

# Paying the Price for a Theory of Explanation: De Regt's Discussion of Trout\*

J. D. Trout†‡

---

In his discussion of “Scientific Explanation and the Sense of Understanding” (Trout 2002), de Regt (2004) argues that Trout “denounces understanding as irrelevant, if not dangerous, from an epistemic perspective. . .” (de Regt 2004, 98), and that Trout prefers “a thoroughly objectivist view, in which understanding has no epistemic function and explanations require only accurate theories” (100). I argue that this is a fundamental misinterpretation of the objectivist proposal of Trout (2002), and that it results from de Regt’s failure to address the distinction announced there between genuine and counterfeit understanding. De Regt also advances a pragmatic alternative to my realist account of understanding in scientific explanation, one that focuses on the skills of scientists and the intelligibility of theories. After supplementing my earlier account of genuine understanding, I argue that de Regt’s pragmatic account obscures the nature of scientific explanation, and is vulnerable on several additional fronts.

---

**1. Introduction: A Preliminary Misunderstanding.** The virtues of explanation are many. Explanations describe underlying causal mechanisms, unify disparate phenomena, clarify the reducibility of a domain, and elucidate relations of causal dependence. Virtually every theory of explanation also places a premium on the power of an explanation to produce understanding. In my article, “Scientific Explanation and the Sense of Understanding” (2002), I argue that the sense of understanding conveyed by many explanations is a poor cue to genuine understanding. We often have the feeling of understanding when we do not genuinely understand

\*Received March 2004; revised June 2004.

†To contact the author, please write to: Philosophy Department and the Parmly Hearing Institute, Loyola University Chicago, 6525 North Sheridan Road, Chicago, IL 60626; e-mail: jtrout@luc.edu; webpage: <http://www.parmly.luc.edu/~jtrout/>.

‡Thanks are due to two anonymous referees for very useful comments on this response. I am also grateful to the National Science Foundation for grant SES#0327104 that supported this research.

Philosophy of Science, 72 (January 2005) pp. 198–208. 0031-8248/2005/7201-0011\$10.00  
Copyright 2005 by the Philosophy of Science Association. All rights reserved.

(counterfeit understanding), and we often genuinely understand when we have no sense or feeling of understanding. In the former case, well-documented hindsight and overconfidence biases are routinely responsible for counterfeit understanding. In the latter case, theories of implicit learning account for understanding without the familiar phenomenology.

By prioritizing the phenomenology of understanding, philosophers of science risk a fundamentally psychological notion of explanation, thereby threatening its objectivity. The science of psychology reveals the sense of understanding as an unreliable cue to accurate explanation, and so naturalistic philosophers of science—whose title is, by now, gratefully redundant—must accommodate the frailties of judgment in their accounts of explanation and understanding.

In his “Discussion Note” concerning Trout 2002, Henk de Regt raises at least two crucial concerns about my argument (de Regt 2004). But we should first dispense with a looming systematic error in his interpretation of my position. In his Abstract, de Regt summarizes my position in the following way: “Trout denounces understanding as irrelevant, if not dangerous, from an epistemic perspective . . .” (98). Nowhere in Trout 2002 do I denounce understanding as irrelevant, or otherwise derogate it. On the contrary, I offer a number of clear examples in which understanding is important to theory construction. In addition, de Regt neglects the distinction between understanding on the one hand, and the *sense* of understanding on the other. I claim that the *sense* of understanding is irrelevant to the success of an explanation, just to the extent that the phenomenology fails to play a role in the explanation’s accuracy. The paper keeps separate the mere sense of understanding from understanding proper, and distinguishes between counterfeit and genuine understanding. Yet, de Regt states that “. . . Trout prefers a thoroughly objectivist view, in which understanding has no epistemic function and explanations require only accurate theories” (100). This misunderstanding is made worse by an unfortunate paraphrase of my statement of a necessary condition for good explanation. Quoting my claim that “when it comes to explanation, there is no substitute for simply being (approximately) right” (230), de Regt draws from it that “The sense of understanding does not do any epistemic work at all—so we had better produce accurate theories without bothering about understanding” (100). But the latter is far too dramatic and sweeping a rendition of the former. Success can have many necessary conditions, and so mentioning one (such as being right—which can happen by accident) does not exclude others (such as bearing an appropriate cognitive relation to the evidence). Because people are intimately involved in the explanatory relation, the cognitive relation that the person bears to the theory (or model) and the world is an important one, as my discussion of implicit learning makes clear.

**2. De Regt's Criticisms of Trout 2004: A Weak and Narrow Account of Explanation.** We can now proceed to the substance of de Regt's concerns about Trout 2002. There seem to be at least two:

1. The objectivist conception of explanation, according to which an explanation must be approximately true in order to be a good explanation, provides a necessary but not a sufficient condition for understanding and accordingly is too weak. One should at least be able to use this knowledge to deduce a prediction (explanation) of the occurrence of the phenomenon.
2. The extremely subjectivist conception of understanding is unduly narrow and is a straw man, and so, my position leaves untouched many other senses of 'understanding'.

In response to the first concern, and in the interest of narrative fidelity, it should be made clear that I say just that—approximate truth is a necessary condition for good explanation. I say nothing about what is required for sufficiency, as I was not attempting an analysis of 'good explanation.' It is not surprising that, if an analysis were the perceived objective, my assertion makes for an analysis that is too weak.

Even so, de Regt illustrates his concern in the following way:

In Trout's discussion of understanding why jets fly, this ability to use the available knowledge in an appropriate manner is missing. In his objectivist view, it is only the logical relations between the pieces of information at hand that count. But if this information is to provide understanding, a subject is involved as well (indeed, Trout introduces himself as the subject: "I understand why jets fly" (222)) and this implies that not only knowledge of principles and background conditions but also particular skills of the subject are involved in the process of achieving understanding. This introduces an element that is *pragmatic*, because it concerns "the use and usefulness of the theory" (van Fraassen 1980, 88) and "requires reference to the persons involved in the process of explaining" (Hempel 1965, 425). (101)

This distinctly philosophical approach sacrifices proper silence for premature conceptual completeness. When our concept of some object, state, property, or process is incomplete, we might attempt to gain clarity in any number of ways. One way would be to define the item by brute force, linguistically legislating its application by introducing necessary and sufficient conditions. For many items with natural definitions, this effort to philosophically sanitize them, however noble, is by now wearily inappropriate. Many items that play a taxonomic role in science do not have necessary and sufficient conditions, and many concepts deployed in a developing theory, such as a naturalistic theory of explanation, or psy-

chological theory of understanding, are at any given moment incompletely defined. The correct reaction to an emerging concept is not to reject it unless it meets certain favored conditions imposed by philosophers (such as usefulness). For, in the end, the world wins. Pretheoretic standards simply sap the resources of advancing research. And these pretheoretic standards descend from a poor descriptive theory possessed by philosophers. If the practical goal is to describe the causal factors contributing to understanding, then we should be looking at accounts of understanding presented by scientifically trained psychologists, not anecdote-bound philosophers.

It is tempting to suppose that the issue that separates de Regt's view from my own is the familiar one between the naturalistic and normative. As the standard taxonomy would have it, my account of understanding and its role in scientific explanation is naturalistic and (therefore) descriptive, and de Regt's account is distinctly philosophical and (therefore) normative. But this taxonomy has only tradition in its favor. This way of carving up the options both underestimates the normative force of scientific inquiry and overestimates the descriptive powers of philosophical observation. Naturalistic theories are often implicitly normative, and many endeavors thought to be normative are based on objectively horrible descriptive theories. The issue that divides us, then, is not the standard distinction between a naturalistic theory of understanding that is at bottom empirical and scientific, and a normative approach that prescribes criteria for understanding by appeal to philosophical principles. The difference is really between an account that is intellectually responsible, and one that is not.

In response to the second charge—that I am attacking a straw man—we might note that Trout 2002 cites and criticizes at length orthodox work on explanation, and those quite standard positions assign a role to the sense of understanding. This fact should suffice to demonstrate that my target is not a 'straw man' position. According to these accounts, the sense of understanding is surely not the only factor that recommends an explanation, and I don't present them as such. (Indeed, the successful use of models is an expression of the appropriate form of understanding. I emphasize this point in footnote 1 of Trout 2002, so my account includes more than just approximate truth.) But the sense of understanding is an important component of each of these orthodox theories.

Perhaps it is the subjective aspect of the sense of understanding that concerns de Regt. After all, the 'sense of understanding' is indeed a subjective state of a person. But it is by no means an uncommon or unfamiliar one. Anyone who has experienced a state of certainty and clarity about any issue—and everyone has—has had a sense of understanding. And, when they turned out to be wrong, they have experienced the *mere* sense

of understanding—with its characteristic overconfidence and low fidelity. Clearly, my claim is not that all extant theories are defective because they all officially endorse an extremely subjectivist conception of understanding. The point, rather, was to show that, to the extent that existing theories of explanation depend (tacitly or explicitly) on a subjective notion of understanding, they are using that sense of understanding as a cue to genuine understanding. Given that the very history that philosophers of science recruit as evidential shows that this inference is treacherous, we must avoid reliance on this ‘sense of understanding’, in favor of more objective measures of accuracy and promise, such as converging lines of independent evidence, and explanatory unification.

**3. Understanding and Intelligibility.** My paper, then, was designed to address a simple issue. If understanding is being in the right cognitive relation to the evidence, then the phenomenology—namely, this ‘sense of understanding’—should not matter, unless the phenomenology plays a significant role in getting you to that relation. An explanation is often said to convey understanding, and this understanding is thought to play a role in both developing a theory and in designing the instruments to test it. Now, this instrumental ability you may call ‘intelligibility’, but in normal circumstances, a theory’s intelligibility is closely associated with its ability to convey understanding. For example, in ordinary settings, when a theory defies understanding we refer to it as ‘unintelligible’, or when we make it easier to understand we refer to that process as rendering the theory intelligible. So you cannot avert these problems by simply shifting the subject from understanding to intelligibility.

If we ask what factors align us with the evidence, there is nothing special about the phenomenology of the experience. De Regt states that I remain “silent about what precisely such an objectivist account of explanation would have to look like” (100). But this is not exactly right. In Section 3, “Understanding, Learning, and Transparent Capture of Information,” I consider a number of cases in which people clearly respond reliably to complex information sources. They internalize a model of the process, or implicitly learn the structure of the objects, without the phenomenology regularly advertised by appeals to explanatory understanding. Unfortunately, de Regt fails to address the objectivist accounts of implicit understanding (the “internalization of the argument patterns,” to use Philip Kitcher’s (1989, 438) apt phrase)—and with it, the internalization of theoretical structure.

By appealing to a piece of conceptual analysis performed by Wittgenstein, de Regt objects that my account fails to capture a Wittgensteinian contention about what ‘understanding’ means (101). But my account is neither based on, nor attempts to deliver, a semantic analysis of ‘under-

standing', still less an antinaturalistic semantic analysis of the sort favored by Wittgensteinians. Moreover, the fact that Wittgenstein defined understanding in terms of *use* as well as knowledge is perhaps historically interesting, but it does little to illuminate the nature of an important cognitive relation such as understanding, even if it comports with our intuitions. For philosophers of science serious about the findings of science, there is a secure body of psychological research on skills and expertise. The question is not whether a proposed account of understanding fits Wittgenstein's intuitions, but rather whether there is a sound scientific conception of cognitive relations that accounts for the epistemic role of understanding. I am optimistic about the prospects here, but the course of theory development must attend to systematic empirical work on the topic. Understanding is, after all, a natural relation. Philosophers may have something to say about the matter, but they do not get to say what understanding is, independent of hard-won scientific evidence.

De Regt is surely right that I could have said more about the conditions of genuine understanding. There is a reason for my relative quiet. It is not clear that I need to offer a positive account of genuine understanding in order to advance my case. Consider, for example, that we sometimes enjoy genuine understanding. And, as Trout 2002 shows, the scientific evidence indicates that the *sense* of understanding routinely accompanies instances of very poor understanding. Therefore, if we want to document the unreliable features of the sense of understanding as a cue to genuine understanding, we need only show that there are many instances of counterfeit understanding prompted by a *sense* that we understand. We don't, for example, need to offer a semantic analysis of "understand," or otherwise set out a list of conditions for genuine understanding. By offering a positive account, one risks the reader confusing the plausibility of the critique with the promise of the positive account. With this risk in mind, however, we can oblige with a positive account.

Naturalistic theories are properly wary of analyses that provide necessary and sufficient conditions, in part because natural phenomena often possess their features contingently. But perhaps a suitably hedged description of conditions will be useful. Possessing genuine understanding depends upon the existence and character of a number of distinct factors. Accordingly, the sense of understanding is a reliable or diagnostic indicator of genuine understanding when the occurrence of that sense is non-accidentally related the following states of affairs:

1. the belief putatively understood is at least approximately true,
2. the agent has sufficient collateral theoretical knowledge or information relevant to that belief, and
3. the belief is produced by a reliable process, perceptual or cognitive.

Notice that genuine understanding does not require the sentimental, “Aha” experience, only an appropriate causal relation to the evidence.

**4. Realism and Pragmatism.** De Regt proposes to replace the realism evident in Trout 2002 with a form of pragmatism. The argument for this pragmatism is that understanding is, in part, instrumental: it is a means to using a theory. This makes usefulness a condition of truthful or accurate attributions of genuine explanation. This is supposed to vindicate the claim that scientists prefer an intelligible theory, because it is one they can use. Here the pragmatist tail is wagging the explanatory dog. Science does not elevate intelligibility as a virtue above all. Certainly you can have theoretical understanding without it thereby allowing you to use the theory to produce certain effects (think of string theory, theory of black holes, etc.) In one sense of ‘use’, this is trivially true. There is always some circumscribed use of theory, at the limit, just like there is a use for fiction. It exercises the mind, perhaps aids in the generation of new ideas, etc.

So the pragmatist twist, by itself, does not advance de Regt’s case. Perhaps it can, when run in tandem with the notion of intelligibility. But here again (in light of the epistemically unreliable relation between sense of understanding and genuine understanding), it is doubtful that anything is gained by introducing the notion of intelligibility into the discussion. The property of intelligibility, a necessary condition for scientific understanding according to de Regt, seems as difficult to define as understanding itself. Intelligibility is, alternately, “the positive value that scientists attribute to the cluster of theoretical virtues that help them in their use of the theory” (103), “a value that is projected onto the theory by scientists” (103), and “a context-dependent value related to properties of the theory but also to scientists’ skills” (103). De Regt concludes: “Hence, there exists an epistemically relevant, pragmatic sense of understanding which, while it is not purely objective, should also not be equated with the subjective, highly individual ‘feels right’ experience . . .” (103). Perhaps, but we still don’t have any idea what intelligibility is, so it is difficult to tell how it is different—if it *is* relevantly different, from the mere *sense* of understanding, distinguished by its phenomenology. There is at least this difference. The sense of understanding is a familiar psychological experience. By contrast, intelligibility, as used here, seems to be a property far less easily described—part psychological projection, part theoretical, and in some way related to context and skills. Intelligibility would not seem to form a natural class of experiences, but rather an unwieldy social product introduced to account for a category of scientific activity: explanation. As such, it doesn’t sound like a property that much illuminates a discussion of understanding.

De Regt closes with a series of challenges. He first questions the rele-

vance of my observation that many scientists were plainly wrong about a hypothesis and yet they experienced a ‘sense of understanding’ when they acquired their beliefs. De Regt replies that “the relevant question is not whether these scientists were wrong but whether they were irrational” (107). I disagree, for two reasons. First, in order to avoid portraying as irrational anyone holding a theory different from yours, standards of rationality are usually notoriously weak, inclusive, or otherwise generous; indeed, they are typically standards of coherence. And so on many accounts, you turn out to be rational as long as you are not actually insane. Second, it is likely that an account of rationality will presuppose a conception of understanding. So a naturalistic account of understanding of the sort I would defend will find that a person is rational if he or she bears an appropriate cognitive relation to the evidence, and not otherwise. This evidence may be in part social, so that it incorporates prevailing practices. As with changing the subject from understanding to intelligibility, there is little to be gained by changing the subject from understanding to rationality.

If the value you attach to explanation is so deeply indebted to the subjective sense of understanding, it may be tempting to suppose that without it you must abandon explanation altogether. But this is a forced choice between artificial options.

Nevertheless, de Regt asserts:

If Trout wants to abandon the sense of understanding because it does not warrant truth claims, he should throw out explanatory power as well. What is more, on the basis of the pessimistic induction from the history of science, we have to acknowledge that our current theories are most likely false, and, if we adopt Trout’s position, concede that we do not have any explanations at all. If this misinterprets his argument, it is because his lack of clarity regarding what his theory of truth is (if not correspondence) leaves him open to misinterpretation. (107)

There is a lot going on in this charge, but let me try to handle it briefly. First, I do not hold that we should abandon the sense of understanding (I assume de Regt means abandoning the sense of understanding *as a theoretical construct*—we could hardly abandon it as a phenomenological experience) on the grounds that it does not warrant truth claims. Indeed, I recommend abandoning it as a necessary condition for good explanation precisely because the truth claims it *does* warrant bear an unreliable relation to (approximate) truth.

Second, de Regt is entirely correct to point out that “false theories can also provide explanations” (107). Some of these theories even provide *good* explanations, especially when the false ones are approximately true

(de Regt seems to rely on a crude distinction between theories as *either* true or false). But such discriminating judgments will depend on determining how to individuate temporally overlapping explanations in the history of science, in the same way that realists have addressed questions about how to individuate theoretical entities across historical theories, and so determine continuity of reference. A cognate question in the area of explanation might be: How much proprietary theoretical machinery is essential to the explanation? I offered a realist account of the role of understanding in explanation, and its realist attachments were announced throughout. The pessimistic induction is a widely discussed problem for realists to which realists have responded at length. It presents no *special* problem for my account of the role of understanding in explanation, and I can find no reason for puzzlement or interpretive difficulty about this realist account. Whatever problems exist are familiar.

With de Regt, I see history as a crucial source of evidence in the philosophy of science. But I believe that the philosophy of science is a sensitive affair. When the population of events is large—as in the history of science—discussion of particular episodes can do little more than show that a certain kind of practice is not literally impossible (because it is historically actual), or that a certain position is not utterly unholdable (because it was actually held by someone). Given that finding an instance of nearly any view we like is inevitable, we are really facing a sampling issue when arguing, as de Regt does, that it is possible to reject scientific realism and yet believe that science supplies explanatory understanding. But allowing this is little concession for the scientific realist if one must mutilate the notion of explanatory understanding in order to both honor the evidence and reject scientific realism.

The core of de Regt's alternative holds that "Scientific understanding is not related to subjective feelings but to the skills of scientists and the pragmatic virtues of theories" (104). In one sense, this is trivially true. The ingenuity and training of scientists, and the instrumental value of a theory, must come close to exhausting the factors that could contribute to scientific understanding. So, from an epistemic perspective, there is little distinctive about the skills of scientists and pragmatic virtues. I too claim that scientific understanding (if we mean by that "genuine" scientific understanding) is not essentially related to subjective feelings. But I don't think the role of understanding is clarified by replacing it with the equally uncertain 'intelligibility' and then identifying isolated cases in the history of science in which scientists self-report that their appreciation for some pragmatic virtue was responsible for a breakthrough. Philosophers of science have occasionally sorted through the history of science, and this post-hoc search can always turn up a successful theory that is simple, elegant, or beautiful according to some dimension of evaluation. There

is no suggestion of a systematic prospective effect of aesthetic factors here, no indication that theories simpler along a certain dimension, or found more aesthetically pleasing by some social standard, are more likely to be true. If you were looking for a methodology that would predict either empirically adequate or approximately true theories, you would not do well rushing to aesthetic criteria.

There are a number of well-recognized problems with such post-hoc searches. In the first place, they don't offer a very severe test of the hypothesis that some pragmatic factor conduced to progress, technical or otherwise. Science rightly avoids the elevation of personal idiosyncrasy to intellectual significance. So it would seem an unpromising approach to pick a handful of aesthetic factors and argue, after marshalling historical detail, that an important discovery would have been impossible without them, and therefore aesthetic factors can, and sometimes do, play an epistemic role in theory-choice. Given the total population of events (however individuated) in the history of science, and the number that might be counted as discoveries, it would be surprising if you could *not* find intriguing cases that combined aesthetic criteria with theoretical success. Of course aesthetic factors can, and do, play a role in theory choice, but this is an actuarial fact.

It was for this reason that I cite Robyn Dawes at the close of my paper. As Robyn Dawes puts it, "The problem is that there is a many-many relationship between antecedents and consequences in the course of human life. As we retrospect, in contrast, we can create many-one relationships" (1999, 37). The same is true of defenses of pragmatic accounts that begin with an outcome in the history of science and show how a sense of understanding might have been instrumental to a discovery.

**5. Conclusion: The Price of a Naturalistic Theory of Scientific Explanation.** The scientific evidence offers a grim prognosis for approaches in philosophy of science that ignore actuarial characteristics of prediction and generalization (Trout 1998, Bishop and Trout 2005). Exactly what makes a theory intelligible is an interesting scientific question, but it is, in the end, a scientific question. Its proper study requires data of archival expanse, and the unleashing of an army of scholars on centuries of theories. There are similarly bleak prospects for approaches to the philosophy of science that disregard the voluminous scientific, psychological studies that have characterized the nature and function of skills or recognitional capacities. No prescientific, philosophical account of skills can go proxy for it.

An adequately naturalistic theory of explanatory understanding, and so of scientific understanding, comes at a price. In this case, the cost is at least equivalent to the burdens of constructing a theory of scientific

explanation that is itself properly scientific. We find the tools for this positive account not in the distinctly philosophical internal dialogues of Wittgenstein, or in the characteristically philosophical analyses of skills and contexts of practice, but in the scientific examination of the processes that underlie successful explanation. Genuine 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 as well. I have attempted to locate one source of failure in our seduction by the sense of understanding. The sense of understanding is produced by overconfidence and hindsight biases, frailties that are well documented in the scientific literature of cognitive psychology. By grounding this source of explanatory failure in the findings of science, this theory of scientific explanation pays at least part of the price of a naturalistic theory.

## REFERENCES

- Bishop, Michael, and J. D. Trout (2005), *Epistemology and the Psychology of Human Judgment*. New York: Oxford University Press.
- Dawes, Robyn (1999), "A Message from Psychologists to Economists: Mere Predictability Doesn't Matter Like It Should (without a Good Story Appended to It)", *Journal of Economic Behavior & Organization* 39: 29–40.
- de Regt, Henk W. (2004), "Discussion Note: Making Sense of Understanding", *Philosophy of Science* 71 (1): 98–109.
- Hempel, Carl G. (1965), *Aspects of Scientific Explanation*. New York: Free Press.
- Kitcher, Philip (1989), "Explanatory Unification and the Causal Structure of the World", in Philip Kitcher and Wesley Salmon (eds.), *Scientific Explanation*, Minnesota Studies in the Philosophy of Science, vol. 13. Minneapolis: University of Minnesota Press, 410–505.
- Trout, J. D. (1998), *Measuring the Intentional World: Realism, Naturalism, and Quantitative Methods in the Behavioral Sciences*. New York: Oxford University Press.
- (2002), "Scientific Explanation and the Sense of Understanding", *Philosophy of Science* 69 (2): 212–233.
- van Fraassen, Bas C. (1980), *The Scientific Image*. Oxford: Clarendon Press.