

Review Essays & Reviews

Philosophical Explanations. By Robert Nozick. Cambridge, Mass.: Harvard University Press. 1981.

I, too, seek an unreadable book: urgent thoughts to grapple with in agitation and excitement, revelations to be transformed by or to transform, a book incapable of being read straight through, a book, even, to bring reading to stop. I have not found that book or attempted it. Still, I wrote and thought in awareness of it, in the hope that this book would bask in its light.

The opening sentences of Robert Nozick's *Philosophical Explanations* suggest that one is embarking on a book of uncommon scope and intention, one liable alternately—or simultaneously—to dazzle, bewilder, and edify. The expectation is amply realized in the 647 pages of text and additional 102 pages of notes. With extraordinary verve and ambition, Nozick embraces in six long chapters problems basic to the philosophic enterprise: identity of the self, why there is something rather than nothing, the nature of knowledge and its challenge by skepticism, free will, value, and the meaning of life. Each of these major topics subsumes dozens of separate investigations, the whole being peppered with digressions and asides.

The intellectual patrimony from which it draws is correspondingly vast. Not surprisingly, the references display easy familiarity with work in the Anglo-American analytic tradition, including its mathematicized variants. Here, though, Hempel and Kripke brush shoulders with Hegel, Heidegger, and Fichte—and assorted rabbis, yogis, evolutionary biologists, psychologists, physicists, aestheticians, and comics. In this case, it is more than a tired cliché to affirm, "There is no other book quite like this."

A mosaic so bold and sprawling cannot be adequately viewed by aiming a light at a few of the pieces that make it up. While each might be singly lustrous, it is the interconnected patterning of the parts that demands attention. Several of the issues addressed by Nozick will be discussed below, and an attempt will be made to identify motifs that interweave their way through the book. But even more than is customary with book reviews, inadvertent misrepresentation lurks. A synopsis may pick out some of the extraordinary

technical accomplishment of *Philosophical Explanations*, but how is it to reflect the exuberant high spirits that animate the book? Philosophizing as encountered here is a joyous activity; grizzled professional philosophers will be reminded of what originally impelled them into their chosen field.

EXPLANATORY RISK-TAKING

The book takes its title from methodological considerations raised in the introduction. Nozick rejects a conception of the business of philosophy as constructing proofs based on self-evident premises that generate inescapable conclusions. He dubs this model “coercive philosophy” and rejects it both because it promises more than it can deliver—one’s interlocutor can either deny a premise or simply walk away from the fray—and because forcing another to believe against his will is morally questionable. The alternative to proof or argument that Nozick holds forth is *explanation*, which he defines as showing how something S is possible, given other facts that apparently exclude S from obtaining. For example, how can it be possible that I know I am sitting in my office at my desk given the skeptic’s possibility that I am a disembodied brain in a vat being stimulated by a mad scientist? A Nozickean explanation will not aim at refutation of the skeptic but rather at the giving of an account that would render understandable how I can know that I am in the office *and* how the skeptical rejoinder can have the unsettling power we feel it to have. *That* S is true is the burden of argument; *how* S can be true is the query/quarry of explanation.

Nozick seems to be proposing a fundamental alteration in the way philosophy is done; yet the distinction between coercive proof and noncoercive explanation seems too tenuous to carry the weight of any decisive shift. It is, of course, only in a metaphorical sense that we can speak of arguments as coercing anyone, and it is doubtful that the metaphor carries as much conviction in philosophical practice as Nozick would have it. True, one speaks of arguments as *powerful*, *forceful*, even *knockdown*, but also as *persuasive*, *attractive*, *eloquent*. The latter are terms of seduction, not rape. And although philosophers are not unmoved by the allure of changing others’ minds, few respond to recalcitrance with cold fury. Philosophical conversations are voluntarily entered into by those who so choose, typically with the understanding that universal and enduring consensus is the outcome least likely to emerge. It is passing strange to see the author of *Anarchy, State, and Utopia* construing *coercive* so broadly.

The history of philosophy provides few instances of individuals claiming to have discovered incontrovertible arguments that lead from self-evident premises to unshakable conclusions, and these few are not necessarily to be taken at face value. (The strategy *is* to be taken at face value within mathematics, but Nozick offers no strictures against a coercive mathematics of axiomatization and proof. Indeed, the book features a number of mathematical proofs. Is this, too, “[not] a nice way to behave toward someone?” [p. 5]) Thoroughgoing foundationalism is so little practiced and so often criticized that a condemnation of coercive philosophy seems moot.

I find the introduction to be doing something rather different from what is advertised. Nozick’s quarrel with contemporary analytic philosophy is less one

of *method* than one of *range*. A lingering heritage from positivism is the reluctance to confront directly global conundrums on which hinges our conception of ourselves as valuable and precious. These are the puzzles that motivate the initial attraction to philosophical reflection and that underpin the world's great religious traditions. Analytical philosophy may have moved beyond issuing manifestos that rigidly segregate sense from nonsense, but discomfort with untidily large questions still constricts its practice. The novice who seeks enlightenment at one of our major universities will be taught that investigation is to proceed by way of manageably small units, and that it will most often take the form of precise scrutiny of language. Issues that are not amenable to treatment in this fashion carry within themselves the seeds of incoherence, and the inclination to pursue them will have been sublimated well before the doctoral dissertation is attempted.

Analytic philosophy is not coercive; it is risk-averse. The outcome it most dreads is foundering in a sea of enigmas from which reason cannot plot a confident egress. Nozick, however, obviously delights in risky activity. He is unwilling to abide by constraints that counsel caution except where a clear line of progress is foreseen. Explanation, as he wields it, eschews the safety of solid premises and methodical reasoning therefrom. Instead, it posits a way of viewing the world that *might be true* and that charms in virtue of its responsiveness to philosophical perplexity. It is not the case that within the context of an explanation, anything goes; Nozick need not apologize to anyone for the analytical rigor of his arguments. An explanation's risk quotient derives not from slapdash syllogizing but rather from the tenacity of questions entertained and the willingness to suspend fixation on the truth of one's premises.

Can one, though, be so noncommittal—some will say cavalier—about the *truth* of one's speculative constructs? Foundationalism is a red herring, but what raises concern is whether this exercise in free-form creativity is continuous with the disciplined search for truth by way of rational reflection that has traditionally gone by the name "philosophy."

Nozick explicitly addresses this concern in the last two pages of the book, but his statement that "philosophy must be true enough to the world" (p. 647) seems deliberately designed to keep this particular cauldron bubbling. Perhaps that is a tactically sound move. If explanation is to be validated, it will not be through metatheoretical considerations but through the ability of particular instances to enliven philosophical activity. That is, Nozick's most important advocacy for this conception of philosophy is implicit in the discussions he offers. If they elevate, captivate, impress, animate—and perhaps jog one closer to significant truths—then explanation is vindicated; otherwise, it is not. Therefore, I now turn to those discussions.

THE CHAIN OF EXPLANATION

Chapter Two, "Why Is There Something Rather than Nothing?" is perhaps the one most representative of Nozick's enterprise. The question appears bleakly unpromising. Anything that figures in an explanation is itself something that potentially stands in need of explanation, and so a noncircular *complete* explanation, one that leaves nothing unexplained, is impossible. If the chain of explanation is nonterminating, then there is no final explanation

in terms of which all else is intelligible. If the chain is finite, then there is a last link that is itself unexplained.

Nozick's crucial move is to press the investigation one step further by considering more closely whether all self-explanation is perniciously circular.

E_1 : p because p

is clearly unsatisfying if put forth as an explanation of p. But consider a quantified proposition of the form:

E_2 : Every statement having characteristic C is true.

If E_2 itself has characteristic C, then E_2 is true in virtue of being an instance of itself. The intuitive distinction between E_1 and E_2 as self-explanatory principles is that the latter possesses a kind of logical depth that the former does not. E_2 as *explanans* is at a different level than E_2 as *explanandum*, and thus self-explanation need not have the feel of uninformatively standing in the same place.

The intuitive case for self-explanation via quantification theory stands in need of considerable sharpening and refinement. Nozick admits that self-subsuming explanation appears strange; but rather than taking that as a mark against the strategem, he counts it in its favor:

The question [of why there is something rather than nothing] cuts so deep, however, that any approach that stands a chance of yielding an answer will look extremely weird. Someone who proposes a non-strange answer shows he didn't understand this question. Since the question is not to be rejected, though, we must be prepared to accept strangeness or apparent craziness in a theory that answers it. [P. 116]

It is only to take the author at his word to agree that the subsequent 50 pages contain an ample quota of strangeness. Along the way he considers whether there might "be" a state that transcends both being and nonbeing, a "reality" (language inevitably stalls in such rarefied atmosphere) that neither is nor is not but that existence and nonexistence alike presuppose. This slides into a discussion of the epistemic status of mystical experience and concludes with what may well be the oddest footnote ever to grace a philosophical manuscript. (Revealing its content would be as unkind as the movie reviewer's giving away the plot of a whodunit. However, readers who might find themselves interested in possible connections among self-subsuming relations, Hatha yoga, the interpretation of esoteric texts, and auto-fellatio will do well to turn to pp. 163-64.)

Is this journey truly necessary? Nozick says that the question why there is anything at all is not to be rejected; but he provides, so far as I can detect, no reason whatsoever for that judgment. That is, even if he is persuasive in suggesting that a *summary* dismissal of the question is too abrupt, a more deliberately rehearsed dismissal may be precisely what is indicated. Indeed, some might take the waywardness of Nozick's rambles to be prime justification for just that course. And how can he say nay? To insist that the question

may not be rejected hints more than a little of that coerciveness he has decried elsewhere.

Rather, the question is embraced because he *wants* to confront it; Nozick's proclivity for bearing risks is nowhere better illustrated. To take as an embarcation point the Vedic Hymn of Creation is not the stuff of which ordinary philosophical activity is made, but ordinary treatments of standard issues are clearly not what Nozick seeks. What he achieves is, however, another matter, one much harder to judge, at least for this reviewer.

Where a range of discourse has a firmly established tradition within a philosophical community, conceptual moorings have been frequently tested, and proposals have benefited from considerable prior criticism and modification, it is with relative confidence that one can judge the merits of a suggestion that is novel yet basks in familiarity inherited from similar conjectures. As innovation becomes more radical, it becomes progressively more difficult to judge with any degree of assurance whether one is confronting a move that has high potential for continued development or instead a hopeless jumble. In at least this respect, philosophical judgment resembles aesthetic judgment.

For what it may be worth, I found the discussion of a realm beyond being and nonbeing mostly impenetrable—and would confess to being mystified were I impervious to the charge of reveling in a bad pun. Nor has Nozick stilled all doubts about the explanatory value of self-subsuming propositions. The metaphor of depth has some resonance, but whether it can intelligibly be construed as providing the room between *explanans* and *explanandum* that any bona fide explanation must have (and that “p because p” blatantly lacks) remains unestablished.

Even a cursory acquaintance with the semantic paradoxes will prompt uneasiness concerning the coherence of a proposition explaining itself through an endless cascade of levels. An analogue to explanation of everything might be a map that maps everything and thus maps itself, thereby mapping itself mapping itself, thereby. . . Or is the appropriate analogy that of a city map that, in virtue of being isomorphic to itself, also maps itself? If so, is every map—is every *thing*—a map of itself? Reflexive mapping seems to involve a breakdown in our concept of what a map is. Perhaps the same is the case for reflexive explanation. Here, however, intuitions also run in the opposite direction: a complete theory of grammaticality in English may be written in English sentences whose grammaticality is explained by that theory. If this isn't incoherent, then perhaps neither is Nozick's explanation (more accurately, package of alternative explanations) of why there is something rather than nothing. I have little confidence in my ability to judge.

With very great confidence, however, I can assert that various of the digressions and byroads of the chapter are splendid. The examination of mystical experience is the most important offered by any philosopher since William James. Nozick's analysis of the distinction between inegalitarian and egalitarian theories amounts to a solid contribution to the theory of explanation. And his skill in utilizing formal relational properties such as reflexivity, self-subsumption, and iterated structures pays philosophical dividends throughout the remainder of the book. This chapter resembles the NASA moon landing venture: the overriding national purpose—getting there before the Russians—may have been of questionable worth, but several of the spinoffs are undeniably positive.

IDENTITY AND KNOWLEDGE

The chapters "Identity of the Self" and "Knowledge and Skepticism" are those that will seem most familiar to analytical philosophers. Standard problem cases and ingenious variations on them are introduced, the literature surveyed is culled much more heavily from philosophy journals than from *Midrash Rabbah* or Vedantic hymns, and the explanations proffered look much like old-fashioned philosophical arguments. That is not to deny originality to these contributions but rather to note that originality is displayed within recognizable forms. In both chapters the results are dazzling. I shall be brief in my scrutiny of them because they are certain to become a touchstone for further philosophical explorations during this century and well into the next.

A vast literature has accumulated since Locke on the criteria for personal identity. Nozick uses it creatively as the jumping-off point for his own proposal, the closest continuer theory. Briefly, it holds that I at time t_1 am identical to the person at t_2 who is the closest continuer of myself at t_1 provided that the person at t_2 is "close enough" to the person I am at t_1 . This is the bare framework of a theory; it is fleshed out by specifying how close is "close enough," what dimensions count in the evaluation of closeness, and how relative weights are to be assigned along these dimensions. It is notoriously the case that any particular way of specifying what should count in judging identity over time, especially what should count *decisively*, is susceptible to problem cases. Nozick's own proposals are no exception, but he is particularly persuasive in elucidating why characteristic problem cases are felt to challenge conceptions of identity.

Were it not the case that we insist on viewing ourselves as uniquely precious and that each recognizes reason for special care about *the future self that will be he*, there would be no acute dilemma of personal identity. That issue would merge into the general problem of identity conditions for temporally extended entities. The identity of the continually rebuilt ship of Theseus requires a theory that is consistent and otherwise logically tight (transitivity, etc.); almost any such theory might do. But personal identity is further constrained. It is not enough that a future being just *turn out* to be me rather than someone else; in a deep sense, the identity ascription cannot be contingent or a matter of arbitrary stipulation. For example, suppose that at t_1 you know that there will exist two persons at t_2 each close enough at t_2 to be you but scoring equally high on the closeness function. Which is you? It seems unsatisfactory to hold either: (1) neither is you, and so you have no reason to care specially about either one, though you would have had reason to care about the one you would have been identical to had the other not existed; or (2) both are you, although they are not identical to each other; or (3) you are identical to the one of them who meets some arbitrary criterion (e.g., being closest to the North Pole at t_2).

No theory can entirely avoid a whiff of the paradoxical when confronted with such riddles; Nozick's efforts are directed not so much at blowing away that whiff as at tracing its source to what it is about our future selves that prompts the caring relationship. Part of his response is a quasi-Fichtean analysis of the nature of the self in which it is created over time through its own acts of referring to itself. Reflexivity strikes again! Here, though, I find it strikingly successful; an adequate account of personal identity cannot construe

identity merely as a passively received endowment from without, but as something a person continually creates and recreates through acts of *identifying* himself with particular objects of care—including, but not limited to, himself. An “identity crisis” is not merely homonymously related to the metaphysical problem of personal identity, in spite of the fact that some familiar accounts leave room for no closer, no more interesting, connection. In short, this chapter *must* be read.

The discussion of knowledge and skepticism is also first-rate. Again, it displays command of the literature and great facility in manipulating problem cases. Nozick proposes that knowledge be understood as actual and counterfactual *tracking* of truth. At first acquaintance, the tracking explanation seems clever but appears, in managing hard cases, to offer no more than a marginal gain over an epistemological theory featuring a best-evidence criterion. That initial impression misleads; the full power of his proposal is revealed only when it confronts the challenge of extreme skepticism.

Stripped of complicating details, the tracking theory analyzes “S knows that p” as:

1. p is true
2. S believes that p
3. If p weren't true, S wouldn't believe that p
4. If p were true, S would believe that p.

Condition (4) requires some clarification. It is not good enough for knowledge that S in fact believe (the true) p; it must also be the case that had circumstances been *slightly* different (the emphasis is crucial), S still would have believed that p. This can be phrased: in those possible worlds close to the actual world and in which p is true, S believes that p. A similar understanding is given to (3): in those possible worlds closest to the actual world in which p is false, S does not believe that p. This is labeled *tracking* because S's beliefs tenaciously track truth (and falsity) across various possible worlds.

Skeptical possibilities challenge knowledge at condition (3). If S were a brain in a vat being appropriately stimulated by the stereotypical mad scientist, S would believe p though p were not true. S doesn't know that he is not a brain in a vat (because he would have precisely the same beliefs he now has if he were/is a brain in a vat); ergo, S does not and cannot know that p.

The crucial move in Nozick's response is to deny that knowledge is closed under known logical implication. S may know that p, know that p entails q, yet not know that q. In the particular case, S knows that he would not know that p were a brain in a vat, S does not know that he isn't a brain in a vat, yet S still knows that p. The skeptical possibility can be granted, yet knowledge survives.

How this is so can be explained in terms of possible worlds. Condition (3) holds because in those possible worlds *closest to the actual world* in which p is false, S does not believe p. To be sure, in the possible world in which he is a disembodied brain, he does believe p, but that world is too distant from the actual world to falsify the counterfactual conditional expressed by (3).

This may seem unacceptably tricky; what is remarkable about Nozick's discussion is that reading it makes the decisive move against skepticism seem almost obvious, not even a slight trick. (Is prestidigitation raised to its highest degree when it seems so commonplace as not to be worth a second glance?) It

may actually be the case that he has carried out his project of showing that skepticism does not defeat knowledge, while simultaneously exhibiting the source of skepticism's Hydra-like power to intrude its disturbing doubts no matter how often it has been "refuted."

One may continue to fear, however, that the pacification of skepticism is a Pyrrhic victory. The tracking theory is content to swallow the result that one does not know, indeed *cannot* know, that one is not a brain in a vat. Knowledge is safeguarded only by conceding that knowing a fact is compatible with inescapable ignorance concerning the necessary conditions for its truth. This is, of course, consistent with Nozick's espousal of philosophical explanation as replacing proof; in both cases, fixation on the truth of one's premises is rejected. Yet for one who is persistently attached to concern for the *reasonability* of premises, that they not rest on thin air, tracking is insufficient to remove the skeptical barb.

VALUE

Many, perhaps most, readers of *Philosophical Explanations* will come to it because of their prior reading of *Anarchy, State, and Utopia*. Nozick's further exploits in political philosophy are what they seek, and they will first turn to the section on value, leaving until later the chapters discussed above.

They will be disappointed. Political philosophy receives a nod only in passing. The connection to the earlier book is predominantly negative: one of its most important reviews was titled "Libertarianism without Foundations."² The current work can be characterized as "Foundations without Libertarianism."

That is not to say that these three chapters are without interest; perhaps the one thing that Nozick could not do is write dull material. But the approach is apt to defeat expectations. Construction of a general theory of intrinsic value is Nozick's primary objective, and he takes the surprisingly old-fashioned route of defining value as degree of organic unity. What is decidedly not old-fashioned is the analysis of the formal properties of value and of the act of valuing value. This is done at a very high level, and if taken simply as a derivation of necessary conditions that must be met by a general axiological theory, it makes a substantial contribution. But Nozick attempts more than an exhibition of the formal structure of value; he wants to argue that the unification of diversity best exhibits those requisite formal properties, that it (largely) suffices to provide the *content* of value. He fails, and I think fails by a wide margin, to convince, because what counts as relevant unities remains mostly opaque.

For example, Nozick confronts the objection that, on his account, a concentration camp emerges as intrinsically valuable, because it collects diverse elements into a tightly organized unity. Not so, he replies; the purpose, the *telos*, of a whole is an important component of unity. A concentration camp aims at the destruction of valuable (i.e., organically unified) beings; therefore it possesses disvalue.

This seems *ad hoc*. Why isn't the destruction of low-level unities in the cause of an all-encompassing unity a net gain in value? Do we know that it isn't *because* we know that concentration camps are bad things? Suppose that we

think not of Hitler's concentration camps but of Pharoah's; people are enslaved but not destroyed, and they unify lots of sand and water into very impressive wholes. Shouldn't Moses have had more respect for *this* kind of unity? To shift gears a bit, how do we compare the respective degrees of unity in diversity of a spontaneous market order with a hierarchically structured planned economy? F. A. Hayek prefers the former, John Kenneth Galbraith the latter; who is right? Does the question even make sense if construed as an inquiry into comparative degrees of organic unity?

Nozick has, I think, allowed fascination with formal structures to overwhelm concern for applicability. The theory is presented at so abstract a level that it could be as easily wielded by an organic state collectivist as by a Nozickean libertarian. This is the *wrong* kind of universalizability to aim at in value theory! Nozick has parsed knowledge as tracking truth; he now wants to explain ethics as tracking bestness. The symmetry has undeniable allure, but while truth is anchored in the firm cement of the way things are, value has been left as otherworldly and intangible as a Platonic form.

Although the major projectile misfires, it throws off sparks that are incandescent. Nozick's analysis of Glaucon's challenge—"Show that being moral pays"—in terms of "ethical push" and "ethical pull" is valuable even if one rejects organic unity as the force exerting the push and pull. An uncommonly persuasive justification of retributive punishment is set forth, in which retribution is explained as a certain way of connecting the malefactor with correct values from which his past actions have "unlinked" him. I think that the analysis ultimately breaks down because justifiable retribution must rest on response to the flouting of *law*, not the more general flouting of value. Nozick has very little place in his retributive account for law, possibly because of his proclivity for state-of-nature theory, within which the justifiability of punishment is logically prior to the formation of civil society as a law-enacting body. Still, for anyone concerned with the theory of punishment, this is required reading.

Space is lacking for even a quick tour of the sections on free will and determinism and the meaning of life. Each is apt to prompt vigorous disagreement, but each will amply repay careful reading and rereading. If the chapters on value disappoint, they do so only relative to Nozick's previous work, his superb achievements in the metaphysics and epistemology chapters, and his own professed aims. Judged against more tolerant standards, they are very good indeed.

RECOMMENDATION

Self-pity is an unattractive trait in a reviewer. Laments that it is impossible to do justice to so vast and sprawling a book, one that mines deeply diverse disciplines and traditions, are likely to be met with sighs of indifference. Perhaps, then, an appeal to the reader's own self-interest will be more effective; *Philosophical Explanations* more than most books exceeds the sum of its parts. Experiencing it second-hand is like reading a menu in place of eating a meal. Because several of its discussions are certain to shape the way philosophy is done in succeeding years, reading it is a high-return professional

investment. But it is also an opportunity to experience someone outstandingly good at what he does, doing it with imagination and unbounded enthusiasm.

LOREN LOMASKY

*University of Minnesota,
Duluth*

1. Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), especially pp. 262-65 on voluntary exchange. However, it appears as less of a deviation when compared with his "Coercion" in *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, ed. Morgenbesser, Suppes, and White (New York: St. Martin's Press, 1969) pp. 440-72.
2. Thomas Nagel, "Libertarianism without Foundations," *Yale Law Journal* 85 (1975): 136-49.