

ROBERT NOZICK

INVARIANCES

The Structure of the
Objective World

THE BELKNAP PRESS OF

Harvard University Press

Cambridge, Massachusetts

London, England

2001

Copyright © 2001 by Robert Nozick
ALL RIGHTS RESERVED
Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Nozick, Robert

Invariances : the structure of the objective world / Robert Nozick

p. cm.

Includes bibliographical references and index.

ISBN 0-674-00631-3 (alk. paper)

1. Relativity. 2. Truth. 3. Objectivity. 4. Ethics. I. Title.

BD221 .N69 2001

110—dc21

2001035102

INVARIANCES

To Trude
The structure of my subjective world

Acknowledgments

Acknowledgments usually begin with intellectual debts, but I have reason to begin more personally here. In their order of appearance in this narrative: Marvin Kolodny, high school, college, and continuing friend, distinguished practitioner and professor of gastroenterology in Los Angeles, urged me to get to the bottom of some physical symptoms, and has provided excellent advice and continuing friendship; Andrew Warshaw, Chief of Surgery at Massachusetts General Hospital and Austen Professor of Surgery at Harvard Medical School, operated brilliantly six years ago, and again one and one half years ago, to remarkable effect; Robert Mayer, Professor of Medicine at Harvard Medical School, laser-minded brilliant clinician and chief of gastrointestinal oncology at the Dana Farber Cancer Institute, provided chemotherapy and radiation, keen medical judgment, good-humored cheer, and continuing excellent medical care (and even as this book goes to press, his knowledge and experienced intuition come to the rescue once again). Without the skill of these three men, this book, and its author, would not now exist. "Grateful" is not the word.

My wife Gjertrud Schnackenberg's love, dedication, and iron-willed determination that illness would not conquer have sustained me in these past few years and have continued to make my life wonderful in innumerable ways. I found that I had absolutely no desire or need to dwell upon the dangers to myself, and the writing of this book over the past seven years took up my spare attention. Consequently, the main emotional task fell upon my wife, who provided constant enormous love, care, and concern. Her own work of writing was seriously interrupted for several years, as mine was not, and I am overjoyed that the world now has her completed book, *The Throne of Labdacus*, which appeared last year. Readers of that magnificent work will understand how substantial are the mind and spirit that I live with daily, and how very privileged I am to do so.

To my parents' (Max Nozick and Sophie Cohen Nozick's) great love for each other and for me, I owe, ultimately, my optimistic view of life and of

its possibilities. Or perhaps it is to their genes. In any case, I have been the great beneficiary.

During the writing of this book, many people offered helpful advice and comments. Timothy Williamson, Dagfinn Follesdal, and Elijah Millgram gave me detailed written comments on the whole manuscript, to its great benefit. I also received helpful comments and criticism on portions of the manuscript from (in alphabetical order) Carla Bagnoli, Ned Block, Scott Brewer, Marshall Cohen, Leda Cosmides, Tamar Gendler, Gilbert Harman, Stephen Kosslyn, Eric Maskin, Thomas Nagel, Derek Parfit, James Pryor, Hilary Putnam, Thomas Scanlon, Martin Seligman, Susanna Siegel, Alison Simmons, William Talbott, John Tooby, and Bernard Williams. My discussions of physics in particular have also benefited from the comments of David Albert, John Earman, David Malament, Timothy Maudlin, Simon Saunders, Abner Shimony, Mark Steiner, and Ady Stern. Neither my philosophical colleagues nor these distinguished and technically proficient philosophers of physics (and physicists) are responsible for any errors the book contains.

Earlier versions of parts of this book were delivered in the spring of 1997 as the six John Locke Lectures at Oxford University. I was first invited to deliver the Locke Lectures in 1978, and I declined. It took almost twenty years before I had material that I was happy to deliver, and I am grateful to Oxford for being willing to reinstitute the invitation. Portions of this book also were presented in three lectures at Notre Dame University, two lectures at the University of California at Berkeley, University College Dublin, Edinburgh University, two at New York University, and in lectures at Dartmouth, the Harvard Kennedy School Ethics Program, the University of California at Los Angeles, the University of Florence, the University of Rome, and the Ecole Polytechnique. An earlier version of Chapter 2 was delivered as the Presidential Address to the Eastern Division of the American Philosophical Association in 1997. I am grateful to these audiences for their attention and for their questions.

While writing this book, I was the recipient of a fellowship from the John Simon Guggenheim Foundation, which I gratefully acknowledge. The manuscript of this book was completed during an intensive and productive two-month stay at the superb and extremely beautiful facilities of the Getty Research Institute for the Humanities.

Contents

Introduction: On Philosophical Method 1

I. THE STRUCTURE OF THE OBJECTIVE WORLD

1. Truth and Relativism 15

- Is Truth Relative? 15
- Who Wants Relativism? 21
- Truth in Space and Time 26
- The Truth Property 44
- Is a Theory of Truth Possible? 48
- Is Truth Socially Relative? 55
- Does Relativism Undercut Itself? 64
- The Correspondence Theory 67

2. Invariance and Objectivity 75

- Objective Facts 75
- Admissible Transformations 79
- Two Types of Philosophical Account 83
- The Ordering of Objectiveness 87
- Intersubjectivity 90
- Objective Beliefs and Biasing Factors 94
- Dimensions of Truth 99
- The Objectivity of Science 102
- The Functional View 106
- Underdetermination of Theory 111
- Rationality, Progress, Objectivity, and Veridicality 114

3. Necessity and Contingency 120

- Epistemology of Necessity 120
- Cross-Classifications 126
- On the Supposed Necessity of Water's Being H₂O 128
- The Withering of Metaphysical Necessity 133
- Explaining Away Necessities 136
- Logical and Mathematical Necessity 141
- Degrees of Contingency 148
- The Nature of Actuality 155
- The Ultimate Theory of the World 161

II. THE HUMAN WORLD AS PART OF THE OBJECTIVE WORLD

4. The Realm of Consciousness	171
The Function of Consciousness	171
Gradations of Awareness	174
The Context of Consciousness	176
The Zoom-Lens Theory	180
Synthesizing and Filtering Data	190
Common Knowledge	197
The Functions of Phenomenology	205
Mind-Body Relations	218
5. The Genealogy of Ethics	236
The Theory of Ethics	236
The Ubiquity of Ethics	238
Coordination to Mutual Benefit	240
Coordination via Ethical Norms	243
The Evaluation of Systems of Coordination	253
The Core Principle of Ethics	259
Normative Force and the Normativity Module	267
Evaluative Capacities	274
Higher Layers of Ethics	278
Ethical Truth and Ethical Objectivity	284
The Unpredictability of Human Behavior	294
Ethics and Conscious Self-Awareness	298
Notes	303
Index	403

Introduction: On Philosophical Method

Philosophy begins in wonder. How much of our view of the world is objective, how much is subjective? How much of what (we think) is true holds absolutely, how much is relative to our situation? Are truths only local or do some hold everywhere and always? Does ethics have an objective basis? Why are we conscious? What is the function of felt experiences in an objective world?

Previous philosophers, almost all of them, have sought to establish permanent truths in an enduring framework of thought; these truths were supposed to be absolute, objective, and universal. Details might remain to be filled in, but the essentials were meant to be in place and to stand firm for all time.

In mathematics, a fixed-point theorem states that all transformations of a certain type (for instance, continuous ones) leave at least one point in a certain space unchanged. The unchanged point is a fixed point of the transformation. It is not the point's intrinsic nature that leaves it fixed, however. Although every transformation may leave some point or other fixed, different points are left fixed by different transformations. No one point is left fixed by every transformation.

Philosophers, however, have sought points that remained fixed under all transformations, at least all the ones that were worth considering. Such points would constitute absolutely secure foundations for knowledge and values. The most prominent candidates for fixity were the *cogito* of Descartes, the sense data of the empiricists, and the necessary metaphysical truths of the rationalists. These points were supposed to be not only fixed but also fertile enough to support the rest of our knowledge and values. One by one, these purportedly certain, indubitable, and unassailable points have been shaken. And their fertility also was undermined. Little could be built upon them using only perfectly secure means of building.

The pragmatist Charles Peirce held, against Descartes, that nothing was indubitable. Each thing was open to doubt, although not all things at once.

Each thing could be doubted on the basis of other propositions that were, at that particular moment, not actually in doubt. Their turn could come later. Otto Neurath, followed by W. V. Quine, likened our situation to that of sailors who have to rebuild a ship at sea: while standing on somewhat rotten planks, they must repair and replace the others. Every plank sooner or later gets repaired. Everything is open to transformation. Nothing stays fixed. Even something as fundamental as the principle of noncontradiction can be open to questioning and to revision.¹

Critics have nibbled at the edges of the position that nothing is immune to being modified or given up. Aren't the criteria for revising views themselves unchangeable? Could we ever give up the weakest principle of noncontradiction, which says that not for *every* statement p , is it the case that both p and not- p ?² But even these objections are not decisive. True, it would take a serious intellectual crisis to motivate a change as far-reaching as holding it is the case both that p and that not- p , for *every* statement p , so it is difficult now to imagine specifically how such a change would be justified or plausible.³ Yet even if we cannot conceive of the details of a (justified) rejection of the weakest principle of noncontradiction right now, nevertheless a gradual process of change might lead to the giving up of any statement or criterion, including that very weakest principle, on the basis of other statements and criteria that have come to be accepted then. The fact that a particular theory has a fixed point does not show that this very same point must or will be held fixed by every plausible theory. And it does not show that the theory under consideration is unable successively to transform itself by using its own then accepted standards, so that this point and even these very standards get modified.

"But won't there be meta-meta- . . . -metacriteria that do not get transformed no matter what?" Even if such objections were accepted, their import would not be very great. You cannot get very far solely on the basis of saying that not every contradiction holds true,⁴ and not much can be built upon (only) such rarified metacriteria. We might as well say that there are no fixed points, and no fixed metapoints either. (And that includes the statement just made.)

Any particular purported principle of reason might be given up, but must not reason itself stand always and eternally as the final arbiter of all intellectual claims? Must not reason remain always (as Thomas Nagel called it) "the last word"?⁵ But what precisely is it that must endure? Not any particular statement or principle, for any one of those might be transformed or replaced or delimited or rejected. If it is said that Reason itself, rather than any particular statement of its content, must remain as the final arbiter,

then we must wonder what precisely *that* is. If not as particular content, then the only sense in which reason must endure is as an evolving chain of descent. Reason will endure as whatever evolves or grows out of the current content of reason by a process of piecemeal change that is justified at each moment by principles which are accepted at that moment (although not necessarily later on), provided that each evolving stage seems close enough to the one immediately preceding it to warrant the continued use of the label “reason” then. (The new stage may not seem very similar, however, to a much earlier, stepwise stage.) That degree of continuity hardly seems to mark something which is a fixed and eternal intellectual point. It is little help to add that what will endure is some process of thought that, in human beings, will continue to have its significant locus in the prefrontal cortex of the brain unless and until genetic alterations and artificial prosthetic supplementary devices come to shift the physical location of reason’s major activity.

My own philosophical bent is to open possibilities for consideration. Not to close them. This book suggests new philosophical views and theses, and the reasons it produces to support these are meant to launch them for exploration, not to demonstrate conclusively that they are correct. Similarly, my criticisms of some major competing theories or positions are not intended to refute them conclusively, merely to weaken them enough to clear a philosophical space in which the newly proposed views can breathe and grow. I state this aim here at the very beginning so that disclaimers do not have to be repeated constantly throughout. This is not a cautious attempt to avoid criticism by a diminution of ambition. The claim to be saying things that are new and philosophically interesting sticks one’s neck out considerably.

The opening and exploring of new views, without aiming at their *proof*, is especially suited for expanding philosophical knowledge. The method of proof starts at a place or with a set of premisses, and it goes on only to what can be proven on their basis. If there are some truths that cannot be reached from that starting place,⁶ this method will limit the truth content of the philosophy that results. Furthermore, it does not make rational sense to restrict one’s philosophical method to proof only, once it has been acknowledged that the view is not built upon certain and unalterable foundations anyway. Since the nonfoundationalist begins at a place that has not been proven or absolutely established, then since it is permissible to accept *some* things that have not been absolutely proven, why cannot one accept other such things as well?

Why might someone require that philosophical views be accepted only if

they are provable from the starting place P ? That requirement does not guarantee the truth of the resulting views, for P itself may not be true. It is true that the method of proof does guarantee that one will not add new independent falsehoods to one's views (although it does allow the addition of new false beliefs that are based, in part, upon falsehoods that P currently accepts). However, we also must notice that the restrictive nature of the method of proof may prevent one from *eliminating* some falsehoods currently residing within P , which the use of certain nondeductive methods might enable one to discard. For a nondeductive method might allow one, starting from P , to reach some new propositions that would be incompatible with those particular falsehoods within P , thereby prompting one to eliminate those falsehoods from one's current view. Through the nondeductive methods, P gets transformed by adding propositions to it and then by subtracting those (false) propositions that are incompatible with the new ones; thus a new view arises. Requiring that philosophers proceed only by the method of proof makes it impossible to eliminate falsehoods in this way. The method of proof is not as unalloyed a friend of truth as might have appeared.

An alternative method, the one I follow here, consists in a series of philosophical forays. Start at your current position P , and consider what is plausible, illuminating, intellectually interesting, and supported by reasons, given P . (These reasons should have weight but they need not be conclusive.) Finding something that is such, suppose that *it* is true and then consider what is plausible, illuminating, intellectually interesting, and supported by reasons, in its light in turn. Finding another thing that is such, consider what is plausible . . . And so on. This method does not mark out one unique path. Different things will be plausible, illuminating, intellectually interesting, and supported by reasons, given P , and these different things are not identical or perhaps even compatible with each other. Assuming one of these things, still further different things will be plausible, illuminating, intellectually interesting, and supported by reasons, given it, and these further things too are not identical or perhaps even compatible. So the proposed method will explore, or allow the exploration of, different pathways.

By following this method as outlined, we may eventually arrive at considering a view that is *not* plausible on the basis of the starting position P . (For the relation " X is plausible given Y " is not a transitive relation.) Should we limit the extent of our philosophical method only to what is plausible given our initial starting point? Some will indeed want to stop there. But I

suggest that plausibility (from the standpoint of the initial position) should not be a constraint on exploration, so long as we follow a chain of reasoning of the sort described, and continue to reach something that is illuminating and intellectually interesting given *P*. No doubt, in seeking philosophical interest and illumination, we will want especially to explore the most plausible of the intellectually interesting and illuminating views.⁷

We should, however, limit the range of our forays to what is philosophically interesting and illuminating given our current position *P*. (This constitutes a limit because the relation “*X* is philosophically interesting and illuminating given *Y*” is not transitive, and so a chain of such connections can reach something that is *not* philosophically interesting and illuminating in relation to the starting point.) Of course, if we come to *accept* some of these new things (and aren’t simply exploring them), then what is our current position will change from *P* to some new *P'*. Thereupon, what we foray into also will change. The domain of forays will both get extended and get restricted—extended, in that new things will be interesting and illuminating given *P'* that previously were not; restricted, in that some once interesting and illuminating views no longer will be so, given *P'*.

A comparison to physics is apropos. Physicists use nonrigorous messy mathematics for extended periods in order to make fast, fruitful, and promising progress on the problems that concern them. (Recall the state of the calculus before Weierstrass, and the path to renormalization methods in quantum field theory.) They are content to leave the task of cleaning up their methods to the mathematicians who follow in their wake. At present, string theorists explore many varieties of string theory, and just now they are beginning to peek at ways of unifying these in M-theory. They find interesting consequences (the theory yields gravity in a very natural way), they explore constraints such as supersymmetry for their fruitful consequences (without any very strong evidence that these constraints hold), they investigate the properties of a theory they have only sketched but not yet explicitly stated (in that *some* of its properties are known but the equations of the theory are not), and they use methods of approximation (perturbation theory) to get some grip on the consequences of the theory (and put little weight on the areas where they know these methods are greatly off the mark). All this theoretical effort, without coming close (at present) to touching down on particular empirical data or experiments. (Since string theory is currently the most exciting area of metaphysics, I return to it in Chapter 3.)

Although arriving at a range of new and philosophically interesting and

illuminating theories is intellectually quite exciting, the mark of significant philosophical progress occurs when some of the different positions that the forays have managed to reach begin, not just to reside alongside each other, but to combine, integrate, and *knit together* to yield a new, illuminating, and interesting structure.

Even this outcome, of course, does not prove that one is on the right track, but it does make it more probable. For one cannot automatically always get some such group of separately illuminating and interesting pieces to knit together into a resulting structure that also illuminates these pieces and perhaps some new things as well. Such a series of disciplined philosophical forays would constitute what Imre Lakatos has termed a “progressive research program,”⁸ if it continues to reach new illuminating positions that knit together into an interesting structure, especially if this structure fruitfully suggests, or provides a platform for, further forays that reach still further positions that satisfy these criteria.⁹

“But should we *believe* the results of this kind of philosophy?” This mode of philosophy is not designed to induce belief in something or the acceptance of it as true (beyond the belief that the theory is illuminating and interesting). Belief may eventually result if the view is plausible and best explains or unifies various things. Yet the philosophy, or that particular pathway to it, is not shown to be unsuccessful when it does not lead to belief (and this, not just because this method’s other pathways might themselves lead to belief). Belief is not the only coin of the philosophical realm. There are new classifications and analyses and understandings, and there is the process of intellectual openness itself. Ones that do not eventuate in belief may get fulfilled in other ways. Aquinas led to Dante.

A philosophy’s process may continue to hold intellectual interest because of its new methods of foray, its new ways of applying old methods, and its imagination and its daring. Future thinkers might be spurred to find analogues of these characters to fit their new intellectual situation. Twenty-five hundred years later, we still find reading Plato exciting—not for his *results*.

When I held earlier that there are no fixed philosophical points, I meant that there are no fixed philosophical concepts either. Yet certain concepts seem central, and locked into apparently immovable place by other concepts that they organize; they seem *guaranteed* to apply fruitfully. Evolutionary psychologists speak of an innate predisposition within us to understand and explain the world through certain categories and “theories,” for instance, that the world consists of continuously moving physical objects, and that beliefs and desires propel other people’s behavior and our own actions as well.

An evolutionary account of some of our fundamental concepts is both reassuring and upsetting. It is reassuring in that it holds that these concepts had enough of a grip upon the world that they aided reproductive success and were selected for. They must have got something right. Yet this selection does not guarantee their complete accuracy. The spatial concepts of Euclidean geometry were good enough to guide us practically, yet we now know that they are not, strictly speaking, accurately applicable to the world. So, when we see the world through an evolutionarily instilled conceptual structure, do we thereby know the way the world really is? Do we know the world—to use Kantian terminology—as it is *in itself*?

Since the time of Aristotle, philosophers have delineated and utilized fundamental categories of understanding: form, content, substance, property, causality, object, belief, desire, space, time, objectivity, truth. Some philosophers hold these categories to be inescapable, and necessarily accurate of the world. Kant retreated to saying that they necessarily apply to the world as we structure and represent it—they apply to appearances rather than to things in themselves.¹⁰ More recently, some philosophers have reinstated the authority and validity of some of these traditional concepts on the basis of the *intuitions* that structure their philosophy; they treat these intuitions as data to which a philosophical theory must conform. However, an evolutionary explanation of these concepts undercuts their unquestionable authority, along with that of any intuitions they bring in their wake. No secure basis for a philosophical theory there!

One piece of the objective world, a piece that is instilled in our heads by evolution, is our theory of that world, and this theory is not necessarily strictly accurate. If someone questions a part of that theory, even calling into question the applicability of its basic categories, it is not sufficient, or appropriate, for philosophers to quickly argue the contrary position in a way that presupposes the valid application of these very concepts or of closely associated ones. Or to react in outrage, even when what is questioned are such central notions as objective truth, objective inquiry, and rationality.

Not even the applicability of the notions of belief and desire can be taken for granted. Two philosophers, Paul and Patricia Churchland, have recently claimed that these categories of “folk psychology” are seriously inadequate and will be replaced by the explanatory concepts of a developed neuroscience. Many philosophers have reacted to this view as earlier generations did to atheists, recoiling from their denial of the existence of sacred things. The denial of the folk-psychological categories might be incorrect, but it is not beyond the pale of possibility.

A philosopher should be open to radically different intriguing conceptual possibilities. At the least, their investigation will lead to deeper insight into the conceptual structure we find ourselves inhabiting or being inhabited by. Too often, philosophers insist that things *must* be a certain way, and they make it their business to close off possibilities. “Damn braces. Bless relaxes.”¹¹

It is a commonplace that advances in physics have radically overturned our evolutionarily instilled theories and concepts, and our common-sense theories. Space is not Euclidean, simultaneity is not absolute, the world is not deterministic, quantum events stand in (an ill-understood) relation to their being observed, and they also can have nonlocal correlates that are not mediated by intervening processes. Such an overturning of our traditional concepts can be disquieting and disconcerting. People go to great lengths to avoid this.

An instance is David Bohm’s formulation of quantum mechanics, which has won some adherents among philosophers but few among physicists.¹² Bohm’s theory is deterministic, particles have definite positions and trajectories, the wave function of the system refers to an existent entity that determines how the particles behave, there is no collapse of the wave function, and (as John Bell has emphasized) the theory does not require reference to an observer and it offers a homogeneous account of the physical world that is not bifurcated into a quantum part and a classical part. However, according to Bohm’s view “the fundamental laws of the world are cooked up in such a way as to systematically *mislead* us about themselves . . . [the theory] recounts the unfolding of a perverse and gigantic conspiracy to make the world *appear* to be *quantum-mechanical*.”¹³

Bohmian mechanics is a far more complicated theory than orthodox quantum mechanics. It is interesting that this deterministic theory of objective existents can be formulated. But what do we learn from that fact? At the beginning of the twentieth century, Henri Poincaré said that since the simplest geometry for space was Euclidean, and since all other facts could be assimilated into the Euclidean framework by complicating the (theories of) physics, Euclidean geometry was unassailable. It could not rationally be given up. Simplicity was important, others agreed, but the important simplicity was that of the overall theory, not just of one theoretical component, even if that one component is as large and extensive as geometry. If the greatest overall simplicity of geometry + physics can best be achieved by including a *non*-Euclidean geometry, then that more complex geometrical component will be included in the interests of the simplicity of our overall