<sup>6</sup> We call the variables that characterize the world on the boundary and outside the sphere exogenous, and those that characterize the interior of the sphere endogenous. Locality entails that the values of variables on the surface of the sphere screen off (render irrelevant) the values of variables outside the sphere. This, together with determinism, entails that if we know the initial state of the matter inside the sphere and all of the forces that impinge on the surface of the sphere over any interval of time, we can predict the behavior of anything inside the sphere with certainty. This holds regardless of the size of the sphere and regardless of what kind of matter it encloses. Let's call this the DL principle (for deterministic-local).

The DL principle holds for the collection of dust particles in this little bit of space. It holds for planets. It holds for toasters, and tree frogs. And it holds also for the human being. So, here's a human being. Here are the exogenous influences impinging on a sphere enclosing it. What physics tells us is that if we have the initial state of the body and we know the forces impinging on the

5. The classical setting simply gives us a precise physical framework in which the challenge takes its sharpest and most pressing form. Until the interpretation of quantum mechanics is settled, it is impossible to say definitively whether it makes a difference to human action, though there are no **positive** physical reasons right now for thinking it will make any significant difference. At the level of brain function relevant to human action, classical physics **is** the effective theory.

6. We make it a sphere just for convenience. We can try to get it as close to the boundaries of the body as possible, though the boundaries of the body are a little vague (is the hair on your skin part of your body? What about the skin cells just flaking off . . . etc.).

surface of the sphere through some interval, we can calculate how the body will move over that interval.

That sounds like a very strong result. But it is actually quite weak for the following reason. The number of degrees of freedom in the microscopic state of the enclosed system =  $(6 \times \text{the number of particles that compose the system})$ . The number of degrees of freedom in the exogenous variables = <sup>roughly</sup>  $(6 \times \text{the number of particles in the rest universe})$ .<sup>7</sup> The DL principle says if we know the precise values of all of those variables, we can calculate with microscopic precision how any physical system enclosed in that sphere will behave. It does not entail that knowing anything less than that will let us calculate anything about that system. It entails, that is to say, that if we know **everything** we can calculate everything. It doesn't entail that knowing **less than everything** will let us calculate **anything**.

Here's what I mean. Let's consider a **quarter**, and let's suppose that the only thing we really care about concerning its behavior is whether, if it is tossed right now, it will land heads. The DL principle tells us that if we know its initial microstate and the values of all exogenous variables impinging on it through the course of the toss, we can calculate with certainty how it will land. But it also tells us—and here is the crucial part—that *nothing less will do*. The reason is that the dynamical laws entail that whether a coin lands heads or tails on a given toss is so sensitive to the microscopic values of so many exogenous variables—e.g., the exact angle at which it is released, the Brownian motion of dust particles in the air, indiscernible fluctuations in speed and direction of wind—that unless we know all of this with perfect precision, or can control their effects, the outcome of the toss cannot be predicted.<sup>8</sup> What is happening here is that coin tosses amplify ignorance. They transfer any ignorance we have about the microscopic state of any of the particles that might make some impact on the boundary of the sphere into ignorance of the result of a given toss.<sup>9</sup>

7. One qualification is necessary to make this strictly correct, but it makes no practical difference here. If the universe is big enough and the sphere is small enough, and we live in a universe in which information cannot be packed densely in any region of space, there may be fewer degrees of freedom on the boundary of the sphere.

8. In the case of coin tosses, we can attach probabilities to outcomes (fifty-fifty for fair coins, different probabilities for weighted coins), because there are stable relative frequencies over the dynamically relevant exogenous variables, but that itself is not something that is generally available. If we can control the effects of exogenous variables, the outcome can be reliably predicted. It is because of the lack of knowledge and absence of control in everyday circumstances that makes coin tosses effectively unpredictable.

9. Note here that this is a different phenomenon from the sort of unpredictability that arises with chaotic systems. In the case of chaotic systems, the unpredictability has its source in the nonlinearity of the

### 9.3. Science Is about Generality; Humans Are All Specificity

When you see how weak the DL principle is, you might be surprised that we can have a predictive science of any open subsystem of the world. And, indeed, the vast majority of open subsystems (if by “open subsystem” we mean the matter contained in any region of space-time around which we can draw a closed boundary) do not exhibit the kind of regularity that makes it possible to formulate simple, predictive laws that express their behavior as a function of the state of the environment.<sup>10</sup> But some do. And we can look at the kind of dynamics that a bit of matter has to have to make a predictive science of its behavior possible.

Let's start with a toaster, and let's suppose that we are interested primarily in its gross, discernible macroscopic behavior. The toaster sits inactive when the lever is up. Depressing the lever lowers a chassis in which bread is placed and initiates a process in which electricity heats internal grills to a certain temperature for a fixed amount of time. When the process is done, the chassis returns to its normal position. There are simple laws for this behavior of toasters because a toaster has a (relatively) fixed internal structure that (more-over) makes only a small number of variables relevant to that behavior. The position of the lever and the knob to determine grill time matter, but not the presence of wind or the absence of noise. Small differences in input don't produce grossly different responses. It doesn't matter *exactly* how hard or fast you push the lever.<sup>11</sup> The internal wiring is hard, and designed to produce the

---

equations and arises for closed systems as well as open ones. When dynamical equations are nonlinear, tiny differences in initial state can lead to radically different outcomes. That means that anything less than perfect precision in knowledge of the initial state can leave us with very great uncertainty (uncertainty spread all over phase space) about the final state. The sort of failure of unpredictability under discussion here has nothing to do with nonlinearity of dynamical equations. It has to do, rather, with the *openness* of the systems and their sensitivity to exogenous variables. It arises even if the dynamical equations are linear.

10. Philosophical usage often counts as a law only perfectly universal, fundamental generalizations. I am using the term loosely, to include defeasible, counterfactual supporting regularities of all kinds. So what we are looking for is a description of the gross behavior of a system (typically its movements) as a function of environmental impact. In the case of living organisms, these kinds of laws are often thought of in terms of its responses to stimuli.

11. Of course, this is all true only if we describe the toaster at a very coarse-grained level, and restrict the predicted behaviors. If we include temperature in “discernible behavior” and allow very fine discriminations, none of this would be true. The behavior would exhibit a high sensitivity to microscopic changes in its environment.

same response for the life of the toaster.<sup>12</sup> We don't need to know very much either about the environment, or what goes on inside a toaster to know how to expect it to behave, so long as it is operating normally.

Using this as a model, we can say that there are simple laws that allow us to express the behavior of an open system as a function of its environment wherever there exists a reduced variable subspace of the physics of the universe that makes only certain variables relevant to behaviors of interest, ~~fixing~~ the system's internal structure so that the impact of those variables is constant (over short time scales), and linear (or approximately so).<sup>13</sup> Toasters break down and wear out.<sup>14</sup> But breakdowns are infrequent, so we can ignore them at ~~some~~ cost in the exactness of the laws, and wearing out is a gradual, predictable change that can be incorporated into predictive laws.

Many systems, often of tremendous complexity, conform to this model. For example, in self-organizing systems like termite colonies or slime molds, even though these systems have an enormous number of microscopic parts and the interactions among those parts contain feedback loops that would make the physical equations for their conjoined behavior effectively unsolvable even if we knew the initial state of each of them, there is emergent behavior regular enough to permit a reduced set of variables in terms of which we can find simple, predictive laws for the configuration. These simple, predictive laws, moreover, can be discerned without a good understanding of the underlying physics.<sup>15</sup> There is no ~~monolithic~~ account of the dynamical underpinnings of systems that exhibit this structure. Scientifically, we've only begun to understand them. Although the open systems that exhibit this structure are the ones that tend to attract our scientific attention, it should be noted

---

12. What I mean by "hard" here is fixed, relative to a range of contexts. One way of making things hard, in this sense, is by making them rigid. But we can also make a connection between A and B hard relative to contexts C by having a lawlike regularity connecting A and B scaffolded by structure present in C. So, for example, the connection between the button on my garage opener and the garage is hard when the two are in spatial proximity and all of the background things that need to be in place for it to function properly are there.

13. There are three things to note: (1) the existence of such laws is relative to behaviors of interest (see fn.13), (2) in saying that the internal structure is fixed, we don't mean that it is static, but that it changes in ways that support a fixed relationship between input and output variables, and (3) simplicity has been left unanalyzed here. More can be said, but an intuitive conception of simplicity will do well enough for our purposes.

14. We do need to know how they work to fix them, or to explain their behavior when they start behaving anomalously.

15. This is not an exhaustive list.

that among all the open systems in the universe, they constitute a very small minority.

Living organisms ~~also~~ often exhibit this structure. Beginning with the macromolecules of DNA and RNA, we can trace step by step, as animal life moved from the simple ability to respond to frequently recurring environmental conditions to much more powerful mechanisms for producing behavior finely attuned to circumstance. Living systems have parts that are bound together in a relatively fixed configuration designed over ecological time to produce advantageous responses to stimuli.<sup>16</sup> The frog brain, for example, is a remarkably well-designed instrument for (among other things) getting frog bodies to respond in reliably predictable and adaptive ways to stimuli. The frog responds to the image of a passing fly with a flick of the tongue because that is what it has been designed to do.<sup>17</sup> There is a lot going on in the frog brain, but the activity is designed (in part) to produce regular macroscopic responses to a particular class of stimuli: to filter out the noise, ignore the differences between one flyspeck and another, and get the tongue where it needs to be. Because this is what the frog brain was designed to do, one frog will do the same as another, and the behavior is more or less constant over time. The same goes for mongooses, ~~and~~ mole rats, and three-toed sloths.<sup>18</sup>

Viewed as part of this progression, what is special about the human being is that in the human mind we see the development of a cognitive platform for the emergence of a new behavior management strategy involving deliberation and choice. Instead of passing through a set of internal filters designed to keep behavior covarying reliably with features of the local environment, the effect of the stimulus on behavior is mediated by a process that seems almost perversely geared to undermine any possibility of general laws of human behavior. Consider a mundane example of choice.

A mother walks into a shop to grab a coffee on her way home. She sees her young daughter, who is supposed to be at the library, talking to a boy that the mother doesn't trust. She retreats before being seen, and walks home ruminating about what to do. The incident makes her realize that her daughter is growing up, a time she has been preparing herself for psychologically,

16. I borrow the phrase "wiring and connections" from Peter Godfrey-Smith (2004).

17. Designed by natural selection or whatever hidden hand shaped the universe.

18. Responses to environmental stimuli can adapt, but adaptation in an individual frog happens slowly, and only with sustained pressure from the outside. Like the wearing out of the wiring of the toaster, it can typically be anticipated and incorporated into the laws.

but it seems too soon. She begins to think about her own life and her plans to have other children. A reconfiguration in her view of her daughter begins to occur. She wonders how much she doesn't know about her daughter, and feels instinctively that this is a delicate time in her daughter's life. By the time she is home, she has resolved to spend more time with her daughter and reaches for the phone to make dinner reservations at an old, favorite place.

The deliberative process is bringing into the causal chain between stimulus and response, here, a whole lot of stored information collected over a lifetime of personal experience. This includes beliefs not only about the world, but also about herself, her daughter, their place in the world, personal plans, memories, intentions, and commitments. If a fly's tongue snaps out unreflectively at a passing fly, the mature human adult runs his experience through a much more complex transformation that can—in the most reflective decisions—call up everything he is and has become. We are not Hamlets at every moment of our lives, but we have Hamlet moments. The result is that the choice-governed aspects of human behavior does not just depend on the immediate stimulus, but is open to influence from an in-principle unlimited number of sources, all stored in memory and encoded in the soft structure of the brain. As if that weren't bad enough, the bearing of this information on behavior is filtered through a quite complex set of higher-order principles for choice (goals, values, priorities, beliefs about who we are and who we want to be) that themselves vary from one person to the next and are constantly evolving. Where do these higher-order principles come from? They are the products of experience, in some sense, but they are forged under the hot fire of personal reflection. And reflection is one of those processes that has the hallmarks of unpredictability. It is holistic, self-feeding, and ongoing. Even if there were a deterministic equation that someone could write down that would describe it, the feedback would make the equation unsolvable within a few steps. So not only does choice make all of one's personal history potentially relevant to one's present behavior, it makes the bearing of personal history on behavior subject to second-order principles that are themselves highly variable both across the population and over the history of a single subject. And even if we held all of that fixed, very small differences in stimulus can produce huge differences in response. We are highly attuned to tones of voice, subtle cues we are scarcely conscious of (e.g., the something in the air that tips the balance between accepting and declining an invitation to speak in New Zealand). Small difference *makers* can lead to very large differences.

The effect of all of this is something that common sense knows, viz., that human behavior is highly individual and deeply unpredictable.



Responses to stimuli vary from one person to the next and over time in the history of a single person.<sup>19</sup> In a toaster, the structure that mediates input and output is “hard,” ~~and~~ ordinarily changing slowly and in predictable ways. The structure that mediates stimulus and response in humans is “soft” and changes at lightning speed in ways that are directly relevant to its discernible behavior. The very processes that are designed to stabilize regular behaviors both across the population and over time in other animals are geared to produce variability and differentiation in the human being. All frogs flick their tongues at passing flies, birds flock, and bees dance. The *choice-governed behavior* of the individual human being, by contrast, is so highly sensitive to the specificities of personal history and belief, all encoded in the soft structure in the brain that, from an external perspective, choice effectively randomizes the effect of stimulus on response. There is no simple, general relationship between environmental stimulus and behavior. Whereas explanations of frog and fish behavior typically refer to the environmental stimulus and general laws of frog and fish behavior, explanations of human behavior typically appeal to accidents of individual history of which there are no traces in the stimulus and that can’t be collected under general laws.<sup>20</sup> Laws seek to capture generalities, and human beings are all specificity.<sup>21</sup>

19. We are still physical systems, and so in principle if we knew the microphysical state of the world at some cross section of a person’s back light cone and we knew the laws that govern the universe as a whole, we could predict their behavior. The same is true for any physical system. The point is that in the case of the human being (or any system whose macroscopic behavior is determined a mechanism that draws on a fund of information evolving as quickly and idiosyncratically as personal belief), nothing *less* will do.

20. Philip Kitcher’s strategy is to break down the difference between science and humanities, to make the difference, as he puts it, one of degree rather than kind. That would make the humanities cousins of the sciences aiming for the same kind of understanding, employing the same kinds of methods, but with a less impressive history of success (though he tries to argue that this too is overstated). I think that he is right in much of what he says, and that the differences that I emphasize might be differences of degree rather than kind.

21. The line between humans and animal cognition is more complex than the contrast between frog tongue-flicking and human choice suggests. On one end of the spectrum, we see simple organisms exhibiting very regular responses to the environment. On the other, we see humans exhibiting highly irregular responses to the environment on the other. As we move from one end of this spectrum to the other, we see increasing complexity both in the character of the stimulus and in the subtlety of the response. Experts disagree on whether there is a hard line to be drawn anywhere along this spectrum. It remains true that choice has the effect of bringing into the chain between stimulus and response information encoded in memory, and so the more information is encoded in memory, and the more that information varies across the population and over time, and the more behavior is governed by choice, the less regular responses to the environment will be.

This is not to say that there aren't very useful algorithms for predicting how people will act. We rely on these algorithms every day of our lives. Their success depends on the fact that although we weren't designed simply to respond in predictable ways to impact from the environment, we *were* designed to be deliberators. If we want to understand systems whose behavior is governed by choice—i.e., to know what makes them tick, to guess how they will respond, to influence their behavior, and to interact with them effectively—we have to get good at *psychological* interpretation. We have to become skilled at understanding what other people believe and desire and feel. We need to understand one another, not in the way that we understand toasters and electrons, but as *persons*, i.e., as subject of experience and emotion, as believers and agents. Anybody that can operate in the social world has rudimentary skill at psychological interpretation, but really learning to see things through other people's eyes, to understand what they are thinking and feeling, what they hope and fear and value, having a rich sense of the complex inner world of another human being in all of its emotional and psychological complexity, ~~that~~ is a kind of understanding that can be fostered and developed by an education in the humanities. Someone who has grown up reading novels will learn to understand the ~~complex hidden internal world~~ that goes on inside another human being. Someone who has studied history will have an appreciation for the complex currents of culture that govern the unfolding of civilizations. Each of the fields traditionally classified as humanities makes a distinctive contribution to this kind of understanding.

Reasons and explanations aren't *just* about prediction. They will also make a contribution to a richer kind of understanding, a kind of understanding that is not just a matter of being able to *predict* how other persons will behave, but being able to see things through their eyes, being fair and generous and empathetic. It enriches our understanding of the world by helping us understand people in the terms in which they understand themselves. That is something that we need to be able to do if we are relate to them not as material systems, but as subjects of experience and sources of agency. For we are social animals: partners, friends, coworkers, mothers, and teachers. The better we are at understanding one another in the way that we each understand ourselves, the better we will be in these capacities. These types of human understanding are not something that one can get from knowledge of physical law. Explanations that invoke reasons teach us how to interact with people *as rational agents* and seekers of value, to affect their behavior by persuasion rather than by trying to control them causally. They teach us to address the rational standpoint and

offer reasons for them to act as we want them to act, so that our effect on their behavior is mediated by their own deliberative processes.<sup>22</sup> They guide in seeing things through other people's eyes and being able to construct a narrative that is fair to all viewpoints. **Understanding** in human affairs requires that kind of understanding. Science is not, and never will be, a substitute for that kind of understanding.

#### 9.4. Science Is about Description; the Humanities Are (Partly) about Guidance

Of course that is only half of the story. What was said above was looking at the processes that mediate stimulus and response in a human being from a third-person standpoint. When we turn from a third-person to a first-person perspective, something else emerges that makes the indispensability of the humanities ~~that much~~ more manifest. Again, it has to do with choice, but this time it concerns the special status that choices have *for the person making them*. To other people's lives, we are observers. We watch them make choices, try to gauge what they are thinking, guess how they feel, what they care about, and why they act the way they do. But we are not mere observers of our own lives. We *live* our own lives and we *make up* our own minds about what to think, and how to act. We bear a special relationship to our own thoughts and experiences that makes it impossible to take a purely detached attitude toward them. There is no way of abdicating our active participation in the making of our choices. From the mundane to the momentous (i.e., whether deciding which socks to wear or whom to marry), the universe will not make those choices on our behalf.<sup>23</sup> The inescapability of choice is our situation in nature. And to be a chooser is to have a special kind of creative role in the production of our lives.

It is worth noticing how little of our own lives, for many of us, is dictated by the practicalities of survival. People need to eat and sleep to survive. In the state of nature, that meant that our daily lives were organized around the necessities of obtaining food, caring for our young, and maintaining shelter in terms largely dictated by our situation. One made a shelter with available

22. The contrast is with manipulation or coercion. Manipulation tries to control the output of the decision process by controlling the input. Coercion tries to bypass the decision process entirely by using physical force to move the body to move directly.

23. We can *choose* to be passive, but choosing to be passive is a way of choosing.

materials, and ate what one could get one's hands on. The structure and content of daily life were not, to a very great extent, a matter of choice. Things are vastly different nowadays. The landscape of opportunities that the world presents is radically expanded. There are countless ways to make a living, and countless ways of maintaining house and home. You get to choose how to make a living, where to live, and whether to have children. And that is to say nothing of what to eat, what to wear, what music to listen to, what newspaper to read, and what to watch on TV. The choices we make are ultimately choices about who we are and who we want to be. Our identities as persons and agents is constituted by them. They give shape and definition to our lives in the way that the hammering of the sculptor gives shape and definition to the unformed stone, transforming an indefinite multitude of potential shapes into a single actuality.

It is in equipping us with tools to address the very personal questions—"What should I do? How should I live?"—that I see the humanities as making an indispensable contribution. A humanities education can, among other things, open up the imagination to the rich array of possibilities of what to be. The bookish child of a farmer who reads *Jude the Obscure* for the first time can see a world open up that he hadn't known. Books—by showing us examples of successful and failed lives—help us decide which of them are worth wanting. What I learned from Plato, Aeschylus, Dante, Goethe, Tolstoy, Joyce, Elliot, and Mann played a very personal role in making me who I am. Once you see all of that—i.e., once you see your role in creating history, rather than being a passive observer—you see that the humanities are not a *recherché* pursuit undertaken in leisure from which people with highbrow tastes draw enjoyment. They help us decide what to make of our lives. They are *the tools of our becoming*.

It is sometimes said that if our scientific knowledge were good enough, at least in a deterministic world, we would be able to predict what we will do and we could just sit back and let it happen. One expression of the threat is embodied in the scientific challenge to free will.<sup>24</sup> The thought seems to be that a completed scientific understanding of the world will push forward the boundaries of prediction, leaving no room for choice. But that there is something wrong with this line of thought should be clear from the observation that without *your* activity, there would be nothing to predict. Our choices don't get made *unless we make them* (see Ismael 2016). Learning physics is not

---

24. This is explicit in Rosenberg (2011), for example.

going to relieve you (in practice or in theory) of the burden of making choices. And it is not going to relieve you (in practice or in theory) of the burden of running your life, or making yourself into what you will be. Understanding *ourselves* will always be an art as much as a science because it involves a form of creation.

## 9.5. Deciding What to Think and How to Feel

When R. W. Hepburn remarked that “One may look upon the “material” of one’s life . . . rather as an artist regards his canvas and paint or a sculptor his stone” (Hepburn and Murdoch 1956, 14–58), he meant to be both calling attention to the creative role that we play in making our lives what they are, and observing that within the bounds of the given facts of our lives, there is a great deal of discretionary leeway in how we *understand* our lives. We each have to sift through the raw materials of our own experience and cull from them an understanding of what has happened to us and who we are. And this observation extends beyond our narrow understanding of our own lives to our understanding of the people around us and the world quite generally. We are always deciding what to think and how to feel about things. This is not a passive matter of simply opening our eyes, but a complex interpretive task that requires imagination and discernment. It requires imagination because it requires to us to be able to see the interpretive possibilities. The exercise of the interpretive imagination is something at which art, history, literature, and autobiography all excel. The right portrait can make the seedy seem romantic, the innocent seem menacing, and the sublime seem ridiculous. It can exalt the debased and deflate the exalted. Tom Waits’s descriptions of waitresses in cheap diners, Knut Hamsun’s scathing portraits of intellectuals, Toulouse Lautrec’s romantic visions of Paris brothels, and Degas’s gorgeous portraits of the ballet have a revelatory character, making us see their objects differently. Interpretation isn’t a matter of getting a complete or detached view of things. It doesn’t strive for the kind of forensic accuracy that is prized in scientific representation. It is a matter of selecting and suppressing, foregrounding and enhancing.<sup>25</sup> It is designed to bring out one particular pattern, suggest an evaluation, and often to encourage us to feel a certain way. New ways of seeing can

---

25. Science is also very interested capturing patterns, but it tends to be interested in patterns that reveal regularity. It is interested in laws, and in induction, so it tries to isolate the shared and generalizable elements in nature.

transform how we experience the world, making us see familiar objects in a novel way and discover value where we hadn't seen it before.<sup>26</sup>

It demands discernment, because entertaining interpretive possibilities is only part of the task. We also have to decide what to think. When we have a fight with a spouse or feel affronted by an interaction with a colleague, we need to sort through what actually happened in our minds and arrive at some interpretation. In doing that, we are not trying to settle the simple narration of events (what he said, what I said, and in what order); we are trying to understand what happened in an evaluatively rich sense. We are trying to understand whether we have been wronged or are in the wrong, whether we should be hurt or apologetic, or whether we should be insulted or indifferent. On one interpretation of the fight with the spouse, I came home after a very difficult day, he attacked me for no reason, I responded defensively, and instead of staying and sorting it out, he left the house, leaving me alone hurting and bewildered. On another interpretation, I walked in the door that night already wounded and needy. When he made a careless remark, I lashed out violently, and he left the house only to keep from angering me further. Getting this right matters in obvious ways, and it is not a simple task. It demands a disciplined and willful effort to see things from other people's point of view, an unwillingness to acquiesce in self-serving interpretations, and the resolve to be fair and honest in our assessment. We portray others in a manner that is flattering to ourselves, and often misdiagnose the sources of our emotions. Iris Murdoch has done more, perhaps, than anyone to describe the moral rigors of what I've called "getting it right." Her most famous example concerns a mother-in-law who undergoes a transformation in her view of her daughter-in-law, and it is worth quoting in full.

A mother, whom I shall call M, feels hostility to her daughter-in-law, whom I shall call D. M finds D quite a good-hearted girl, but while not exactly common yet certainly unpolished and lacking in dignity and refinement. D is inclined to be pert and familiar, insufficiently ceremonious, brusque, sometimes positively rude, always tiresomely juvenile. M does not like D's accent or the way D dresses. M feels that her son

---

26. It is tempting to say that interpretation is about evaluation rather than fact, but this is to suppose a separation between fact and evaluation that is not psychologically faithful. The most basic descriptive terms that we use to represent a situation—'seedy' vs. 'romantic', 'innocent' vs. 'menacing', 'sublime' vs. 'ridiculous'—don't have a clearly identifiable, shared factual core.

has married beneath him. Let us assume for purposes of the example that the mother, who is a very “correct” person, behaves beautifully to the girl throughout, not allowing her real opinion to appear in any way. . . . [T]ime passes, and it could be that M settles down with a hardened sense of grievance and a fixed picture of D, imprisoned . . . by the cliché: my poor son has married a silly vulgar girl. However, the M of the example is an intelligent and well-intentioned person, capable of self-criticism, capable of giving careful and just attention to an object which confronts her. M tells herself: “I am old-fashioned and conventional. I may be prejudiced and narrow-minded. I may be snobbish. I am certainly jealous. Let me look again.” Here I assume that M observes D or at least reflects deliberately about D, until gradually her vision of D alters . . . D is discovered to be not vulgar but refreshingly simple, not undignified but spontaneous, not noisy but gay, not tiresomely juvenile but delightfully youthful, and so on. (Murdoch 1970, 16–17)

We are made to understand from Murdoch’s discussion that the altered interpretation of D comes closer to getting it right, and that getting it right is an epistemic matter, though not one that is simply a matter of looking. I think that we all understand what she is pointing to here. We have to interpret people and events all the time, and suffer the interpretations of others. Sometimes (as M’s case) getting things right is a matter of being kinder and gentler in our vision of others, but sometimes it is a matter of seeing them under the cold, harsh light of sober assessment. It is not easy to recognize that your lover is an overconfident bore, or that your child is a cruel bully.

Getting things right doesn’t come easily. It brings with it a kind of truth and rigor that are quite different from the sort that one finds in the sciences. Murdoch said that it consists of “a refined and honest perception of what is really the case, a patient and just discernment and exploration of what confronts one, which is the result not simply of opening one’s eyes but of a certain and perfectly familiar kind of moral discipline” (Murdoch 1999, 330). Humanistic disciplines—literature, history, and art, most obviously, but also philosophy, anthropology, and languages—engage the interpretive muscles. In so doing, they help us see more deeply into ourselves and others, and cultivate the kind of understanding that helps us get it right in our own lives. There is no monolithic account of what the humanities are and what they do. These are just a couple of examples of the myriad complex ways in which the humanities help us make sense of ourselves and the world in which we live.

One of the complaints or frustrations that people who gravitate to science or math sometimes make about the humanities is the squishiness of the subject matter. They complain that everything is qualitative, impressionistic, and a matter of discretion or judgment. They say that there is no proof, no certainty, and no truth. It is correct that there is no proof and no certainty, but that does not mean that there is no truth. The kind of truth ~~they~~ strive for is softer than the kind of truth we have in math or in the sciences. One has to get comfortable with ambiguity and squishiness and the lack of full resolution—i.e., with interpretation rather than calculation—to operate in that environment. But it resembles in that respect the messy world of human affairs. Not everybody needs to be a writer, artist, or historian to make some knowledge of literature, art, and history valuable. Reading Dostoyevsky, Thomas Mann, Milton, Joyce, Herodotus, Omar Khayyam, Plato, and Charlemagne teaches ~~us all kinds of~~ things that will enhance our perception and enrich our experience of the world. It will deepen our understanding of ourselves and other people, ~~teach~~ us how to live, how to love, and how to feel. That is enough to make it a valued part of ~~the academy~~.

The upshot of all of this is that sciences and humanities don't compete or conflict. They are entirely complementary, answering to different needs. The humanities provide a type of ~~understanding~~ that is both essential to human living and not readily attainable from science. The everyday notion of understanding makes room for both. Of course, that leaves open the vexed question about what universities are for. That is a question that we need to address as a culture. We do need to prepare a workforce and produce researchers who will help us cure cancer and save the earth. But education does not have to be *just* about that. It can also be about helping us make better choices, getting us to care about the right things, and opening up the imagination to new ways of thinking and being. The academy should be a place where all of these things can happen. If we ask ourselves what we would like to pass on to our children, I think that many of us would say that, alongside hoping that they learn a trade or a job that will support them, we hope that they learn how to be imaginative and resourceful and fair, that they become the kind of human beings that know how to love and that approach other people with subtlety, perceptiveness, and understanding. These are valuable qualities within the workplace and without.

## 9.6. Conclusion

It has taken time for science to mature so that we can see the importance of the humanities as emerging from within the scientific conception of the human



being. If there was a time when the sciences and the humanities seemed to offer competing visions of the human being, that time is past. It is now possible to say on scientific grounds what is wrong with the idea that the sciences will ever replace (or displace) the humanities. The structure of human knowledge is complex, but it forms a single fabric, in which the humanities have their place alongside the sciences, and which every part makes a contribution to understanding of ourselves and our place in the universe.

## Acknowledgments

This work was supported by the generous support of the Templeton Foundation. I owe a large debt of gratitude to Stephen Grimm. The chapter would not have been written without him, and his insightful comments made it much better than it would have been otherwise. I would also like to thank Elisabeth Camp and Barbara Tversky for very helpful comments and the audience at the “Varieties of Understanding” conference, Fordham University, June 2015.

## References

- Godfrey-Smith, Peter. 2004. “On Folk Psychology and Mental Representation.” In *Representation in Mind: New Approaches to Mental Representation*. Ed. Hugh Clapin, Phillip Staines, and Peter Slezak. Amsterdam: Elsevier. 147–62.
- Hepburn, Ronald W., and Iris Murdoch. 1956. “Vision and Choice in Morality.” *Proceedings of the Aristotelian Society, Supplementary Volumes* 30: 14–58.
- Ismael, Jenann. 2016. *How Physics Makes Us Free*. New York: Oxford University Press.
- Kitcher, Philip. 2012. “The Trouble with Scientism.” *New Republic*, 24 May. <http://www.newrepublic.com/article/books-and-arts/magazine/103086/scientism-humanities-knowledge-theory-everything-arts-science>.
- Murdoch, Iris. 1970. *The Sovereignty of Good*. New York: Routledge.
- . 1999. *Existentialists and Mystics: Writings on Philosophy and Literature*. New York: Penguin.
- Nussbaum, Martha. 2010. *Not for Profit: Why Democracy Needs the Humanities*. Princeton, NJ: Princeton University Press.
- Rand, Ayn. 1999. *Why Businessmen Need Philosophy*. New York: Ayn Rand Institute.
- Rosenberg, Alex. 2011. *Atheist’s Guide to Reality*. New York: Norton.
- Zakaria, Fareed. 2015. “Why America’s Obsession with STEM Education Is Dangerous.” *Washington Post*, 26 March. [https://www.washingtonpost.com/opinions/why-stem-wont-make-us-successful/2015/03/26/5f4604f2-d2a5-11e4-ab77-9646eea6a4c7\\_story.html](https://www.washingtonpost.com/opinions/why-stem-wont-make-us-successful/2015/03/26/5f4604f2-d2a5-11e4-ab77-9646eea6a4c7_story.html).

## 10

---

# Understanding and Coming to Understand

*Michael Patrick Lynch*

## 10.1. Introduction

Suppose I ask you how to get to Larissa, and you give me the right answer. Suppose further your answer is not a guess; you have some grounds for it. There are lots of different ways that could happen. For example, you might:

- Look it up on your phone
- Remember how you got there last year
- Do both of these things but also explain why certain routes that look good on the map are actually impossible or difficult because of the geography and road conditions

As I see it, all three of these actions might result in conditions that could ground your knowledge of how to get to Larissa. Such conditions represent three different ways our opinions can be grounded, by being based on:

- Reliable sources
- Experience or reasons that we possess
- A grasp of the causal relations between local conditions and the feasibility of local travel routes

If, like me, you are tolerant of a sensible pluralism about knowledge, you'll be comfortable saying that these different kinds of grounding give rise to different

kinds of knowing.<sup>1</sup> The first sort of knowing is the sort we engage in when we absorb information from expert textbooks or good Internet resources. The second is the sort of knowing that occurs whenever possessing reasons or experience matters. And the third is different still—it is the sort of knowing we expect from experts—even if those experts are more intuitive than discursive in their abilities. This is what I’ll call understanding. Understanding in this sense is what we have when we know not only the “what” but the “how” or the “why” (see Kvanvig 2003; Grimm 2006). Understanding is what the scientist is after when trying to find out the root causes of Ebola outbreaks (not just predict how the disease spreads). It is what you are after when you want to know *why* your friend is so often depressed (as opposed to knowing that she is).<sup>2</sup>

Many philosophers take understanding to be a distinctive kind of knowledge that is particularly valuable. The kind of knowledge in question concerns grasping dependency relations. This chapter aims to investigate and address two well-known puzzles that arise from this conception. The first concerns the nature of understanding itself—in particular, the nature of the “grasping” relationship that understanding is thought to involve. The second concerns the source of understanding’s distinctive value. In what follows, I’ll argue that we can shed light on both puzzles by recognizing, first, the importance of the distinction between the act of coming to understand and the state of understanding; and, second, that coming to understand is a creative act.

## 10.2. Understanding: The Functional Role

“Understanding,” like “perceiving,” displays a typical state/act ambiguity. Taken in the first sense, it refers to a particular kind of epistemically valuable intentional cognitive state.<sup>3</sup> While the details differ, most views of understanding agree on certain common features they take the state to have. These

1. Ernest Sosa is the leading pluralist about knowledge in this sense. See Sosa (2010).

2. Philosophical accounts of the state of understanding often differ over whether to take understanding as a form of knowledge or not. This is an important difference, although how important may depend on one’s account of knowledge; pluralist accounts, like the one I favor, are willing to take “knowledge” as multiply realizable. While I will continue to take understanding as a form of knowledge in what follows, the substantive contributions of this view are consistent with holding that the concepts are more distinct. For discussion, see Zagzebski (2001), Kvanvig (2003), Grimm (2006).

3. The state I go on to describe is degree-theoretic. One can understand more or less.

commonly cited properties of understanding can be used to help fix the reference of the “state” use of the term, and help us get clear on what we are talking about.

First, the state obviously is meant to be capable of having positive epistemic status—moreover, it can convey such status on other states. The state of understanding has probative force.

Second, it is directed at *how* or *why* something is the case. Thus, you understand more about the civil rights movement if you understand why and how it came about; you understand string theory if you understand why it predicts certain events; you understand a person to the extent you don’t just know that she is unhappy, but what makes her unhappy.

We can take it that what is common between understanding *how* and understanding *why* is that we know something about the *structure* of the whole (see Grimm 2006; 2011). This sounds grand, and it can be, as when we understand how a proof works or why a great historical event occurred. But it can also happen on a smaller scale. Consider, for example, the lucky person who understands how her car works. She has this understanding in part, as we’ll discuss more fully in a moment, because she has certain skills, skills that give her the ability to see how various parts of a machine depend on one another: you can’t get the car to move without the battery and the battery won’t be charged without the alternator. You understand when you see not just the isolated bits, but how those bits hang together. Similarly with understanding why something is the case. When we understand why something is the case, such as why a certain disease spreads or why your friend is unhappy, or why a given apple tree produces good apples, we do so because we *grasp* various relationships. These relationships are what allow us to see the difference between possibilities, between one hypothesis and another.

If this is so, we might say that a third common thought about understanding is that understanding why or how is the result of *grasping* actual *dependency* relations, not just correlations. An instructive example is Plato’s Euthyphronic contrast:

$x$  is holy when, and only when,  $x$  is loved by the gods.

Instances of this schema will be universally true. They might be true in all possible worlds. But simply grasping the instances doesn’t add up to understanding why what is holy is loved by the gods, or how holiness and the will of the gods are metaphysically situated in terms of one another. Therefore, it

doesn't add up to understanding the nature of holiness. To truly understand, you also need to know what *depends* on what.

The dependency relations we grasp when we understand can come in different forms. Some relations might be about cause and effect. Think of a game of chess: if I move my bishop to a certain square, I cause it to change its position. But they might also be logical: if I move my bishop to this square, it will be vulnerable to your pawn. Or semantic: the bishop can move to that square because the rules define it as being able to move diagonally across the board. In other words, the first important element of understanding a game like chess is grasping dependency relations: having systematic knowledge of how things both fit together and depend on one another, causally, logically, and otherwise.<sup>4</sup> Of course, knowledge of certain kinds of dependency relations might be particularly relevant to certain kinds of understanding. It seems plausible that scientific understanding, for example, gives pride of place to knowledge of causal relations.

Understanding, seen as stemming from the grasp of dependency relations, is consistent with, if it does not entail, holding that understanding can be directed at both theories and persons. Understanding a theory, on this view, would involve understanding the dependency relationships between the principles and theorems that constitute the theory. Likewise, understanding a person would amount to understanding the relationship between their cognitive and emotional states and their behavior in certain contexts. In this sense, understanding a person amounts to grasping what "makes them tick," as we say.<sup>5</sup>

The state of understanding is also seen as being related, directly or indirectly, to other cognitive states and an agent's behavior. Thus the idea that understanding is the grasping of dependency relations supports the idea that understanding, as numerous commentators have noted, is tied to explanation. On some views, understanding of certain kinds involves having the grasp of a *correct* explanation, or at least having the potential for such a grasp (see Strevens 2013). But even if one does not take the (potential) to supply a

4. This is a broadly Aristotelian account of understanding. See Greco (2014). See also Grimm (2006). (Not everyone sees understanding as involving knowledge. See Zagzebski 2001.)

5. One complicating factor is whether, as Stephen Grimm has noted, one needs to also take a person's beliefs and desires to be intelligible in order to understand her. Here, I think it is relevant to remember that understanding is a matter of degree. I understand you to some degree if I understand why you do what you do. I understand you more if I understand why you do what you do and find your beliefs and desires intelligible. See Grimm (2016).

correct explanation as a necessary condition for being in the state of understanding, it is plausible, at the very least, that understanding is conducive of good explanations.<sup>6</sup>

A related point is that the agent who understands thereby has certain abilities. The Oracle of Delphi supposedly announced that no one was wiser than Socrates. He famously replied that he only knew that he knew very little, or what he didn't know. So what sort of knowledge did he have? Well, consider what he was truly good at. One thing, surely, was asking questions. This came from a combination of knowing facts and the ability to draw connections between them. As a result he had *know-which*, as it were. He knew *which* questions to ask.

This is suggestive. The person who understands is, to some degree, discerning not only the actual situation, but also why various hypotheses and explanations *won't* work as well as how to ask what would (again, see Strevens 2013). They know that kicking the refrigerator *here* and not *there* will help get it working. This is something that experts in general can do. Indeed, experts—those who understand a given subject best—are often able to increase their understanding even further because they have the ability to know which question they should ask in the face of new information. By so doing, they can, for example, reveal that Euthyphro knows nothing of piety.

Arguably, however, the skill of being able to ask good questions itself hinges, at least in part, on a simpler (and less overtly verbally orientated) cognitive capacity: the ability to make inferences and draw out a position's consequences—and not just the actual consequences of, say, a given position on what causes apples to be tasty, but also the consequences of that position in certain counterfactual situations. This is precisely the skill that a good doctor employs when considering whether to administer a drug, or a lawyer uses when considering an argument. It is also, arguably, the skill a good mechanic employs when considering whether to disassemble a head gasket, or an apple farmer uses when deciding whether another farmer's advice is reasonable. And those who have the capacity to cognitively engage, should they have the requisite verbal and linguistic abilities, will know which questions they should ask in order to carry their inquiries even further.

This list of commonly cited characteristics of understanding is hardly exhaustive. But it can be used to give a partial functional characterization of

---

6. Thus understanding need not be factive, although the deeper it becomes, the more it will approach factivity. To understand perfectly, perhaps, *is* factive (For further discussion see Elgin 2009; see also Zagzebski 2001.)

the state of understanding. Even a partial functionalist characterization can be used to fix the reference of the term. One suggestion might be this:

U: A state of some agent plays the understanding-role with regard to some subject when its content concerns dependency relations between propositions or states of affairs relevant to the subject; it is conducive of the agent's ability to offer justified explanations of the relevant subject;<sup>7</sup> and it disposes the agent to make further justified inferences both factual and counterfactual about the subject.

Such a state would presumably have positive epistemic status and probative force.

Like any functional description, this one still leaves much to be said. In particular, it leaves open the underlying psychological nature of the state or states that can play the role, and the value those states may or may not have when playing that role. Moreover, as we shall see, there is more to be said about the etiology of understanding—the distinctive causal antecedents of states playing the understanding-role.

### 10.3. Grasping and Coming to Understand

So what is the state that plays the *u*-role? A natural suggestion, given what we've said so far, might be a distinctive cognitive attitude we've called "grasping." While I think this suggestion is intuitive, and has something to be said for it, I don't think it is productive, for two reasons.

First, and as Stephen Grimm has noted, the psychology of such relations is difficult to parse—especially when we take it as constitutive of a stable state (Grimm 2011).

One reason for that is that the root metaphor at work in "grasping" is obviously active. Grasping is something we do, and insofar as we think of it as having a distinct phenomenological character, it is a cognitive act, available to conscious attention. In contrast, understanding, taken as a state of mind, seems (like belief) dispositional. Just as one might be in a state of decision (or indecision, as the case may be) without doing anything in particular,

7. This clause is intended to be neutral with regard whether agents have such an ability in the first place. If they do have such ability, the state playing the understanding-role will be conducive of that agent manifesting it. Children, for example, might lack that ability in certain contexts while still possessing some understanding. Thanks to Stephen Grimm for discussion.

or even being conscious of being in that state, we can understand why something is the case without being consciously aware of that fact, and without the understanding being available for conscious attention. Your understanding in such a case is tacit or implicit. Arguably, much of what we understand we understand in this way, and we can forget that we understand something that we do understand.

As noted above, understanding is also thought to have distinctive value. A second problem with the suggestion that grasping is what plays the u-role, is that it doesn't particularly help answer this question.

One might think at first that the reverse is the case. The thought would be that understanding is valuable because grasping dependency relations is valuable. That seems true, but it doesn't say what is *distinctively* valuable about understanding. If cognitive contact with dependency relations is valuable, then it seems possible that we might have epistemically positive cognitive contact with such relations in other ways. That is, we might know about the relations without understanding them. If so, what constitutes the *distinctive* value of the grasping of those relations? One tactic is to argue that the value of grasping rests in the fact that understanding is "active" and not passive: as a result understanding, unlike other epistemic states, is a cognitive achievement (see Pritchard 2008). That is plausible. But it raises some questions all on its own.<sup>8</sup> First, we might wonder what type of achievement understanding is. Second, if being in the state of understanding is an achievement, then being in that state must itself be the result of an act—the act of *achieving* understanding. And generally, when we talk about achievements, we think that part of what makes the achievement admirable is that the agent voluntarily did something to achieve the goal. Achievements are partly valuable because, well, they were . . . *achieved*.

In sum, I think that it is implausible that grasping is what plays the u-role. On the one hand, it seems to overcomplicate matters: it posits an active process to play the role of what is plausibly a dispositional or implicit state. Second, it underexplains: it doesn't by itself, at least, explain why understanding is held to have distinctive value over and above other kinds of knowledge.

That said, I think it clear that grasping should figure in any account of what understanding is. The question is where.

---

8. Similar problems arise with the possibility that understanding is valuable because knowledge is valuable, and understanding either is a type of knowledge or leads to knowledge. That just seems less than illuminating, since it doesn't actually tell us what is distinctively valuable about understanding itself and/or makes that value instrumental to another.



I suggest we can make a promising start on this project by paying attention to the other half of the “understanding” state/act ambiguity. More precisely, I think it is worth thinking about the process of *coming to understand*. This should shed light on understanding (the state) itself. That’s because like other kinds of mental states, the state’s etiology, or typical causal antecedents, help to pick out its functional role. Think of the state of believing that certain things are in your visual field. Being in that state is the causal result of having visual experience. Likewise, to understand, one must first come to understand. My hypothesis is that an analysis of the cognitive act of coming to understand can help shed light on why we are tempted to say that understanding (the state) is a cognitive achievement, and why it involves an activity like grasping.

In the view we’ll entertain here, grasping is constitutive not so much of the state of understanding, *but of the causally prior act of coming to understand; moreover, this prior act is partly definitive of the state it produces.*

If this is right, then the state of understanding is distinctive in part because of its etiology; one comes to be in that state only in virtue of having been caused to do so by first engaging in an active psychological process which is available to conscious attention.<sup>9</sup> But what is it to come to understand?

Coming to understand is a mental act in the same way that reflecting or deciding are mental acts. They are activities that your mind engages in. They take effort and increase the total cognitive load. A full description of the act of grasping is of course the job of empirical psychology and cognitive science; but prior theoretical reflection sharpens, here as elsewhere, our empirical inquiries. The sharpening I suggest is this: coming to understand, and therefore the grasping that helps to constitute it, is a creative act.

In order to begin to see why this is plausible, and how it sheds light on understanding itself and its value, consider a (probably apocryphal) story about Descartes. Descartes was a late riser. His habit, when possible, was to stay in bed till around noon—musing. One day, according to legend, he was watching a fly zoom around above his head when, suddenly, he realized that he could track its position by measuring its distance from the walls and the ceiling. He understood how to plot its flight path in space . . . and voilà! We get Cartesian coordinates, or so the story goes.

The story of Descartes’s fly—and others like it, such as those about Newton’s apple or Einstein’s clock—are instructive because they emphasize that the

---

9. Something can be available to conscious attention without, at that moment, being attended to.

moment of coming to understand can involve sudden insight. Such moments are often called “aha moments” and, in the psychological literature, are collectively taken to signify the “Eureka effect” (so named after Archimedes, who after a moment of great insight shouted “Eureka!”). Of course, most acts of understanding do not require the sudden novel inspiration that Descartes had. But all of them do involve some level of insight. Having such an insight is part of why understanding is fundamentally a creative act.

Creativity, or creative acts, are complex. They are marked by having a combination of characteristics, characteristics that other acts can have but which taken together help to distinguish creative acts from other things humans do. One subset of creative acts—a particularly important subset—are creative mental or psychological acts. It is their characteristics I discuss here.

First, creative mental acts are new or novel. As Margaret Boden as famously emphasized, creative ideas needn't be *historically* novel—like Descartes's new geometrical ideas—but they are *psychologically* novel or novel to the creator (see Boden 2004). Thus, being creative isn't the same as being original. People can have ideas that are creative *for them*. As Boden says, “Suppose a twelve-year-old girl, who'd never read *Macbeth*, compared the healing power of sleep with someone knitting up a raveled sleeve. Would you refuse to say she was creative, just because the Bard said it first?” (2004, 2). I don't think so, and neither does Boden. Creativity is relative to a person.

Second, creative mental acts are generative of valuable, not just psychologically novel ideas. Creative ideas are valuable to the person's cognitive workspace. They move things forward on the conceptual field on which they are currently playing. They are useful and fecund. They have progeny, and they contribute to the problems at hand.

Third, creative mental acts are typically the result of the cognitive effort distinctive of synthetic imagination. The psychological act involved in composing a song requires the ability to put together a complex string of different ideas about harmony, melody, and rhythm; the process of creating a coherent and believable fictional character likewise involves psychologically combining ideas of personality and physical description. That is, we think that creative mental acts often put things together in new ways.

So far, coming to understand seems to fit this model of creative psychological acts: it involves generating in a synthetically imaginative way new and valuable ideas. Which ideas? Those that concern dependency relationships—how things fit together. The “grasping” of those relationships, which lies at the heart of understanding, is what makes understanding creative. This is most obvious in paradigmatic, historic cases of new understanding, like Descartes's

insight into how point location in Euclidean space can be plotted algebraically or Einstein's flash of understanding relativity upon seeing a clock. But what about less historically original acts of understanding? Consider again a child who comes to understand, for the first time, why 0.150 is smaller than 0.5. At that moment, the child is also having an insight—a realization of how things are related. Or consider again our student above, coming to understand for the first time why Lady Macbeth sees blood on her hands, or why sailing is more pleasant and efficient when the wind is not behind you. Each of these acts of coming to understand are creative insights for the person in question, even though they are in no way novel.

The three characteristics of creative mental acts we've canvassed so far are not exhaustive; there is a fourth—one emphasized by Boden and more recently by Nanay (2014). Creative mental acts have a distinctive phenomenological feel. The phenomenology in question might be described as something akin to surprise. Boden calls this their "impossible" aspect—that is, an idea is creative for a person when she has a felt sense that the idea could not have been had prior to the moment of creation. Conditions were right, and the person suddenly "sees."

It seems clear that the act of grasping/coming to understand shares this characteristic as well. Indeed, it is particularly striking in this case. Coming to understand has a particular phenomenological appearance. In cases of sudden insight, this phenomenological aspect of creativity either constitutes, or leads to, the "eureka" feeling. But creative acts can be surprising even if they do not necessarily provoke that "aha" moment. Consider coming to slowly understand, for example, why a particular theorem followed from a particular set of premises—you understood, as we say, the proof. Even when coming to understand happens gradually over time it still feels "new"—as if you couldn't have understood it prior to that moment. It feels as if you've made forward progress. That's why it makes sense to say that the act of coming to understand is also surprising—again, not necessarily in the "eureka" sense—because the person who comes to understand feels as if they could not, relative to their past evidence and cognitive context, have understood it before that moment.

It might be thought that not all acts of coming to understand can be creative in this sense. Surely, one might think, coming to understand the simple logical entailments of what else I understand cannot be a creative. In many cases, those entailments, and seeing why they follow, is too trivial to count as truly creative.

In some cases, coming to know what follows from a proposition you understand, and why it follows, is not at all trivial. It requires great effort

and insight. In those cases you do come to understand the entailment. But in other cases such knowledge may well be trivial, and nothing in the account rules that out. That's because coming to know a given entailment may not be an act of *coming to understand*. Understanding is a matter of degree, and our account of the state implies that the greater one understands, the more one is able to draw the relevant inferences. Part of drawing the relevant inferences is knowing why certain entailments hold from what you do understand. Thus, if one understands  $\langle p \rangle$  to a sufficient degree one will, *just by virtue of being in that state*, know why certain entailments hold. That's consistent with saying that while coming to understand  $\langle p \rangle$  might be creative, one doesn't *separately* come to understand (in the target sense)  $\langle q \rangle$ . That's because if one understands  $\langle p \rangle$  sufficiently, one is by virtue of that fact already disposed to know both  $\langle \text{if } p \text{ then } q \rangle$  and  $\langle q \rangle$ .<sup>10</sup>

Some might protest that this account of creative mental acts is too permissive. According to this line of thought, originating a new proof is creative. You are the first person that comes up with it. But simply coming to understand why the proof works isn't creative.

This objection confuses ways something can be creative. A novel discovery or origination of a new proof is undoubtedly creative. Call this special type of creativity, which is very rare, o-creativity (for "original"). But as I pointed out above, not all creative acts are o-creative. (Consider, for example, the fact that someone might originate or discover a proof, the proof be forgotten for a thousand years, and then someone else might originate or discover it again.) Thus your act of first originating a proof might be o-creative, but my act of coming to understand it for myself might be creative for me. And that's creativity enough.

Finally, it is worth emphasizing that the surprising or "impossible" aspect of creativity makes creating seem at once something we do (which it is) and at the same time something happening to us. The muse suddenly strikes. Realization comes in a flash. Coming to understand is like this as well. It involves insight, and insight, as the very word suggests, is like the voluntary opening of a door, a "disclosing," as Heidegger said. One acts by opening the door, and then one is acted upon by seeing what lies beyond.

---

10. In addition, it bears noting that even if one understands  $\langle p \rangle$ , knows  $\langle \text{if } p \text{ then } q \rangle$ , and on that basis, comes to know  $\langle q \rangle$ , that fact alone doesn't mean you come to understand  $\langle q \rangle$  since you may not understand  $\langle \text{if } p \text{ then } q \rangle$ .

## 10.4. The Value of Understanding

We can summarize the hypotheses floated in the last section as follows: To understand requires first coming to understand. And coming to understand involves actively grasping certain dependency relations, where grasping is a conative state of mind (both directed and active) with features associated with a creative mental act. In particular, the act of coming to understand/grasping is creative for a person to the extent that it generates ideas that are, for that person:

- Novel
- Valuable
- Resultant from synthetic imagination
- Distinctive in their phenomenology: their “surprising” or “impossible” aspect.

Obviously, this description of the mental act of coming to understand/grasping doesn't tell us everything about its nature. But it does help us see what is distinctive and special about the act of grasping, and as a result, can help to guide further investigation. When we “look for” grasping in our psychological theorizing and experimentation, the present suggestion is that we look for a mental act with these characteristics.

The above suggestion also allows us to adopt a more straightforward answer to the question of what occupies the u-role—that is, what kind of mental state is at play when we are in the state of understanding. The straightforward answer is that we are in a state of belief whose properties (both epistemic and psychological) allow it to play the understanding role. To understand is to believe in a certain way. But part of what it is to believe in that way is *to arrive at* the belief in a distinctive manner, to come to understand by grasping the relevant dependency relations. If so, then we can revise our functional analysis as follows:

U\*: A belief (of some agent) plays the understanding-role with regard to some subject when the agent has been caused to be in that state by grasping dependency relations between propositions or states of affairs relevant to the subject; the ensuing belief is about those relations; having the belief is conducive of the agent's ability to offer justified explanations of the relevant subject; and it disposes the agent to make further justified inferences both factual and counterfactual about the subject.

The above analysis is consistent, of course, with some kind of psychological state other than belief playing the understanding role in some agents. But it seems likely that it is a kind of belief that most often realizes the role in human beings.

I began by suggesting that a focus on the act of coming to understand can shed light on two puzzles about understanding. The first puzzle was metaphysical, and concerned its nature. The above analysis provides an overall lesson. Like many other targets of psychological and epistemic analysis, understanding is distinguished by (1) how we come to be in the state; (2) its properties and content; (3) its effects on our behavior and dispositions to behave. Nonetheless, understanding is still a deeply interesting and important state of mind, one that our analysis predicts as requiring cognitive effort to achieve, the result of the act of grasping how things hang together.

The second puzzle about understanding our suggestion may help to solve concerns its value. But before dealing with this question directly, it is illuminating to look at another long-standing issue about understanding: its relation to testimony.

Understanding is often said to be different from other forms of knowledge precisely because it is not *directly* conveyed by testimony—and thus not directly teachable (see Zagzebski 1999; 2001). The thought is that you can give someone the *basis* for understanding via testimony, including the knowledge that they must have in order to achieve that understanding. But in the usual cases, you can't directly convey the understanding itself.

The question is *why* this should be. What is it about understanding that makes it difficult or impossible to convey via testimony? The present suggestion supplies an answer: in order to first understand, one must come to understand. And coming to understand is a creative act. As such, it requires a cognitive, generative psychological action on the part of the agent over and above whatever knowledge might be conveyed by another. An art teacher, for example, can give me the basis for creative thought by teaching me the rudiments of painting. She can give me ideas of what to paint and how to paint it. But I did not create these ideas; I create when I move beyond imitating to interpret these ideas in my own way. Likewise, you can give me a theorem without my understanding why it is true. And if I do come to understand why it is true, I do so because I've expended some effort—I've drawn the right logical connections. Coming to understand is something you must do for yourself.

Let's contrast this with other kinds of knowledge. I can download ordinary factual knowledge directly from you. You tell me that whales are mammals;

I believe it, and if you are a reliable source and the proposition in question is true, I know in the receptive way. No effort needed. Or consider responsible belief: you give me some evidence for whales being mammals. You tell me that leading scientists believe it. If the evidence is good, then if I believe it, I'm doing so responsibly. But in neither case do I thereby directly understand why whales are or aren't mammals. You can, of course, give me the explanation (assuming you have it). But to understand it, I must first grasp it myself. Understanding can't be outsourced.

Earlier we noted that while it is intuitive that understanding has distinctive, perhaps intrinsic, value, it is unclear what the basis of that value happens to be. Our reflections on why understanding isn't conveyed by testimony rely on the fact that understanding is partly defined by its etiology; to be in the state, one must first come to understand. This same fact helps to explain the distinctive value of understanding as well.

Earlier we noted that we typically take understanding to be a cognitive achievement, and that fact is part of the explanation for why we think it is valuable. But seeing understanding as an achievement, we noted, means that the state must be something we do out of an act of will. It has to be, as it were, achieved. The present account dovetails with, and explains, this fact. We achieve understanding because we first come to understand—an act that requires effort.

Moreover, coming to understand is a creative act. And the creativity of that act helps to explain our intuitive sense that understanding is a cognitive state of supreme value and importance, not just for where it gets us but in itself. Creativity matters to human beings. That's partly because the creative problem-solver is more apt to survive, or at least to get what she wants. But we also value it as an end. It is something we care about for its own sake. And that goes for coming to understand as well. It is an expression of one of the deepest parts of our humanity.

## Acknowledgments

Thanks to John Greco, Duncan Pritchard, and Nathan Sheff for comments and discussion of this chapter, and to audiences at the University of Edinburgh, St. Louis University, Yale University, and Henderson State University.

## References

Boden, Margaret A. 2004. *The Creative Mind: Myths and Mechanisms*. New York: Routledge.

- Elgin, Catherine. 2009. "Is Understanding Factive?" In *Epistemic Value*. Ed. Adrian Haddock, Alan Millar, and Duncan Pritchard. Oxford: Oxford University Press. 322–30.
- Greco, John. 2014. "Episteme: Knowledge and Understanding." In *Virtues and Their Vices*. Ed. Kevin Timpe and Craig A. Boyd. Oxford: Oxford University Press. 285–301.
- Grimm, Stephen R. 2006. "Is Understanding a Species of Knowledge?" *British Journal for the Philosophy of Science* 57: 515–35.
- . 2011. "Understanding." In *The Routledge Companion to Epistemology*. Ed. Sven Berneker and Duncan Pritchard. New York: Routledge. 84–94.
- . 2016. "How Understanding People Differs from Understanding the Natural World." *Philosophical Issues* 26.1: 209–25.
- Kvanvig, Jonathan L. 2003. *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press.
- Nanay, Bence. 2014. "An Experiential Account of Creativity." In *The Philosophy of Creativity: New Essays*. Ed. Elliot Samuel Paul and Scott Barry Kaufman. Oxford: Oxford University Press. 17–35.
- Pritchard, Duncan. 2008. "Knowing the Answer, Understanding and Epistemic Value." *Grazer Philosophische Studien* 77: 325–39.
- Sosa, Ernest. 2010. *Knowing Full Well*. Princeton, NJ: Princeton University Press.
- Strevens, Michael. 2013. "No Understanding without Explanation." *Studies in History and Philosophy of Science Part A* 44.3: 510–15.
- Zagzebski, Linda. 1999. "What Is Knowledge?" In *The Blackwell Guide to Epistemology*. Ed. John Greco and Ernest Sosa. Oxford: Blackwell. 92–116.
- . 2001. "Recovering Understanding." *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. Ed. Matthias Steup. Oxford: Oxford University Press. 235–51.



## Against Understanding (as a Condition on Explanation)

*Bradford Skow*

### 11.1. Understanding as a Condition on Explanation?

It is often said that there is an important relationship between explanation and understanding. Like so many ideas about explanation, this one makes an appearance in Carl Hempel's "Aspects of Scientific Explanation" (1965), the founding document of contemporary thinking about the topic. In the second section of that essay Hempel introduces the Deductive-Nomological Model of explanation, according to which (I'll risk the reminder) an explanation of a fact  $X$  is a sound argument for  $X$  that essentially contains a law-stating premise. Hempel then writes:

[A] D-N explanation answers the question "*Why* did the explanandum-phenomenon occur?" by showing that the phenomenon resulted from certain particular circumstances, specified in  $C_1, C_2, \dots, C_k$ , in accordance with the laws  $L_1, L_2, \dots, L_r$ . By pointing this out, the argument shows that, given the particular circumstances and the laws in question, the occurrence of the phenomenon *was to be expected*; and it is in this sense that the explanation enables us to *understand why* the phenomenon occurred. (337)

What is going on in this passage? In particular, why did Hempel think it important to say the things he says in that last sentence? One natural

interpretation has Hempel assuming, in the background, a necessary condition on explanation:

**The Explanation-Understanding Condition:** Something E is an explanation of some fact F only if someone who possesses E understands F—at least in normal circumstances.

Then, this interpretation continues, what Hempel does explicitly in this passage is assert (it would be too much to call it an argument) that the DN model satisfies the necessary condition the Explanation-Understanding Condition states.

The Explanation-Understanding Condition, if true, looks like it could be a really useful tool for evaluating theories of explanation. One basic way to evaluate a theory of explanation is to “directly check it against your intuitions”: find a body of fact that the theory counts as an explanation, and ask yourself, *is* it an explanation? Or, find an explanation, and then ask yourself, does the theory count it as an explanation? If the answer to either question is no, that’s a mark against the theory.<sup>1</sup> But it would be nice to have other ways to evaluate a theory, besides directly checking it against intuitions. The Explanation-Understanding Condition provides another way. Find a body of fact E that the theory counts as an explanation of F, and check if someone could possess E without understanding F. If that is possible, then the theory must be false.<sup>2</sup> In this way, it looks like the phenomenon of understanding can be used as an independent check on theories of explanation.

The Explanation-Understanding Condition has been, and continues to be, widely accepted. Michael Friedman, for example, in his 1974 paper “Explanation and Scientific Understanding,” asks, “What is it about . . . scientific explanations . . . that [give] us understanding of the world?” (5). A page later he asks again, “what is the relation between phenomena in virtue of which one phenomenon can constitute an explanation of another, and what is it about this relation that gives understanding of the explained

1. I take it that this is the method philosophers *should* call “checking your intuitions,” even though my statement of the method never uses the word “intuition” or any similar word. I realize this is controversial, but that controversy is not relevant here. (I am largely in agreement with Williamson [2008] and Cappelen [2012] about the role of so-called intuitions in philosophy.)

2. Of course, this method involves using your “intuitions” about whether someone in some hypothetical circumstances understands F.

phenomenon?” (6). Friedman’s questions presuppose the Explanation-Understanding Condition. I could go on with examples like this for a long time. Here are two more: in 1984 Elliott Sober asserted that “Explanations afford understanding” (78). In 2003 James Woodward asserted that “It is a plausible constraint on what an explanation is that it must be something that provides understanding” (179). (I will have more to say about these authors’ use of the Explanation-Understanding Condition later.)

I am “against” the Explanation-Understanding Condition. On one interpretation, the condition is false. On another interpretation, it can’t, after all, be any use for evaluating theories of explanation.

A certain idea has shadowed philosophical debates about explanation like an assassin. That idea is that theories of explanation really are, or really should be, theories of answers to why-questions. I believe in this idea. I believe that all that philosophers of science should want out of a theory of explanation is a theory of answers to why-questions. And I hold that thinking about the phenomenon of understanding is no help at all in evaluating theories of answers to why-questions.

## 11.2. Explanations as Answers

No theory of explanation a philosopher has proposed has ever really been a theory of *all kinds* of explanation. Instead they have always been, at best, theories of explaining *why*. Suppose Smith asks Jones why the moon is waning, and in response Jones explains why the moon is waning by asserting some propositions (presumably about, among other things, the earth’s and moon’s orbits). The DN model, and every theory that has followed, can be read as aiming to say something about what it took for Jones to have succeeded in explaining (to Smith) why the moon is waning. But now suppose that Smith asks Jones what a gene is, and in response Jones explains what a gene is by asserting some propositions (presumably about inheritance, maybe also about chromosomes and DNA). There is no way to read either the DN model or any theory of explanation that has followed as aiming to say anything about what it took for Jones to have succeeded in explaining (to Smith) what a gene is.

I said in the last paragraph that every theory of explanation since the DN model can be read as a theory of explaining why. That’s not entirely true. Bas Van Fraassen’s theory of explanation, in *The Scientific Image* (1980), can’t. His theory takes the form of a theory of why-questions and their answers. That’s not the same thing as a theory of explaining why. Assuming that the

“canonical form” of an answer to the question why  $Q$  is something of the form “ $Q$  because  $R$ ,”<sup>3</sup> a theory of answers to why-questions will fill in the schema

(S<sub>1</sub>)  $Q$  because  $R$  iff . . .

A theory of explaining why, by contrast, will fill in the schema

(S<sub>2</sub>) Person  $P_1$  explained (to person  $P_2$ ) why  $Q$  iff . . .

Since explaining is a speech act, to complete (S<sub>2</sub>) is to lay down the conditions required for someone to have successfully carried out this speech act. One does not have to lay down such conditions to complete (S<sub>1</sub>).

What I just said might suggest that a theory of explaining why must do “more” than a theory of answers to why-questions, in the sense that a theory of explaining why must contain a theory of answers to why-questions as a part. This suggestion is not true. A theory of explaining why will certainly *make use* of the notion of an answer to a why-question. But it can do this without containing, or entailing, any particular *theory* of answers to why-questions. In support of this, here is how I think the true theory of explaining why, the true completion of (S<sub>2</sub>), will start:

(S<sub>3</sub>)  $P_1$  explained to  $P_2$  why  $Q$  iff  $P_1$  told  $P_2$  the answer to the question why  $Q$  in the following way/manner: . . .

Right there after “told  $P_2$ ” is the reference to the answer to the question why  $Q$ . But (S<sub>3</sub>), while it mentions answers to why-questions, doesn’t contain a theory of such answers. Of course (S<sub>3</sub>) is not itself a theory; it trails off into ellipsis. But I don’t think that the true completion of (S<sub>3</sub>) will contain a theory of such answers either. The stuff that comes after “way/manner” will (obviously) specify a way of telling someone a proposition, but I can’t think of any way of specifying a way of telling that involves stating a theory of answers to why-questions.

When Van Fraassen proposed his theory of explanation, which, again, was a theory of why-questions, he did not say that the criteria for judging his

3. Certainly not all answers have this form; the answer to the question why that plant is closing its stomata may be that it is doing this in order to conserve water. I happen to believe that every “in order to” answer—sometimes called teleological answers—is equivalent to a because-answer, but will not defend this claim here. (I defend it in chapter 6 of Skow 2016.)

theory were different from those for judging earlier theories of explanation, like the DN model. But if his is a theory of why-questions and their answers, while earlier theories were theories of explaining why, shouldn't the criteria be different? The answer to this question would be yes—if the question's presupposition, that earlier theories were theories of explaining why, were true. The fact that Van Fraassen did not say that his theory should be judged on different criteria is some evidence that he, at least, thought that this presupposition was false. Anyway, it is my view that the presupposition is false. Although earlier theories can be read as theories of explaining why, I don't think they should be read that way. I don't think they're best interpreted that way. It's not only Van Fraassen's theory of explanation that is really a theory of answers to why-questions; that's what all of them really are.<sup>4</sup>

I don't take these observations to be terribly novel or even all that revolutionary. They've been around in some form or another from the beginning: Hempel himself says that explanations are answers to why-questions.<sup>5</sup> But even if the idea that philosophers of science should really be after answers to why-questions is in fact widely accepted, it is not really taken to heart. For philosophers rarely phrase their theories as theories of answers to why-questions. They phrase them as theories of explanation, phrasing them as completions of one of these:

- Fact X explains fact Y iff . . .
- Fact X is an explanation of fact Y iff . . .

At one level this is just a matter of terminology: there is, I think, a well-established practice of using the noun "explanation," in some contexts at least, as a general term for answers to why-questions (only). In some contexts, "I was confused about the photoelectric effect; then Professor Smith gave me the explanation" means "I was confused about why the photoelectric effect

4. There are some exceptions. For example, Achinstein is explicit that the theory he presents (1983) is a theory of the speech act of explaining. I think this makes his theory a kind of theory that philosophy of science should not be interested in. I will defend this claim below.

5. However, Hempel did not say that every answer to a why-question was an explanation; he isolated (or tried to isolate) a special class of why-questions that he called "explanation-seeking" (see Hempel 1965, 335). I argue against the existence of such a special class, and defend more generally the idea that theories of explanation really are, or at least should be read as, theories of answers to why-questions, in chapter 2 of Skow (2016). Some of the distinctions I've been drawing were first drawn by Bromberger, in the 1960s. He observed that one can explain things other than answers to why-questions, and emphasized that explaining an answer is different from telling someone the answer; see the papers collected in Bromberger (1992).

happens; then professor Smith told me (the answer to the question) why the photoelectric effect happens.” Still, overuse of the word “explanation” can seduce us into false claims and mislead us into accepting bad arguments. The Explanation-Understanding Condition is a good example of this.

### 11.3. Against the Explanation-Understanding Condition

Here is the Condition again:

Something E is an explanation of some fact F only if someone who possesses E understands F—at least in normal circumstances.

Is this right? That depends on how it is interpreted. The first thing we need to do, to make it more precise, is to reword it so that it does not speak of *facts* as the things that are explained, or understood. We do talk of explaining or understanding facts, but it is misleading to talk this way in this context. An explanation of the fact that the moon is waning is, in this context, just an explanation of *why* the moon is waning. Similarly, to understand the fact that the moon is waning is, in this context, to understand *why* the moon is waning. So let’s make the “whys” that belong in the statement of the Explanation-Understanding Condition explicit (from now on I’m going to leave off the qualification about normal circumstances):

(EU) Something E is an explanation of why Q only if someone who possesses E understands why Q.

Now I want to distinguish two readings of (EU). One reading takes “explanation of why Q,” as it appears in (EU), to mean nothing more than “answer to the question why Q.” Another reading takes talk of an “explanation of why Q” to be talk of acts of performing the speech act of explaining.

Here is the first reading:

(EU<sub>1</sub>) A proposition P is the answer to the question why Q only if anyone who knows P understands why Q.

I should note that to get (EU<sub>1</sub>) from (EU) I did more than replace reference to an explanation with reference to an answer. There’s another difficult bit

of wording in (EU): it speaks of “possessing” an explanation. Once we eliminate “explanation,” we are left with talk of someone “possessing” an answer to a why-question. But what is it to “possess” an answer to a why-question? The most natural thought is that possessing an answer is just knowing that answer.

Now we have one interpretation of the Explanation-Understanding Condition. Do we have in it an interesting constraint on theories of answers to why-questions? I don’t think so, because I don’t think (EU<sub>I</sub>) is even true. Since knowing why Q is the same thing as knowing the proposition that is the answer to the question why Q, (EU<sub>I</sub>) amounts to saying that knowing why Q is sufficient for understanding why Q. But this claim is false. Understanding is a greater achievement than knowledge, not a lesser one.

Consider, for example, Lester. Lester has never taken a chemistry class, or studied the subject on his own. But like all of us he has heard people use the words “acid” and “base,” and words related to them, like “acidic.” He has heard people say things like “Lemon juice is acidic,” and “Baking soda is basic.” He’s had enough exposure to these words for them to be part of his vocabulary. For example, when he called his doctor recently with a stomach ache, and she asked him if he’s had anything acidic to eat or drink recently, he understood her question, and replied that he had had some orange juice earlier in the evening. Now the other day Lester’s niece was showing him the lab experiment she’d done in her chemistry class. She dipped a piece of litmus paper into a liquid, and it turned red. Lester was very curious about this result. “Why did the litmus paper turn red?” he asked her niece. “Because this stuff I dipped it into is an acid” she replied.

Here Lester’s niece told him the correct answer to the why-question he asked. Moreover, she knew that was the correct answer, having studied this experiment in her chemistry class. Furthermore, Lester has no reason to doubt her testimony. So after hearing his niece’s answer Lester knew that the paper turned red because it was dipped in acid. In other words, Lester knew why the litmus paper turned red. But, I claim, he did not *understand* why the litmus paper turned red. This is a counterexample to (EU<sub>I</sub>).

What more would it take for Lester to understand why the litmus paper turned red? I am not really sure. It doesn’t really matter for what is to follow. But for what it is worth, here is some speculation. One possibility is that Lester needs more knowledge. He needs to know more than just the answer to the question why the litmus paper turned red in order to understand why the litmus paper turned red. Maybe what he needs to know, in addition, is some chemistry. More specifically, maybe he needs to know something about why dipping litmus paper into acids turns it red. If he knew that, he would

know something about the “connection” between the event the why-question concerns (the color-change), and the fact offered in the answer to that why-question (that the paper was dipped in acid). And this is not just any old connection; instead, it appears to be the connection *in virtue of which* that answer counts as the answer. That is, the fact that acid causes litmus paper to turn red via such-and-such chemical process appears to be the (or at least an) answer to the question “Why is it true that the paper turned red because it was dipped in acid?” If these speculations are on the right track, then they suggest a generalization: maybe, in general, understanding why Q requires not just that one know the proposition that is the answer to the question why Q, but also know something about why that proposition is the answer. In terminology I introduce in (Skow 2016), the claim is that understanding why Q requires not just knowing some or all of the reasons why Q, but also knowing why those reasons are reasons.<sup>6</sup>

So the first interpretation of (EU), namely (EU<sub>1</sub>), is not a principle that can be used to evaluate theories of answers to why-questions, because it is false. I turn now to a second interpretation of (EU). This interpretation focuses, not on the answer to the question why Q, but on the act of explaining why Q. It interprets (EU) as stating a necessary condition on having performed this speech act. Here is one way to state this condition:

(EU<sub>2</sub>) Person P<sub>1</sub> explained to P<sub>2</sub> why Q only if, as a result of what P<sub>1</sub> did, P<sub>2</sub> understood why Q.

What (EU<sub>2</sub>) is saying is that to count as having *explained* to P<sub>2</sub> why Q, as opposed to, say, *telling* P<sub>2</sub> why Q, one must do more than get him to believe the answer to the question why Q. What more that is (EU<sub>2</sub>) does not say—except that it is whatever it takes for P<sub>2</sub> to end up understanding why Q.

So is (EU<sub>2</sub>) right? I honestly have no idea. I can see reasons to doubt it. There are plenty of principles that compete with it. Here are a few:

6. Stephen Grimm holds that understanding why E happened requires knowing, not just that C (a cause of E) happened, but also knowing something about the “modal relationship” between C and E (Grimm 2014). It may be that the connection between C and E in virtue of which “because C happened” is the answer to the question why E happened just is the modal connection Grimm focuses on. But I will not pause here for a detailed comparison of the two views. Michael Strevens holds that understanding why E happened requires directly mentally apprehending the explanation of E, where knowing a proposition is not sufficient for directly mentally apprehending it (Strevens 2013). If we assume that “the explanation of E” here denotes the answer to the question why E happened, then I disagree; one need not directly mentally apprehend the answer, but one does need to know other propositions in addition to the answer.



(EK) P<sub>1</sub> explained to P<sub>2</sub> why Q only if, as a result of what P<sub>1</sub> did, P<sub>2</sub> *knew* why Q.

(EP) P<sub>1</sub> explained to P<sub>2</sub> why Q only if, as a result of what P<sub>1</sub> did, P<sub>2</sub> *was in a position to know* why Q.

When I say that these principles compete with (EU<sub>2</sub>) I don't mean that they are incompatible with (EU<sub>2</sub>); in fact, assuming that understanding why Q entails knowing why Q, (EU<sub>2</sub>) entails each of them. They compete with (EU<sub>2</sub>) not for truth, but for acceptance. Maybe having one's audience (merely) know why Q is the strongest (relevant) necessary condition on successfully explaining why Q; maybe one's audience doesn't also need to have achieved so much as having come to understand why Q.

When I compare (EU<sub>2</sub>) to (EK) and (EP), I just don't know how I would go about choosing between them.

But I also think it doesn't matter. What reason does a philosopher of science have to care whether (EU<sub>2</sub>), rather than one of the others, is true? The people who have reason to care are philosophers who aim to have a theory of the speech act of explaining. But, as I argued above, that's not what philosophers have been looking for under the heading of a theory of explanation.

Of course, it's possible that while philosophers of science *haven't* been looking for a theory of the speech act of explaining, they *should have* been.

Now if philosophers of science should be looking for a theory of explaining, that must be because a complete philosophy of science requires such a theory. But in fact I think that a complete philosophy of science shouldn't include a theory of explaining.

A complete philosophy of science would have to include a theory of explaining if, but only if, explaining were in some way part of the nature of science. Now we do sometimes say things that suggest that it is. Some philosophers say that one of the aims of science is to explain why things happen. They say this in opposition to those who say that science aims to provide (just) a comprehensive description of what happens. Whether they are right is contentious; this debate is part of the debate over "scientific realism." But just what claim is it that the debate is about? Does "science aims to explain why things happen" mean just "science aims to answer why-questions," or does it mean "science aims to inform the public (or whomever), by means of performing the speech act of explaining, why things happen"? Once this distinction is made, I think it's obvious that only the first interpretation has a chance of being true. Or I should say, it is only the first interpretation the scientific realists mean to affirm.

What exactly is scientific realism? It doesn't seem to be the same thing to everyone. But certain doctrines appear to be central: the doctrine that our current scientific theories are close to being true; the doctrine that it is a constitutive aim of science, as a practice, to produce true theories; and—I think—the doctrine that it is a constitutive aim of science to produce theories that are not just true, but also contain answers to why-questions. But I just can't see why anyone would think it was a constitutive aim of science that its practitioners perform any particular speech acts.

Couldn't a mute Robinson Crusoe have been an excellent scientist? Maybe he finds himself alone in Geneva in a postapocalyptic world. He fixes up the Large Hadron Collider and does a bunch of experiments. He improves on the current formulation of quantum field theory, devising a theory that is closer to the truth. He discovers the answer to the question why electrons, or Higgs bosons, do this or that. But he never says anything (not to anyone else, there being no one else there, and not to himself either). He never performs any speech acts. In particular, he never explains anything to anyone. Would any scientific realist really maintain that Crusoe was in some way defective as a scientist—that he was failing to pursue all of the aims of science?

Let's take stock. I distinguished two readings of the Explanation-Understanding Condition. One reading, (EU<sub>1</sub>), says that some fact about understanding is a necessary condition on something's being an answer to a why-question. But (EU<sub>1</sub>) is false. The other reading, (EU<sub>2</sub>), says that some fact about understanding is a necessary condition on someone's having (successfully) performed the speech act of explaining. Maybe (EU<sub>2</sub>) is true, but whether it is is of no interest to the philosophy of science.

#### 11.4. An Example: Sober on Explanation and Causation

What I want to do next is look at a few examples of philosophers invoking the Explanation-Understanding Condition. We are now in a position to see the mistakes these invocations have led to.

I selected the examples to look at from important works in the philosophy of science, but the selection was otherwise random. There are a great deal more I could have looked at.

In his 1984 book *The Nature of Selection* Elliott Sober took up the question of whether an organism's fitness causes, or explains, its survival. This led him to a discussion of whether a disposition, say a sugar cube's being soluble,

causes, or explains, its manifestation (the sugar cube's dissolving). Sober accepted that the solubility causes the dissolving, but claimed that "we feel somewhat disappointed when told that the sugar lump dissolved in water because it was water-soluble" (77). Being told that the sugar lump is water-soluble is satisfying when we have asked what caused it to dissolve, but is (in his words) "disappointing" when we have asked why it dissolved. To reconcile the claim that the solubility is a cause with the claim that citing it does not make for a "satisfying" explanation, Sober set out to "disentangle issues of causation from issues of explanation":

Consider the question "What caused Y?" One correct answer would be "the cause of Y." However, this answer might rightly be classified as unexplanatory, since it does not provide us with any better understanding of why Y occurred than we started with. Explanations afford understanding; therefore, claims asserting that X explains Y are true or false partly because of the relation of those two terms to a third—namely, *us*. Causality is not similarly "consciousness-dependent." Whether or not X caused Y is not in general influenced by whether *we* might find it interesting to be told this. (1984, 78)

Let's look closely at the argument Sober is offering. Here it is as he stated it:

1. "Explanations afford understanding."
2. "Therefore, claims asserting that X explains Y are true or false partly because of the relation of those two terms to a third—namely, *us*."
3. "Causality is not similarly 'consciousness-dependent.'"
4. [Implicit conclusion:] It is false that, or at least not always true that, if X is a cause of Y, then X explains Y.

I don't think that this is a very good argument. Its problems start with line 1. Line 1 is a version of the Explanation-Understanding Condition. I have proposed two readings of this condition. But the first reading, (EU<sub>1</sub>), is false, and the second, (EU<sub>2</sub>), is not an interesting thesis (from the perspective of the philosophy of science anyway). So the first premise of Sober's argument is either false or uninteresting.

The charge of uninterestingness might seem itself uninteresting. Surely what matters is how interesting Sober's *conclusion* is, not how interesting his premises are. So before dismissing the reading of Sober's argument that takes

line 1 to be (EU<sub>2</sub>), we should at least see what the argument's conclusion looks like on that reading.

Since (EU<sub>2</sub>) comes from (EU) by interpreting talk of explanations as talk of the speech act of explaining, we should interpret line 4 to also make a claim about the speech act of explaining. Here is one stab at an interpretation of line 4 like that:

- 4\*. The following is false: for any description C of a cause of E, someone can explain to someone else why E happened by saying 'E happened because C happened.'

I think that 4\* may well be true, but that whatever interest it has does not essentially depend on its focus on the speech act of explaining. Here is how I would put the main thought Sober wants to get at in the quoted passage above, that 4\* seeks to capture: if I ask you why some event E happened, and you reply that E happened because its cause happened (let's pretend that E has only one relevant cause), you will have failed in some way to respond to my question appropriately.

Now I can see a route one might try to take from this claim to an interesting conclusion, a conclusion about what can and cannot be an answer to a why-question. Here is how it goes, starting from the claim I ended the last paragraph with:

5. If, in response to the question why E happened, someone X responds by asserting that E happened because its cause happened, then X will have given a bad response to the question.
6. If, by asserting P, someone gives a bad response to a question, then P is not an answer to that question.
7. So the proposition that E happened because its cause happened is not an answer to the question why E happened.
8. So even though the cause of E is a cause of E, it is false that E happened because its cause happened; it is false that for every description C of a cause of E, 'E happened because C happened' is true.

I want to make two points about this argument. The first is about who this argument is supposed to target. Elsewhere in his book Sober takes issue with the thesis that to answer the question why E happened one must cite, or describe, causes of E. The argument (5)–(8), if sound, would show this thesis to be false, *under one interpretation*. That interpretation says that 'E happened

because  $R^7$  is true if and only if the sentence  $R$  in some sense describes a cause of  $E$ —where any way of describing that cause is permitted. But this does not seem to me a strong point against those who think that answers must cite or describe causes. They will just opt for a weaker interpretation of their thesis; they will find a principled way to restrict which descriptions of the  $E$ 's causes are allowed as values of  $R$ .

My second point is that the argument's premise 6 is false. Asserting  $P$  can be a bad response even if  $P$  is an answer. How? Here is one way. It could be that  $P$ , while an answer, is not the answer the asker is looking for. I might watch a window break after being hit by a rock,<sup>7</sup> but not know whether the window was fragile, or whether instead it was made out of superglass that had just one point of weakness, the point where the rock was lucky enough to hit. I ask why the window broke. If you tell me that it broke because it was hit by a rock, that's a bad response. But the problem with it is not that what you said was *false*. "Because it was hit by a rock" is true, and is (therefore) an answer to the question why the window broke. The problem with your response is, instead, that it was not the answer I was looking for.

The point can be made without using a why-question as an example. For any question, there are many true answers to that question that it would be wrong to assert when asked that question. Suppose you and I are in some Boston greenspace or other, but I, being from out of town, don't know which. Are we on the Common, or in the Public Garden, or what? I asked you where we are. If you reply "We are in Boston," you give a true answer to the question "Where are we?" But it was wrong of you to assert that answer. It's not the answer I was looking for. Similarly, suppose I ask you who came to the party, and you answer "The people who came to the party came to the party." You haven't said something false. You've given me an answer to my question. But it's an answer I already knew. It's not the answer I wanted.

This could be what is going on with the response Sober is interested in, the response to the question why  $E$  happened that one gives by saying "E happened because its cause happened." That's certainly a bad response; but it could be bad not because it's not an answer, but only because it's not the answer the asker wanted. If this is right, then Sober's observation that it would be bad to say "Because its cause happened" in response to the question of why  $E$  happened does not show that there is any restriction on what descriptions of  $E$ 's causes can appear in an answer to the question why  $E$  happened.

---

7. After the *window* is hit by a rock, that is.

I have said that “We are in Boston” and “The people who came to the party came to the party” are correct answers to the relevant questions. This might be challenged. (I believe there are theories of questions and answers on which they are not answers.) But I don’t think Sober is in a position to challenge them. He accepts, right at the beginning of the long quotation above, that “The cause of E” is a correct answer to “What caused E?”

To sum up: Sober, in the quoted passage, tried to use a connection between explanation and understanding to reach a conclusion about when descriptions of the causes of E can, and cannot, constitute explanations of the occurrence of E. I have argued that his argument was based on a false claim about the connection between explanation and understanding, and also sketched a way that someone who thinks that any description of the causes of E can appear in an answer to the question why E happened can resist his conclusion.

### 11.5. Another Example: Woodward against the DN Model

We saw Hempel claim that the DN model satisfies the Explanation-Understanding Condition. In *Making Things Happen*, James Woodward turns this around: he argues that the DN model is *false* on the ground that it *fails* to satisfy the Explanation-Understanding Condition. Woodward in his argument focuses on explanations that do not appear to cite any laws, such as Michael Scriven’s canonical example,

INK BOTTLE: The carpet is stained because Jones knocked over the ink bottle.<sup>8</sup>

Hempel’s view about INK BOTTLE, and other similar examples, was roughly that the sentence in INK BOTTLE is true, even though no law-statements follow the word “because,” because that sentence, or someone who uses that sentence, in some way conveys information about the DN argument that is the “ideal” explanation of the carpet stain. Woodward calls the strategy behind this response “the hidden structure strategy.” The “hidden structure” is the ideal explanation; it is hidden because someone could truly assert that the carpet is stained because Jones knocked over the ink bottle without knowing what laws appear in the ideal DN argument. The hidden structure strategy is to say

8. It became canonical by being quoted at length and discussed by Hempel (1965, 360).

that a “because” statement can be true by “conveying information” about a structure that is hidden from (not fully known to) the person asserting that statement, or his audience.

Woodward does not like the hidden structure strategy, and thinks that any theory of explanation that makes use of this strategy will fall afoul of the Explanation-Understanding Condition. Here is how Woodward puts his argument:

There is yet another reason for rejecting the hidden structure strategy. This derives from a general point about the epistemology of explanation and the connection between explanation and understanding. . . . It is a plausible constraint on what an explanation is that it must be something that provides understanding. To say that certain information is “part” of an explanation or contributes to its explanatory import is to say that this information contributes to the understanding provided by the explanation. This in turn imposes an epistemic constraint on what information can be part of an explanation and can contribute to its explanatory import: such information must be epistemically accessible to those who use the explanation. Put slightly differently, the idea is that the features of the explanation that endow it with explanatory import—that make it an explanation—must be features that can be known or grasped or recognized by those who use the explanation; if not, it isn’t in virtue of possessing those features that the explanation produces understanding. On this way of looking at matters, there is something deeply puzzling about the suggestion that claims like [INK BOTTLE] explain or convey understanding in virtue of providing information about the existence of some underlying epistemically hidden structure, whether a DN argument or an ideal explanatory text. The mere obtaining of this structure, independently of anyone’s awareness of its existence, cannot be what accounts for people’s judgment that, for example, the impact of the knee on the desk is explanatorily relevant to the tipping over of the inkwell. If this line of thought is correct, it seems to follow that to the extent that information about the laws or structures that underlie singular-causal (or other sorts of causal) claims is epistemically hidden from those who use such explanations, it cannot be that this information contributes to the explanatory import of these explanations. (179–80)

While I agree with Woodward that the DN model is false, I don't think there is a good argument against it here in this passage.

Let's walk through Woodward's argument slowly. Woodward starts with a version of the Explanation-Understanding Condition:

1. "It is a plausible constraint on what an explanation is that it must be something that provides understanding."

He also asserts a version of this condition about "parts" of explanations:

2. "To say that certain information is 'part' of an explanation or contributes to its explanatory import is to say that this information contributes to the understanding provided by the explanation."

Woodward then says that claim 2 "imposes an epistemic constraint on what information can be part of an explanation and can contribute to its explanatory import." The epistemic constraint that he infers from claim 2 is this:

3. Information can be part of an explanation and can contribute to its explanatory import only if it is "epistemically accessible to those who use the explanation."

Presumably the argument proceeds from here by appealing to the implicit premise:

4. Plenty of people use, or have used, the explanation in *INK BOTTLE*, to whom no law of nature is epistemically accessible.

The argument's conclusion then follows from lines 3 and 4:

5. Therefore, no law of nature is part of the explanation in *INK BOTTLE*, and no law of nature contributes to its explanatory import.

How does all this look if we eliminate "explanation" in favor of "answer to a why-question"? It is not entirely obvious how to translate each of the premises. Here is one attempt. Line 1 becomes (EU<sub>1</sub>):

- 1\*. A proposition A is an answer to the question why Q only if someone who knows A understands why Q.



Line 2 takes line 1 and turns it into a claim about parts of an explanation, so line 2\* should take line 1\* and turn it into a claim about parts of an answer:

- 2\*. If P is part of an answer A to the question why Q, then anyone who knows A-P (“A minus P,” the information left over when P is “subtracted” from A) understands why Q to a lesser degree than someone who knows (all of) A.

Line 3 required parts of explanations to be epistemically accessible, and was inferred from line 2, so from line 2\* we should infer a claim about parts of *answers* being epistemically accessible:

- 3\*. If P is part of an answer A to the question why Q, anyone who answers the question why Q by asserting A knows P.

Let me pause briefly: premise 3 spoke both of being part of an explanation, and also of contributing to that explanation’s explanatory import. But it’s hard for me to see why adding the stuff about explanatory import is needed; how could something be part of an explanation without contributing to that explanation’s explanatory import? In terms of answers, how could something be part of an answer without contributing to that answer? As far as I can tell, it couldn’t. So I’m leaving out the stuff about explanatory import in the \*-ed interpretation of Woodward’s argument.

Let’s continue. There’s one premise left, and then the conclusion. I’ll put the whole argument here, with the last premise and the conclusion at the end:

- 1\*. A proposition A is an answer to the question why Q only if someone who knows A understands why Q.
- 2\*. If P is part of an answer A to the question why Q, then anyone who knows A-P understands why Q to a lesser degree than someone who knows (all of) A.
- 3\*. (So,) if P is part of an answer A to the question why Q, anyone who answers the question why Q by asserting A knows P.
- 4\*. Plenty of people who have answered the question why the carpet is stained by asserting the proposition INK BOTTLE did not know any law of nature.
- 5\*. Therefore, no law of nature is part of the answer to the question why the carpet is stained in INK BOTTLE.

Is Woodward's argument, interpreted this way, any good? Premise 1\*, again, is (EU<sub>1</sub>), which I have rejected. That's enough for the argument to fail. But let's look at the rest of it. Maybe there's a way to fix the argument, so that it doesn't rely on 1\*?

I think there is. The argument's intermediate conclusion, line 3\*, looks plausible to me. Surely if you assert A you know A—or at least this is usually the case. And if you know A you know (or at least are in a position to know) its parts.

Of course, the idea of a “part” of a proposition is not an ordinary one. It is some kind of technical notion.<sup>9</sup> But the paradigm case must be conjunctions: the proposition that X is part of the proposition that X and Y. In this case a proposition entails each of its parts (though not vice versa). Presumably this holds in general: if P is part of a proposition A, then A entails P. And it is not too wild to assume that if you know A, and A entails P, you are at least in a position to know P. (Maybe this looks better if we require you to *know* that A entails P; but such niceties don't matter here.)

So, interestingly, the \*-ed version of the argument arrives at a plausible intermediate conclusion, line 3\*, an intermediate conclusion that, as I have stated it, says nothing about understanding; this despite the fact that it is reached from suspect claims about the connection between understanding why Q and being an answer to the question why Q.

In fact, the rest of the \*-ed interpretation of Woodward's argument goes fine as well. I think that 4\* is true: plenty of people *have* used INK BOTTLE without knowing any laws (or, at least, we can easily consider a hypothetical scenario in which this is so). And the conclusion 5\* follows from 3\* and 4\*.

Despite all this, there is still a big problem with the \*-ed argument. The problem is not that it is unsound; the problem is that its conclusion is not the conclusion Woodward wants. Sophisticated defenders of the DN model, or of theories that descend from the DN model, Peter Railton, for example (whose views Woodward discusses at length in the chapter from which I am quoting), do not say that some law of nature is *part of the answer* in INK BOTTLE. They say instead that INK BOTTLE is *itself* only part of a larger and more complete answer, and that that larger and more complete answer contains laws of nature (see, e.g., Railton 1981).

9. It receives a detailed development in Yablo (2014).

Is there a better interpretation of Woodward's argument, an interpretation that gets him to the right conclusion? Well there is at least a different interpretation; let us see if it is better.

If you look back at the long quotation above, you will see that after Woodward drew the intermediate conclusion 3, he wrote:

Put slightly differently, the idea is that the features of the explanation that endow it with explanatory import—that make it an explanation—must be features that can be known or grasped or recognized by those who use the explanation; if not, it isn't in virtue of possessing those features that the explanation produces understanding.

I actually do not think that this is conclusion 3 “put slightly differently.” I think that this is a very different idea. To ask *what features of an explanation endow it with explanatory import*, or to ask *what features make it an explanation*, sounds to me like asking *why it is an explanation*. If we focus, as we should, on answers to why-questions, not explanation, the question we are attending to becomes the question of *why a given answer to a why-question is an answer*. What do Woodward's remarks suggest about the answer to this question? Here's what I think they suggest:

6. If P is the answer to the question of *why is it that R is the answer to the question why Q*, then anyone who asserts R as an answer to the question why Q must be in a position to know P.

Here P, the answer to the question of why R is the answer to the question why Q, describes the “features of the explanation”—that is, the answer—“that make it an explanation,” or answer.

Unlike claim 3\*, claim 6 *can* get us to the conclusion Woodward wants—if it is true. Let's first see how it gets Woodward where he wants to go, and then turn to the question of whether it is true.

The argument from 6 against the DN model is, I think, straightforward. We need just two more premises:

7. If the DN model is correct, then some law of nature is (at least part of) the answer to the question of why INK BOTTLE is the answer to the question why the carpet is stained.

8. Plenty of people who have answered the question why the carpet is stained by asserting the proposition INK BOTTLE did not know (and were not in a position to know) any law of nature. (This is premise 4\* from the \*-ed argument, slightly revised.)

From 6, 7, and 8, the desired conclusion follows:

9. The DN model is not correct.

What about premises 7 and 8? Line 8 is certainly plausible, for the same reason that premise 4\* from the \*-ed argument was. I myself don't know of any laws of nature that could be added to INK BOTTLE to turn it into a DN argument. The plausibility of line 7 is harder to assess. Rather than dwelling on line 7, though, I will skip to the end: ultimately this argument does not succeed, because its crucial premise, 6, is false.

At least that's what I think. I think that someone could know the answer to the question why Q, and could tell someone who wants to know why Q that answer, without having any idea *why* that answer is the answer.

Suppose Jones is on a tour of the Old Mill. The tour guide points to some charred wood, and says that these are the parts of the mill that burned in the fire of '03. Jones asks why the mill caught on fire in '03. The tour guide replies that it caught on fire because it was struck by lightning. Jones then persists in his questioning. He says, "That's great, but I have another question. Why is that the answer? Why is it that 'because it was struck by lightning' is the answer to the question of why the mill caught on fire?" The tour guide is baffled, and a little annoyed. "I don't know" he says. He gives Jones a look to get him to shut up. Still, despite not knowing the answer to Jones's second question, the tour guide knew *why the mill caught on fire*. He also successfully told Jones why the mill caught on fire.

It is important to distinguish the baffling question Jones asked from another one. Jones did not ask *why the answer is true*. If, after the tour guide had said "Because it was struck by lightning," Jones had said, as my two-year-old automatically says in response to pretty much everything I say, "Why?" we would most naturally take him to be asking *why the mill was struck by lightning*. To ask that is to ask why the tour guide's answer is true. That's not the same as asking why it is an answer.

I extracted premise 6 from some things Woodward said. But he didn't just say some things that suggested 6; he also offered a little argument for those things. Maybe 6 will seem more plausible if we look at that argument? Here, again, is what Woodward wrote; I've italicized the relevant bit:

the idea is that the features of the explanation that endow it with explanatory import—that make it an explanation—must be features that can be known or grasped or recognized by those who use the explanation; *if not, it isn't in virtue of possessing those features that the explanation produces understanding.*

This argument is worthy of our attention not just because Woodward offers it, but also because it appeals to a connection between explanation and understanding, and such connections are the main concern of this chapter. So what is the argument? I think it is a good idea to start by making the argument more explicit using Woodward's own terminology, before trying to see how to put it as an argument for 6. Here is what the argument seems to me to be:

10. The features of an explanation that make it an explanation are (also) features in virtue of which that explanation produces understanding.
11. The features of an explanation in virtue of which it produces understanding are features that those who use that explanation can know it to have.
12. Therefore, the features of an explanation that make it an explanation are features that those who use that explanation can know it to have.

The conclusion 12, of course, corresponds to premise 6 of the earlier argument:

6. If P is the answer to the question of *why is it that R is the answer to the question why Q*, then anyone who asserts R as an answer to the question why Q must be in a position to know P.

What we need to do is “translate” lines 10 and 11 so that they can serve as premises in an argument for 6. I don't think this is hard to do, in light of my earlier discussions of what could be meant by “explanation produces understanding,” and my earlier claim that “X in virtue of W” can be rendered as “W is the answer to the question why X.” I'll label the translations of lines 10 and 11 with \*s:

- 10\*. If P is the answer to the question of why is it that R is the answer to the question why Q, then P is the answer to the question of why anyone who knows R understands why Q.
- 11\*. If P is the answer to the question of why anyone who knows R understands why Q, then anyone who offers R as an answer to the question why Q is in a position to know P.

6. (Therefore,) if P is the answer to the question of why is it that R is the answer to the question why Q, then anyone who asserts R as an answer to the question why Q must be in a position to know P.

This argument is valid, but the premises 10\* and 11\* don't seem to me to have anything going for them. Start with 10\*: the problem with it is that it presupposes that it is *true* that anyone who knows the answer to the question why Q understands why Q. I've said why I think this is false. I also don't see much reason to accept premise 11\*. I can see no reason to think that you can only answer a why-question if you know why the answer you're giving produces understanding in those who hear it.

I have been examining Woodard's argument that the DN model fails to respect some connection between explanation and understanding. I found two different arguments that Woodward might have been giving; I have argued that both of them fail.

## 11.6. Conclusion

The idea that there is an important connection between explanation and understanding is a common one. The alleged connection, vaguely stated, is that explanations produce understanding. If it existed, this connection would open up a new avenue for searching for the true theory of explanation: just investigate what things produce understanding. But I think that the promise that comes with this alleged connection is not fulfilled. On one precisification of the vague idea, the connection does not exist—knowing the answer to the question of why some event E happened is not sufficient for understanding why it happened. On another precisification the connection is this: someone has performed the speech act of explaining with respect to the question of why E happened only if their audience came to understand why it happened. Maybe so, but a theory of this speech act is not what philosophers of science have been after under the heading “theory of explanation.”

After all this negativity I want to draw attention to the fact that my title is not “Against Understanding.” While I have opposed a certain use of the notion of understanding in theorizing about answers to why-questions, I do not think that it is a useless notion for other purposes, or that the question of what it takes to understand why some event E happened is not itself independently philosophically interesting. In fact at least one place where the notion of understanding may be important is not hard to find. I said earlier that one of the aims of science is to answer why-questions (at least, according to some

scientific realists). But it is plausible that science aims at more. Certainly science aims not just to “have” the answer to the question of, for example, why the dinosaurs went extinct, in some thin sense of “have,” where believing it on very weak evidence is enough. Science aims to *know* why the dinosaurs went extinct. But even this might undersell how lofty the aims of science are. I find it plausible that science aims to *understand* why the dinosaurs went extinct; more generally, it aims, for each fact in its domain, to understand why that fact obtains. Even if we do not need to understand understanding in order to figure out what it takes to be an answer to a why-question, we do need to if we want to understand science.

## Acknowledgment

I’d like to thank Fred Feldman for discussing this material with me.

## References

- Achinstein, Peter. 1983. *The Nature of Explanation*. New York: Oxford University Press.
- Bromberger, Sylvain. 1992. *On What We Know We Don’t Know*. Chicago: University of Chicago Press; Stanford, CA: Center for the Study of Language and Information.
- Cappelen, Herman. 2012. *Philosophy without Intuitions*. New York: Oxford University Press.
- Friedman, Michael. 1974. “Explanation and Scientific Understanding.” *Journal of Philosophy* 71: 5–19.
- Grimm, Stephen R. 2014. “Understanding as Knowledge of Causes.” In *Virtue Epistemology Naturalized: Bridges between Virtue Epistemology and Philosophy of Science*. Ed. Abrol Fairweather. Heidelberg: Springer. 329–45.
- Hempel, Carl. 1965. “Aspects of Scientific Explanation.” In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press, 331–496.
- Railton, Peter. 1981. “Probability, Explanation, and Information.” *Synthese* 48: 233–56.
- Skow, Bradford. 2016. *Reasons Why*. New York: Oxford University Press.
- Sober, Elliott. 1984. *The Nature of Selection*. Chicago: University of Chicago Press.
- Strevens, Michael. 2013. “No Understanding without Explanation.” *Studies in History and Philosophy of Science* 44: 510–15.
- Van Fraassen, Bas C. 1980. *The Scientific Image*. New York: Oxford University Press.
- Williamson, Timothy. 2008. *The Philosophy of Philosophy*. Oxford: Wiley-Blackwell.
- Woodward, James. 2003. *Making Things Happen*. New York: Oxford University Press.
- Yablo, Stephen. 2014. *Aboutness*. Princeton, NJ: Princeton University Press.

---

# Understanding and Fluency

*J. D. Trout*

## 12.1. Introduction

In philosophy and psychology, there is now a widely used notion of the *sense of understanding*, typically a *feeling* conveyed by an explanation that may be true or false, and invoked to explain why people make the choices they do. This sense is most prominent when we experience wonder, or moments of insight. The notion of understanding itself, however, may not include this feeling of fluency, or possess a distinctive phenomenology. As a description of its use in both technical and nontechnical contexts, ‘Understanding’ seems loosely associated with properties like transparency (things we understand we can also introspect), or voluntary (cognitive) control (things we understand we can turn over in our mind).

I will discuss recent empirical literature on attention and memory in which there are many kinds of candidate cases of understanding that lack these properties of transparency and voluntary (conscious) control. In fact, ‘understanding’ may be a term that denotes an unprincipled stew of states, processes, capacities, and goals that are sometimes present and sometimes not when we apply the term or deploy the concept. In light of the evidence I consider, understanding may be a metacognitive motley of unobservable states of memory and attention, a hidden assemblage of mechanisms that, together, yield cognitive capacities that can contribute to intellectual control and success. There is nothing in this description that supports the bold pronouncements of some philosophers that understanding must possess a specific property, usually one available to introspection. There may be value in seeking a unified



account of understanding, but it is an open question whether it consists in a set of necessary and sufficient conditions.

We may experience a sense of understanding, a feeling of cognitive control or fluency that inclines us to believe we have mastered a skill or concept. Yet, in ordinary life, that cue is notoriously unreliable. Any unity we find in this sense comes *not* from the involvement of common mechanisms across diverse cases, but rather the phenomenological residue of these messy cognitive activities in the common goal of pursuing the truth. Accordingly, philosophical declarations about the requirements of understanding are not premature but misplaced. In my view, the very fact that people address such proposals at length shows that philosophical fluency does not reliably identify and weed out the products of unpromising methods and frivolous constructs in the discipline.

## 12.2. The Conceptual Analysis of ‘Understanding’

Imagine that you are given a rich assortment of measurement tools and theoretical models to characterize some prototheoretical process or event in our psychology, like understanding: reaction times and error rates to measure processing speeds and accuracy, and MRIs and PET scans to chart areas and levels of neural activation. Instead of using those tools and models, and remaining open to the possibility that the nature of the underlying processes may not be traceable by our introspection and brute associations, we reject those tools and models and rely on methods of conceptual analysis that haven’t changed in centuries.

Sadly, this is no parody. In fact, this would seem an accurate description of the philosophical study of understanding.

The concept of understanding, like the concept of love or anger or joy or sorrow, is many things to many people. ‘Understanding’ is a notoriously difficult concept to define. But then again, so are lots of words that have a life at once in the lay and the scientific worlds that serve many purposes—words like ‘belief’, ‘cloud’, ‘desire’, etc. But that is not news. If the model of philosophical analysis is to propose candidate definitions until one meets a certain conception of philosophical precision, we shouldn’t be surprised to find that most of these concepts just can’t carry the freight.

Understanding is a kind of lurking presence in contemporary work in epistemology and on explanation. It is sufficiently related to these

important topics to bear mentioning, but not concrete enough to get any traction. This has left work on the nature of understanding half-baked, tentative, impressionistic, and just plain programmatic. It is also in a terrible state. Unlike epistemology and explanation, which enjoy broad agreement about many of the basic issues (truth matters to knowledge, justification is an important feature of knowledge), participants can't agree on even the most basic features of understanding. Is 'understanding' a success word (does it require "downstream successes"?) Is understanding a representation? Is it propositional? Is it a recognitional capacity, and if so, why doesn't its cue sensitivity make the process of understanding just like the epistemic process of truth tracking, and so a species of knowledge? Existing analyses have, accordingly, expanded the senses of 'understanding' to understanding *how*, understanding *why*, understanding *of*, and understanding *that*.

As with other concepts that populate the inventory of lay psychology, like belief or desire or intention or intelligence, its boundaries are fuzzy and its uses are driven by many forces of different strength and opposing or orthogonal direction. Such forces have little to do with truth or satisfaction conditions, like the affective need for cognitive closure, the pragmatic demands of workaday communication, and personal idiosyncrasy. Not surprisingly, then, understanding is a mongrel concept, used ambiguously, vaguely, and unevenly across languages, cultures, and ages. Accordingly, the research on understanding oriented to capture this character and variability is carried out by intellectual hybrids who have at once a deep interest in foundational issues and a deep respect for empirical approaches (see, e.g., Wilkenfeld, Plunkett, and Lombrozo 2016).

Still, contemporary analytic philosophers tend to toil in their narrow and traditional specializations, and they have some surprisingly specific and often conflicting ideas about the extension and proper application of the concept of understanding. All of them are held with great conviction. On the basis of their own phenomenological and narrative evidence, some philosophers seem to have their own primitive theories of how psychological capacities are characterized and interact. For example, understanding is "a state that is constituted by a state of conscious transparency" (Zagzebski 2001, 246). You are aware that you understand whenever you do. Understanding is also said to have a special relation to curiosity; it supplies the "aha" moments that act as "legitimate closure of inquiry" (Kvanvig 2011, 89). Only understanding, not knowledge, has the power to slake our curiosity. (Never mind that some of the best-known "aha" moments in history are fantastic blunders, marking

closures of inquiry that are stupendously illegitimate.)<sup>1</sup> On yet another view, also apparently, understanding has nothing special to do with transparency or curiosity, but is characterized primarily by reflective equilibrium: “The individual commitments that comprise such a system must be reasonable in light of one another, and the system as a whole must be at least as reasonable as any available alternative in light of relevant antecedent commitments.”<sup>2</sup>

It is hard to know what status these claims have. None of them strike me as obviously true, but even if they did, we should recognize that many false, even stupid, assertions, strike humans as obviously true. Nor are these assertions the consequence of scientific theorizing. What, then, are we to make of their seemingly authoritative delivery and with it, the idea that this approach properly addresses deep and durable question?

I will identify the origin of this conceit in a sense of understanding, a feeling of fluency, conveyed by entrenched methods and concepts in contemporary philosophy. In the meantime, let me suggest an alternative vision of many protoscientific concepts, including understanding. ‘Understanding’ is like a chimera. It is not a unified or monolithic ‘thing’, but a mongrel, an awkward and unprincipled mixture of traits. In its objective use, ‘understanding’ covers a psychologically loose assemblage of cognitive achievements. I will show how a systematic empirical perspective shows by decomposition how the term ‘understanding’ has no role to play in a working science, but it is worth noting that the term also has few of the hallmarks of unification even in ordinary usage. What tendencies it has in that direction are largely superficial artifacts of some prominent and compelling phenomenological experiences associated with general cognitive achievements.

Philosophers engaged in this conceptual analysis might cast their efforts not as the final act, but a “place to start” in the analysis of the concept of understanding. But that is the standard claim of conceptual analysis, and it now has only the ring of ceremony. The specific conceptual analysis of understanding we’ve considered is not oriented toward projects confirmable by

1. Trout (2002); also see Trout (2016). This tendency reaches into many domains, such as finance and investment. Increases in the sense of understanding, for instance, reduce perceived risk (Long, Fernbach, and De Langhe 2016).

2. Elgin (2007, 2). Not surprisingly, philosophers more likely to take a philosophy-of-science approach to the study of understanding are less prone to this basic, distinctly philosophical, conceptual analysis. For example, marking his own view off from Zagzebski’s, Grimm says: “[U]nderstanding requires truth, is not transparent, and can be Gettiered” (2006, 516). Also see Strevens (2013) for a somewhat more naturalistic approach. For more work by philosophers who canvas empirical approaches to the notions of understanding and explanation, see Wilkenfeld (2017) and Khalifa (2017).

scientific (as opposed to anecdotal) support. It is not couched in a vocabulary deployed in research on memory, attention, or computation generally. It does not cite those sources. It does not entertain the possibility that its characteristics are moderated by individual variables, by demographic facts, by culture, or by epoch. In short, there is no evidence in these conceptual analyses that they are conceived as a place for psychologists and other interested parties to begin their investigation. On the contrary, the firewall of neglect they have built around the concept indicates that such empirical investigations are at best unnecessary, and at worst, unwelcome.

In our common uses of ‘understanding’, lots of consequences are not just unarticulated, but undetermined. It is not as though there is some settled extension of the term, but we have not fully captured it; the extension is no more unfinished than the extension for ‘thought’ or ‘intention’ or ‘desire’. Focused efforts advertise the extension as under construction, and there isn’t much evidence of progress. Rather than the philosopher unearthing the implicit consequences of the notion of understanding, philosophers are often trying to make up a coherent notion of understanding that nothing in our use or theoretical practice dictates.

Humans have inbuilt cognitive limitations that place some explanations permanently beyond their grasp. This is especially so for accounts of explanation that demand much from understanding. Our faculties of attention and memory arise from a rigid neurological architecture. The ordinary standards of understanding—normally requiring that the causes responsible for the effect be introspectable and separately trackable—cannot be met by an architecture that squeezes efficiency out of processing by rendering so much processing automatic, shallow, and opaque. Our perceptual and cognitive systems are capable of maintaining only very small numbers of memory and attention contents from just a few information sources. If we had undistorted and deliberate access to the content of attentional windows of arbitrary size, could track those events with less constrained powers of identification and discrimination, and could integrate the many temporal levels and modal qualities of information, we might be able to meet the standards of grasping or understanding normally imposed when we attempt to ‘understand’. In other words, if we had capacities that we don’t, we might have been able to do things we can’t.<sup>3</sup> I don’t think you want an account of understanding that requires that you appreciate things like separable

---

3. A conceptual treatment of this issue is found in McGinn (1991).

dependence relations, because almost nothing in the world is understandable in that sense.

I will focus on two features of existing philosophical accounts of understanding. The first is that they include concepts like transparency, awareness, and control that enjoy a certain fluency in the philosophical tradition but belong to no scientific category. Second, and relatedly, the existing philosophical accounts of this fundamentally psychological notion—understanding—are not formulated in way that could be confirmed by scientific evidence. Instead, they characterize understanding by performing conceptual analysis on the experience of understanding, dissecting and reassembling it using only those capacities that our best psychological science already demonstrates are unfit for routine identification and estimation.

Although conceptual analysis aspires to *yield* a clear and specific product, some philosophers *begin* with very specific, even tailored, ideas about what understanding requires. Understanding, Zagzebski says, “is a state that is constituted by a type of conscious transparency” (2001, 246). But some lay and technical uses conflict with this conviction. We often say that a person understands a function even when they can’t articulate it or be aware of its features. Zagzebski continues:

It may be possible to know without knowing that one knows, but it is impossible to understand without understanding that one understands. . . . [U]nderstanding is a state in which I am directly aware of the object of my understanding, and conscious transparency is a criterion for understanding. Those beleaguered by skeptical doubts therefore can be more confident of the trustworthiness of putative understanding states than virtually any other state. (2001, 246–47)

When philosophers announce a condition for the application of the term ‘understanding’, it can be hard to tell whether they are playing a normative role like the prim grammarian, or a merely descriptive one of the anthropological lexicographer. Is this a claim about how we use the word? Is it an attempt to summarize a supposed consensus about the nature of understanding?

With an evidence base so uncertain, we might think we should treat every assertion of the meaning of ‘understanding’ as speculative or risky, until the necessary evidence is specified.

Nevertheless, Zagzebski claims that understanding is “The state of comprehension of nonpropositional structures of reality” (Zagzebski 2001, 242). Sooner or later, advocates admit to a strategy behind their specific treatment.

Some hold that an account of understanding is best treated as a virtue. Some seek an account continuous with our philosophical heritage in ancient Greece. Some believe that an account must properly navigate a particular strain of usage. “In the case of understanding, the objects would be something along the lines of ‘structures’ (Linda Zagzebski), or ‘systems’ (Julius Moravcsik), or ‘information chunks’ (Jonathan Kvanvig), or ‘dependency relations’ (Jaegwon Kim and [Stephen Grimm])” (Grimm 2012, 105).

Other accounts are very broad, leaving open the reasonable possibility that there are no features of understanding common to the cases we might be inclined to apply the term: “As a very crude first approximation, I suggest that understanding is a grasp of a comprehensive body of information that is grounded in fact, is duly responsive to evidence, and enables non-trivial inference, argument, and perhaps action regarding the subject the information pertains to” (Elgin 2007, 9).

Although it is good to be open about one’s philosophical strategies, admission is not absolution, still less a substitute for a reason that one might think understanding deserves a unified treatment. Other approaches fare no better. One might try to gain insight into understanding generally by looking at scientific understanding specifically, suggesting that scientific understanding is the state produced, and only produced, by grasping a true explanation.<sup>4</sup> Much depends on how “grasping” gets worked out. But taken at face value, “grasping” does not belong to any scientific category. And, if it is circular to define ‘understanding’ in terms of the synonym ‘grasp’, you also don’t want it defined in terms of ‘comprehension’ either. “While these descriptions differ in various ways, if there is a common idea here it seems to be that understanding is directed at a complex of some kind—in particular, at a complex with parts or elements that depend upon, and relate to, one another, and that the mind grasps or apprehends when it understands” (Grimm 2012, 105). If we had a psychologically informed and empirically rigorous account of ‘grasping’, that would go a long way toward characterizing understanding. But we don’t.

In keeping with these dependency approaches, I have a special place in my account of understanding for causation. After all, with the exception of a few special domains, only causal taxonomy promotes understanding. But my account gives priority to science. When it comes to uncovering causal taxonomies, science has no peers. The goal of this search for the causal details of

---

4. An account friendly to this one can be found in Strevens (2013).

cognitive achievement is to provide the best theoretical systematization of the empirical facts delivered by cognitive and comparative psychologists.<sup>5</sup>

### 12.3. What the Concept of Understanding Is Made of

The philosophical study of understanding involves a circle of interdefined concepts. Appreciating the structure of the situation, grasping, comprehending, manipulating representations—all with their conscious and unconscious variants—these are the central intellectual notions designed to elucidate understanding. Although the sense of understanding, when present, is palpable, it is really hard to know when *genuine* understanding is present. We can't elucidate understanding until we look at the processes of memory and understanding beneath it. This is not a plea for reductionism; it is a call for explanatory independence,<sup>6</sup> a description of notions that seem related to, if not constitutive of, understanding, that are implicated in a wide range of independently specifiable notions. The bottom line is, there are simply too many ways of representing cognitive achievements and failures equally as understanding.

Memory and attention also have dual lives in the lab and the public square, and both make understanding possible. When we consider the application of those words by the many, we see the assumption that in order to understand a statement, concept, theory, rule, etc. we must use our memory capacity. In order to understand, we must recall the characteristics of the object of understanding. And in order to characterize those properties, we must be able to focus our attention on them.

At the same time, memory and attention are not endless resources (Cowan 2005). There are limits on the number of items in memory that can be constantly activated (Just and Carpenter 1992), the number of relations among

5. De Regt and Dieks (2005) suggest two principles can explicate the notion of understanding, albeit of a scientific variety:

1. A phenomenon P can be understood if a theory T of P exists that is intelligible.
2. A scientific theory T is intelligible for scientists (in a context C) if they can recognize qualitatively characteristic consequences of T without performing exact calculations.

Principle 2 is too weak, of course, because there is a potentially infinite number of “qualitatively characteristic consequences of T” that are deductive consequences but utterly uninformative about the causes producing those consequences. In those cases, I doubt we would want to say that we understand phenomenon P.

6. For one account that emphasizes unification, see Kitcher (1989).

items in memory that can be integrated (Halford, Wilson, and Phillips 1998), and the limit enforced to avoid interference between items in working memory (Oberauer and Kliegl 2001). Feelings of understanding aside, cognitive success depends on the delicate alignment of multiple psychological mechanisms, far more intricate than crude schema like “activated portion of long-term memory,” “focus of attention,” and a set of “time-locked sensory buffers.” For example, behind long-term memory is a process called “long-term potentiation” in which a memory is encoded by a biological process made stable by a chemical state. Potentiation does not have phenomenological correlates, and so we cannot characterize its theoretical properties by the crude gazes of unaided introspection.

Memory is not the only complex process whose components are difficult to track by introspection or casual inspection. To cite another case, the faculty of attention proceeds by activating three broad component stages: Alerting, Orienting, and Executive. Both imaging and lesioning studies indicate that each of these processes is served by distinctive neural pathways, pathways whose functions are unlikely to be captured by analytic constructs arrived at through introspection of phenomenological contents (Geva et al. 2013). When we acquire understanding through observation, for example, an observation of even simple objects requires attentional capture of separable dimensions of shape, surface contour, color, and motion, to mention just a few factors. That capture process is measured in tens of milliseconds. Any less time, it cannot be captured in the window, and it is lost to later processes of storage and report. Any more time, and it interferes with processes that come after or crowds out other analyses that co-occur, placing so much load on working memory that we are unable to retain the event. Even the sensory analysis of a solitary dimension of color places measurable demands on one of many distinctive attentional processes (see, e.g., Belopolsky and Theeuwes 2010). Recent studies have confirmed and extended these observations to other dopamine genes and to the orienting network. In two different studies employing other conflict tasks, the catecholamine-O-methyltransferase gene was linked to the mental operations related to resolving conflict (Blasi and Mattay 2005; Diamond and Briand 2004). Different alleles of cholinergic genes were also related to performance on orienting tasks such as visual search (Parasuraman and Greenwood 2005), thus confirming the link between orienting and the neuromodulator acetylcholine.

It may be that there is no understanding without explanation (Stevens 2013). But it is also true that there is also no understanding without memory, without attention, and without awareness. This causal understanding portrays



the object's actual or potential changes over time, and it typically takes time to represent or imagine temporal sequences. We must be able to focus on some part of the problem and relate it to other parts. Sustained focus requires that we remember the details of those components for as long as we need to in order to represent the problem. All existing theories of attention and of memory treat these processes as incremental, and composed of identifiable ingredients.

Many common explanations for complex phenomena are true but people find them hard to understand. The idea is this. Humans have severe limits on memory and attention, and this places limits on what they can (explain and) understand. But many phenomena that we try to explain involve many variables of different causal direction and potency. I have in mind domains like population ecology, or models of watershed pollution, which include dozens of variables. Understanding these is very different from understanding, say, why a billiard ball moves the way it does after impact, in which only a few factors are involved.

Models of these domains can potentially contain hundreds of candidate predictor variables (Bishop and Trout 2005)—demographic factors (age, sex, race, education level, region of the country), genetic profile factors (sometimes dozens of them), behavioral factors (nutrition, exposure to certain chemicals, etc.). With many diseases, no single variable is a good predictor on its own; most of the variables are weakly predictive. So weak, in fact, that it is pretty routine to use some kind of statistical variable selection process to prune down the number of variables in the model in order to arrive at something that's "clinically interpretable"—in other words, to work around the limits of our processing ability!

Consider a simple model for predicting coronary artery disease (CAD), in which the model identifies patients at high risk for a coronary event. By allowing us to track changes in the risk of coronary events with the addition of each variable, the model helps us to understand the causes that bring about the coronary effects. But with twenty-two variables in the model, there is not a chance in the world that we could have understood this risk anecdotally, nor could we have followed the causal influences without the aid of a computational model. The number of items we would have to recall is more than 300% of the number we can attend to tracking them unassisted. To make our infirmity more emphatic, we cannot track the potency of the causes and their direction. Some of the causes are orthogonal and there are interaction effects.<sup>7</sup>

---

7. The twenty-two predictors in the model were age, sex, CAD diagnosis, deprivation, smoking, hypertension, diabetes, lipids, heart failure, peripheral arterial disease, atrial fibrillation, stroke, chronic kidney disease, chronic pulmonary disease, liver disease, cancer, depression, anxiety, heart rate, creatinine, white cell count, and hemoglobin. See Rapsomaniki et al. (2014).

This model edges us toward scientific understanding of CAD, but also refines and corrects our understanding of CAD in lay cases. Our lay experience with the effect of coronary artery disease may be a massive coronary event, open to casual inspection: A man of about sixty years of age clutches his chest, struggles to breathe, and falls to the ground, unconscious. We try to understand. Why did this happen? Nonexperts are told “Heart attack.” That may seem like a useful category, but if we are looking for causes, it is a term of convenience used to pick out causes, conditions, and events, too numerous and tedious to list, and too diverse to be useful. We could continue to build an ontology on it, or employ a taxonomy of heart ailments based on the casually observable actions. That is the philosopher’s discretion, and nothing about the organization of the profession prevents it. But no effort at understanding could rationalize doing so. Like calling a manic depressive “crazy” or “possessed,” this amounts to the willful imposition of archaic categories of proven inaccuracy.

The same goes for theories of understanding. The limitations on memory and attention are best seen if you have theoretical knowledge of the causes. Philosophers propose accounts of understanding and are silent about the constitutive role of memory and attention. But they could see that they were mistaken about prospective outcomes had they simply considered cognitive limitations under ordinary stresses.

Consider a common example of cognitive limitation showcased by Paul Meehl:

Surely we all know that the human brain is poor at weighting and computing. When you check out at a supermarket, you don’t eyeball the heap of purchases and say to the clerk, “Well it looks to me as if it’s about \$17.00 worth; what do you think?” The clerk adds it up. There are no strong arguments from the armchair or from empirical studies . . . for believing that human beings can assign optimal weights in equations subjectively or that they apply their own weights consistently. (Meehl 1986, 372)

Notice that in Meehl’s grocery example, we know that a simple addition is the right calculation to apply and the variable values (i.e., the prices) are usually stamped right on the products. But suppose that the computation required was much more complex. This of course would make matters even worse.

Suppose instead that the supermarket pricing rule were, “Whenever both beef and fresh vegetables are involved, multiply the logarithm of

0.78 of the meat price by the square root of twice the vegetable price”; would the clerk and customer eyeball any better? Worse, almost certainly. When human judges perform poorly at estimating and applying the parameters of a simple or component mathematical function, they should not be expected to do better when required to weigh a complex composite of those variables. (Dawes, Faust, and Meehl 1989, 1672)

These processes require attention to the kinds and quantities being measured. In order to attend, people must hold in working memory the items like beef and fresh vegetables. And while doing so, we must have also made a decision about what kind of meat to attend to, or whether to classify candidate plants as “vegetables.” If these processes are constitutive of understanding grocery price, then the fact that we can’t fluently or transparently estimate a value shows us that the opacity of understanding is routine.

To dramatize the human memory/attention limits involved in understanding any complex topic, consider routine efforts to understand genetic risks. It is quite common now for people to explain, and so purport to understand, a medical condition in terms of “having the gene for it.” Perhaps the local character of causation creates the impression that its forces are easier to cognitively process. However, most conditions that have clear genetic influences (say, breast cancer) aren’t just a result of a single point mutation but instead are related to hundreds of single-nucleotide polymorphisms (SNPs). Each mutation alone is not going to give you breast cancer, but as a whole, the aggregation of many relevant mutations can strongly affect risk. I think there really might be some human memory/attention limits involved with failing to understand this, because the news media and personal genetic profiling companies always seem obsessed with finding “the” [insert disease] gene, even though it is very rarely that simple.

It may be that philosophers, too, suffer from an errant fluency in their construction of theories of understanding. Even a few stops on a tour of philosophical theories of understanding reveals a troubled landscape: Definitions of understanding that appeal to grasping, pronouncements that understanding is nonpropositional, transparent. Compared to a psychological theory that traces the temporal and spatial contours of attention, the powers of memory, and the chemistry of encoding—all features of a complete scientific theory of understanding—distinctly philosophical approaches to theorizing about understanding seem proudly medieval.

Consider a more distinctively empirical way that an approach to understanding might take. Begin with situations that have similar structure. You

build a kind of model of some phenomenon. This model or schema incorporates the forces thought to be most important to the effects considered. If it is a weather pattern, the structure will include such factors as wind turbulence, dew points, and temperature. If the system or model concerns the stress on a steel beam, the structure will include the strength of the material, the length of the run, and the weight it bears. If it is crime in an urban neighborhood, you will focus on the kind of infraction, the demographics of the perpetrators, and the temptations in the environment that transforms onlookers into criminals. The motion of a projectile may be depicted by a model that includes forces that propel the projectile and produce recoil, factors like gravity that operate on its weight and air that creates drag.

If understanding seems a difficult achievement, at least in some domains, it may be because some phenomena in the world are complex in ways that challenge the mind's delicate balance of attentional gaze, perceptual buffer, and memory capacity. So we simplify. We conceive of an atom as a solar system, or electricity as a liquid. This abbreviation is, of course, a mnemonic, not an accuracy-inducing device. But without it, understanding even something like electricity would be difficult for most of us. Now think of understanding the operation of complex systems—population ecology, evolution, the psychology of morality, disease, poverty, and consciousness. What would be required in order to say that we “understand” any of these systems?

## 12.4. Philosophical Fluency

When a concept is accessible to our thoughts, like a triangle, a dog, or for the physicist, an electron, we can use it fluently. We can talk about the concept, hold it before our attention, relate it to others, and gauge its many powers. This experience of fluency, gives rise to confidence in a judgment or decision. Psychologists call this experience by a variety of names: Feeling of Rightness, Feeling of Knowing, and Judgments of Learning, and this experience is a memory-based judgment, and is based on experiences that are cued automatically. But the causes of these experiences—the processes of attention, and the mechanisms of memory workspace—are not introspectable (Brewer and Sampaio 2006; also see Koriati et al. 2004; also see Matvey et al. 2001). Thus, the experiential component of these Feelings—of Rightness, Judgment, and Learning—leaves one with a feeling of confidence without knowledge about the basis of that confidence.

As an indicator of truth, how reliable is the Feeling of Understanding? It is a good thing that difficult problem-solving settings provoke our analytic,

deliberative skills, because if left unchecked humans will succumb to overconfidence and hindsight biases to embrace the subjective cue that a given task is easier than it actually is. The sense of understanding feels good. We rush to get the feeling, and we are hooked on it. But it is a tragic addiction. Stephen J. Gould viewed it instead as comical farce, as though we have a homunculus in our head, jumping up and down, and shouting out the wrong answers (Gould 1991, 469). Whatever image we choose, these errors stemming from not actually understanding are large, and they are often about weighty matters.

Surely the feeling of understanding is at least sometimes right, even right much of the time. But in the common cases of knowing simple facts, we get no jolt of satisfying understanding. When I prepare for a left turn in my car, I have the belief that oncoming traffic will adjust, and that I have correctly judged the distance. I believe that eggs will cook if I heat them, that the local bodega around the corner is open, that I can barely ski, but that a stick frame wall built to code has plenty of strength to carry a heavy snow load in a temperate zone. We could continue this list indefinitely.

You might think that monitoring these biases is pointless, because their behavior is so difficult to control. But even mediocre performance can be an improvement. And some people have more room for improvement than others. People experience the Feeling of Rightness with different regularity, because they monitor and control their feelings differently.<sup>8</sup> Some are better at spotting and correcting errors. Did you catch the typo in the last sentence? If so, that is an example of successful monitoring. People who monitor their psychological processes better are typically also better on selected reading tasks (Lin, Moore, and Zabrocky 2001), better in science classes when given optical principles to learn (Prins, Veenman, and Elshout 2006), and better at mathematical problem solving (Desoete and Roeyers 2006; also see Lucangeli, Tressoldi, and Cendron 1998).

The extent to which we monitor our feelings determines whether we accept an existing outcome or seek another (Mazzoni and Cornoldi 1993; also see Son and Metcalfe 2000; also see Nelson 1993; also see Son 2004). If you feel that your swing is good, you may stop taking batting practice; if you don't, you may change stance. If you are confident that you have understood the above passage, you won't reread it. If you are confident that you have correctly remembered the name of a person you have just run into, you will address that person by name; if not, you may choose a more generic greeting. If you doubt

---

8. See Hertzog and Robinson (2005) for a review.

that you will remember the bread on the way home, you might arrange a cue to trigger your memory. The success of your judgment depends not only on the accuracy of your memory, but also on your ability to monitor your mental processes and take appropriate action (Koriat and Levy-Sadot 1999). This is the function of metacognition.

Philosophers have their own conceptual repertoire, and they traffic in it with great fluency. The courting of intuitions is one example, leading to judgments about “what we want to say” about a particular case. In response to the claim that understanding must be transparent, we might generate intuitions by imagining a mathematician working on a proof that the square root of 2 is irrational. She works at it so persistently that by the time she solves the problem she is too tired to appreciate the solution. She falls asleep and when she awakens she looks at the solution she had written. The proof is perfect, but the last series of steps in reasoning are unrecognizable to her. We might want to say that she solved the problem but doesn’t understand the solution. She doesn’t understand it because she doesn’t have an awareness of the factors responsible for the solution. Appeal to awareness has the satisfying sound of transparency, but it fails to move the ball. We might understand how to ride a bike but not have an awareness of the factors responsible for this skill. Proliferating senses of ‘awareness,’ however, is no way to provide a unified account of understanding. The option seems to be a concept of understanding that captures only a small class of intellectual and practical achievements—the ones that exhibit a kind of transparent awareness of the principles in virtue of which the achievement is accomplished—or treat recognitional capacities or discriminative sensitivities too, as expressions of understanding. Understanding feels like it goes with awareness, because we expect people who understand a subject to be able to instruct others about it, and instruction typically requires that we use that awareness to articulate the factors that contribute to the subject of instruction. Control or manipulability feels like it goes with understanding, because understanding an object seems to require knowing its causal powers. If an object can surprise us, acting unpredictably or beyond our abilities to regulate its conduct, we would not say that we understand the object, however we might fear or mistrust the object.

But aren’t these all just expressions of conceptual fluency, a fluency that requires not correspondence but mere coherence? I develop this psychological view at greater length in *Wondrous Truths: The Improbable Triumph of Modern Science*. If we get the same good sense of understanding whether our belief is true or false, or our theory is good or bad, how does that feeling advance the theoretical cause? After all, the feeling supplies conviction

rather than truth, the kind of conviction that people have that going outside in cold weather with wet hair will make you sick (it won't), or that shaving will make your hair grow in thicker (it won't). But after thousands of hours of implementing familiar concepts, laypeople, like scientists, have a conviction that is fully automatized, even when their relevant beliefs are false. In "The Will to Believe," William James dramatizes the power that ritualized behavior has not just to sustain belief but to create it. As he says: "Go, then, and take holy water, and have masses said: belief will come and stupefy your scruples" (James 1896/1907, 6).

The processing benefits of fluent concepts is clear. Familiar concepts, like familiar methods, are used most easily when most familiar. When identifying objects from an array, people display fluency (through faster reaction times) for living categories like human faces (Rhodes and Tremewan 1996; also see Rhodes et al. 2001) as well as fish, dogs, and birds (Halberstadt and Rhodes 2000; 2003). But the drive is so strong to find the center that humans do it even with categories of nonliving objects, like color patches (Martindale and Moore 1988), and even artifactual objects like furniture (Whitfield and Slatter 1979), wristwatches, and automobiles (Halberstadt and Rhodes 2000; 2003). So robust is this drive, in fact, that objects that clearly do not admit of gradations receive this treatment from humans. Whole numbers are one such example: certain odd numbers are reported by people to be "more odd" than others (Armstrong, Gleitman, and Gleitman 1983). Who would have guessed that 7 and 13 are the oddest of odd numbers, and 15 and 23 the least odd of them? Or that 8 and 22 are the eveniest even numbers, 30 and 18 are the least even of the even numbers? And in science, just like in everyday life, explanatory prototypes that free up processing space in the brain are deemed more attractive, and more accurate, whether or not they actually are. Fluency feels good, and disfluency feels bad.

Something like understanding has a healthy epistemic role to play in theory testing and development. But this naturalistic theory of scientific explanation should also account for explanation's failures. I have described a number of the psychological sources of failure that lie behind our tendency to be seduced by the sense of understanding (also see Trout 2002).

Other attempts to specify effective conditions for the acceptance of an explanation underestimate the theory-dependence of these judgments. It may be that people are more likely to accept an explanation if it seems applicable to other settings of prediction and manipulation (Lombrozo and Carey 2006). But this is a deeply theory-dependent judgment, and if your theories are poor you are likely to identify poor explanations as useful. Surely this was part of

the unifying promise of supernaturalism, of alchemy, or of the humoral theory of health. There seems no merely instrumental or pragmatic measure, such as usefulness, that tells us when an explanation should be accepted, even if there are more robust theoretical standards that might recommend acceptance (for example, that a theory is mature and has unified a diverse range of phenomena).

Moreover, the understanding conveyed by a good explanation may be a community achievement. Except for the simplest of events, explanatory understanding is not essentially an achievement of an individual. And any alternative account of explanation that requires the transmission of a sense of understanding must address this criticism. My positive account of scientific explanation asserts that, as a contingent matter of fact, the only feature of an explanation that can render explanation epistemic is its systematic tendency to produce increasingly accurate theories. In effect, only explanations capable of sustaining theoretical progress are good explanations. This pronouncement may not help us to decide now, rather than in retrospect, which explanation to take seriously. But this is not the job of an account of explanation. A theory of scientific explanation should not attempt to predict the future history of science, but set out what scientific explanation is, and what standards should be met by a proper explanation.

In order to accord explanation the epistemic role it seems to play in successful theory selection in contemporary science, we must abandon our attachment to the comforting idea that the “sense of understanding” is a cue to at least a working version of the truth. But this will not be easy. Explanation is a backward-looking affair and thirty years of research on judgment shows both that people are not good at tracking how they are affected by knowledge of outcomes and that they are not good at admitting this limitation.

A distinctly philosophical analysis of explanatory understanding may include a role for the sense of understanding. But it is unlikely that this role will be a justificatory one. In the series of cognitive steps that lead to understanding, phenomenology is a latecomer. In the absence of independent evidence of its reliable role—that its presence covaries importantly with progressive findings—and in light of the psychological and historical evidence that it is an unreliable cue, the sense of understanding is not a promising route to genuine understanding.

In fact, no one has the vaguest idea how this phenomenology is related to getting things right, so it is a field ripe for exploration. Scientific realists can assign a robust role to objective factors in explanation—such as statistical and causal relevance—and value the contribution of explanation to scientific progress. If your focus is not balanced by a positive account of the sense of



understanding in a scientific theory of explanation, then it is easy to portray any criticism of the sense of understanding as the first step toward explanatory nihilism. But explanatory nihilism is surely premature.

On the one hand, the sense of wonder when explaining fills us with joy. But on the other, the illusion of explanatory depth shows that our joy may be misplaced. Why would evolution have designed us so that the sense of understanding feels good, and yet so that we regularly stop before we fully understand? Alison Gopnik has the first half of the answer (Gopnik 1998). Evolution has designed us so that we have a drive to learn. And what better way to design that drive than to have a body and mind that responds to the world with wonder, so that we want to find out how these wonderful things work? But why the paradoxical termination of inquiry prior to actually fully understanding? Why would evolution make us feel so giddy when we seem to be making cognitive progress, yet at the same time hinder our learning through phenomena like the Illusion of Explanatory Depth?

Of course, few people can accurately boast a complete understanding of the workings of a toilet, tumbler lock, or tectonic stress. Even the best scientists, if they are honest, will cry uncle at some point when probed about their understanding of their favorite natural phenomenon. The Illusion of Explanatory Depth, while sobering, does not resolve our question. How do we develop such beautiful and accurate theories when our understanding is so routinely incomplete and our attitude so robustly overconfident? And how can we develop a scientific account of understanding with the feeble instruments of conceptual analysis, no matter how confidently wielded?

## 12.5. Conclusion

If understanding begins in flights of wonder, the limits of human understanding bring it back to earth. The need for speedy processing imposes forbidding constraints on memory and attention, two of the chief faculties underlying understanding (Cowan 2005). Understanding requires a complicated choreography of moves whose execution limits the number of items in memory that can be constantly activated (Just and Carpenter 1992), the number of relations among items in memory that can be integrated (Halford, Wilson, and Phillips 1998), and the ability to mentally maintain the distinction between items in working memory (Oberauer and Kliegl 2001). Just like in juggling, the error profile is what psychologists call catastrophic; you drop one ball, you drop them all. The process that manages the juggling of attention and memory is metacognition, and metacognition has its limits too.

The evidence for the nature of understanding is empirical (even when the empirical evidence is anecdotal). You can certainly choose to ignore the scientific evidence, and generate categories tied to bygone traditions like introspectionism, respects of similarity associated with magical thinking, and theorizing based on ancient categories. You could use concepts like transparency, nonpropositionality, and grasping, notions unrecognizable to scientific cognitive and comparative psychologists. Philosophers are, of course, at leisure to use whatever methods and concepts they choose. But that is only to observe that philosophy has no certification procedure that prohibits the use of fruitless methods, and no barrier to entry whose standard of success is based on truth.

By now it should be clear that the sense of understanding, by itself, is an unreliable cue to truth. True beliefs may carry more or less of the sense of understanding, a sense of fluency or metacognitive control. But this does not add up to unifying diverse cognitive achievements.

It is convenient to have a name for the event or state when our struggle with a particular cognitive challenge strikes the proper combination of attentional capture, available working memory, accurate shortcuts, or heuristics to reduce the conflict. But we shouldn't become so smitten that we attribute to the nominal orthography 'understanding' a unifying power, or suppose we can squeeze a new pedigree out of an old mongrel.

## Acknowledgments

I would like to thank Maya Mathur for providing, and discussing with me, some of the key examples of opaque explanation provided in this chapter. Kathleen Parker provided several apt examples as well. Thanks to Clinton Neptune for detailed and useful comments, Paul Abela for discussion of several examples, and Michael Strevens for many conversations about the relations between explanation and understanding. Support from the Templeton Foundation is gratefully acknowledged.

## References

- Armstrong, Sharon Lee, Lila R. Gleitman, and Henry Gleitman. 1983. "What Some Concepts Might Not Be." *Cognition* 13: 263–308.
- Belopolsky, Artem V., and Jan Theeuwes. 2010. "No Capture outside the Attentional Window." *Vision Research* 50: 2543–50.
- Bishop, Michael A., and J. D. Trout. 2005. *Epistemology and the Psychology of Human Judgment*. New York: Oxford University Press.

- Blasi, Guiseppe, and Venkata S. Mattay. 2005. "Effect of Catechol-O-methyltransferase val158met Genotype on Attentional Control." *Journal of Neuroscience* 25: 5038–45.
- Brewer, William F., and Cristina Sampaio. 2006. "Processes Leading to Confidence and Accuracy in Sentence Recognition: A Metamemory Approach." *Memory* 14: 540–52.
- Cowan, Nelson. 2005. *Working Memory Capacity*. Hove, East Sussex, UK: Psychology Press.
- Dawes, Robyn M., David Faust, and Paul E. Meehl. 1989. "Clinical versus Actuarial Judgment." *Science* 243: 1668–74.
- De Regt, Henk W., and Dennis Dieks. 2005. "A Contextual Approach to Scientific Understanding." *Synthese* 144: 137–70.
- Desoete, Annemie, and Herbert Roeyers. 2006. "Metacognitive Macroevaluations in Mathematical Problem Solving." *Learning and Instruction* 16: 12–25.
- Diamond, Adele, Lisa Briand, John Fossella, and Lorrie Gehlbach. 2004. "Genetic and Neurochemical Modulation of Prefrontal Cognitive Functions in Children." *American Journal of Psychiatry* 161: 125–32.
- Elgin, Catherine. 2007. "Understanding and the Facts." *Philosophical Studies* 132: 33–42.
- Fonseca, Brenda A., and Micheline T. H. Chi. 2011. "Instruction Based on Self-Explanation." In *The Handbook of Research on Learning and Instruction*. Ed. Richard E. Mayer and Patricia A. Alexander. New York: Routledge, Taylor and Francis. 296–321.
- Geva, Ronny, Michal Zivan, Aviv Warsha, and Dov Olchik. 2013. "Alerting, Orienting or Executive Attention Networks: Differential Patters of Pupil Dilations." *Frontiers in Behavioral Neuroscience* 7 (October 14): 145. doi:10.3389/fnbeh.2013.00145.
- Gopnik, Alison. 1998. "Explanation as Orgasm." *Minds and Machines* 8: 101–18.
- Gould, Stephen J. 1991. *Bully for Brontosaurus*. New York: Norton.
- Grimm, Stephen R. 2006. "Is Understanding a Species of Knowledge?" *British Journal for the Philosophy of Science* 57: 515–35.
- . 2012. "The Value of Understanding." *Philosophy Compass* 7: 103–17.
- Halberstadt, Jamin, and Gillian Rhodes. 2000. "The Attractiveness of Non-face Averages: Implications for an Evolutionary Explanation of the Attractiveness of Average Faces." *Psychological Science* 11: 285–89.
- . 2003. "It's Not Just Average Faces That Are Attractive: Computer-Manipulated Averageness Makes Birds, Fish, and Automobiles Attractive." *Psychonomic Bulletin and Review* 10: 149–56.
- Halford, Graeme S., William H. Wilson, and Steven Phillips. 1998. "Processing Capacity Defined by Relational Complexity: Implications for Comparative, Developmental and Cognitive Psychology." *Behavioral Brain Sciences* 21: 803–31.
- Hertzog, Christopher, and A. Emanuel Robinson. 2005. "Metacognition and Intelligence." In *Understanding and Measuring Intelligence*. Ed. Oliver Wilhelm and Randall W. Engle. London: Sage. 101–23.
- James, William. 1896/1907. "The Will to Believe." In *The Will to Believe: And Other Essays in Popular Philosophy*. New York: Longmans, Greens and Co. 1–31.

- Just, Marcel Adam, and Patricia A. Carpenter. 1992. "A Capacity Theory of Comprehension: Individual Differences in Working Memory." *Psychological Review* 99: 122–49.
- Khalifa, K. 2017. *Understanding, Explanation, and Scientific Knowledge*. Cambridge: Cambridge University Press.
- Kitcher, Philip. 1989. "Explanatory Unification and the Causal Structure of the World." In *Scientific Explanation*. Ed. Philip Kitcher and Wesley C. Salmon. Minneapolis: University of Minnesota Press. 410–505.
- Koriat, Asher, Robert A. Bjork, Limor Sheffer, and Sarah K. Bar. 2004. "Predicting One's Own Forgetting: The Role of Experience-Based and Theory-Based Processes." *Journal of Experimental Psychology: General* 133: 643–56.
- Koriat, Asher, and Ravit Levy-Sadot. 1999. "Processes Underlying Metacognitive Judgments: Information-Based and Experience-Based Monitoring of One's Own Knowledge." In *Dual Process Theories in Social Psychology*. Ed. Shelly Chaiken and Yaacov Trope. New York: Guilford Press. 483–502.
- Kvanvig, Jonathan. 2011. "Millar on the Value of Knowledge." *Aristotelian Society Supplementary Volume* 85: 83–99.
- Lin, Lin-Maio, Dewayne Moore, and Karen M. Zabucky. 2001. "An Assessment of Students' Calibration of Comprehension and Calibration of Performance Using Multiple Measures." *Journal of Reading Psychology* 22: 111–28.
- Lombrozo, Tania, and Susan Carey. 2006. "Functional Explanation and the Function of Explanation." *Cognition* 99: 167–204.
- Long, A. R., P. M. Fernbach, and B. De Langhe. 2016. "The Influence of Subjective Understanding on Risk Perception." Working paper.
- Lucangeli, Daniela, Patrizio E. Tressoldi, and Michela Cendron. 1998. "Cognitive and Metacognitive Abilities Involved in the Solution of Mathematical Word Problems: Validation of a Comprehensive Model." *Contemporary Educational Psychology* 23: 257–75.
- Martindale, Colin, and Kathleen Moore. 1988. "Priming, Prototypicality, and Preference." *Journal of Experimental Psychology: Human Perception and Performance* 14: 661–70.
- Matvey, Greg, John Dunlosky, and Robert Guttentag. 2001. "Fluency of Retrieval at Study Affects Judgements of Learning (JOLS): An Analytic or Nonanalytic Basis of JOLS?" *Memory and Cognition* 29: 222–33.
- Mazzoni, Guiliana, and Cesare Cornoldi. 1993. "Strategies in Study Time Allocation: Why Is Study Time Sometimes Not Effective?" *Journal of Experimental Psychology: General* 122: 47–60.
- McGinn, Colin. 1991. *The Problem of Consciousness: Essays toward a Resolution*. Cambridge, MA: Blackwell.
- Meehl, Paul. 1986. "Causes and Effects of My Disturbing Little Book." *Journal of Personality Assessment* 50: 370–75.

- Nelson, Thomas O. 1993. "Judgments of Learning and the Allocation of Study Time." *Journal of Experimental Psychology: General* 122: 269–73.
- Oberauer, Klaus, and Reinhold Kliegl. 2001. "A Formal Model of Capacity Limits in Working Memory." *Journal of Memory and Language* 55: 601–26.
- Parasuraman, Raja, and Pamela M. Greenwood. 2005. "Beyond Heritability: Neurotransmitter Genes Differentially Modulate Visuospatial Attention and Working Memory." *Psychological Science* 16: 200–207.
- Prins, Frans J., Marcel V. J. Veenman, and Jan J. Elshout. 2006. "The Impact of Intellectual Ability and Metacognition on Learning: New Support for the Threshold of Problematicity Theory." *Learning and Instruction* 16: 374–87.
- Rapsomaniki, Eleni, Anoop D. Shah, Pablo Perel, Spiros Denaxas, Julie George, Owen Nicholas, Ruzan Udumyan, et al. 2014. "Prognostic Models for Stable Coronary Artery Disease Based on Electronic Health Record Cohort of 102,023 Patients." *European Heart Journal* 35: 844–52.
- Rhodes, Gillian, Sakiko Yoshikawa, Alison Clark, Kang Lee, Ryan McKay, and Shigeru Akamatsu. 2001. "Attractiveness of Facial Averageness and Symmetry in Nonwestern Cultures: In Search of Biologically Based Standards of Beauty." *Perception* 30: 611–25.
- Rhodes, Gillian, and Tanya Tremewan. 1996. "Averageness, Exaggeration and Facial Attractiveness." *Psychological Science* 7: 105–10.
- Son, Lisa K. 2004. "Spacing One's Study: Evidence for a Metacognitive Control Strategy." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30: 601–4.
- Son, Lisa K., and Janet Metcalfe. 2000. "Metacognitive and Control Strategies in Study-Time Allocation." *Journal of Experimental Psychology* 26: 204–21.
- Strevens, Michael. 2013. "No Understanding without Explanation." *Studies in History and Philosophy of Science* 44: 510–15.
- Trout, J. D. 2002. "Scientific Explanation and the Sense of Understanding." *Philosophy of Science* 69.2: 212–33.
- . 2016. *Wondrous Truths: The Improbable Triumph of Modern Science*. New York: Oxford University Press.
- Whitfield, T. W. A., and P. E. Slatter. 1979. "The Effects of Categorization and Prototypicality on Aesthetic Choice in a Furniture Selection Task." *British Journal of Psychology* 70: 65–75.
- Wilkenfeld, Daniel. "MUDDy Understanding." *Synthese* 194 (4): 1273–93.
- Wilkenfeld, Daniel, David Plunkett, and Tania Lombrozo. 2016. "Depth and Deference: When and Why We Attribute Understanding." *Philosophical Studies* 173: 373–93.
- Zagzebski, Linda. 2001. "Recovering Understanding." In *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. Ed. Matthias Steup. New York: Oxford University Press. 235–51.