

# STRAWSON'S NO-SPACE WORLD

What does the suggestion that we explore [a] No-Space world amount to? What is it to imagine ourselves dispensing with outer sense? . . . The only objects of sense-experience would be sounds. Sounds of course have temporal relations to each other, and may vary in character in certain ways: in loudness, pitch and timbre. But they have no intrinsic spatial characteristics. . . . I shall take it as not needing further argument that in supposing experience to be purely auditory, we are supposing a No-Space world. . . .

The question we are to consider, then, is this: could a being whose experience was purely auditory have a conceptual scheme which provided for objective particulars?

Source: P. F. Strawson. *Individuals*. Garden City, NY: Anchor Books, 1959. 56–58.

Strawson states that in *Individuals*, he is engaging in “descriptive metaphysics”—“describing the actual structure of our thought about the world” (xiii). “We think of the world as containing particular things some of which are independent of ourselves” (2), Strawson says, and we seem to accord special importance to their spatiotemporal position. Why is that so? He replies that “the system of spatiotemporal relations has a peculiar comprehensiveness and pervasiveness, which qualify it uniquely to serve as the framework within which we can organize our individuating thought about particulars” (13). Every particular thing seems to have its place in this system, which makes it easy to identify them, to refer to them, to communicate with each other about them. However, Strawson suggests, other conceptual schemes could exist; materiality (spatiotemporal position) is not a necessary condition of objective particulars. And his No-Space world is intended to demonstrate such an alternative scheme.

One might raise the objection that sound is, in fact, spatial—doesn't it come from the right or left, from near or far? Yes, Strawson would reply, but this seems so only because we have our other space-based senses (such as touch); if we had *only* auditory sense, sounds wouldn't seem spatially located.

Would sounds be identifiable particulars in his No-Space world, as he claims (for he answers his question with a “yes”)? That is, if there were nothing else but sound, how could the beings in that world distinguish between them and not-them? Strawson replies that audible continuity or discontinuity could be used as a criterion.

Another interesting question is whether, given that there is no spatial sense at all (for the beings themselves or for the things in their world), such beings would be able to distinguish between themselves and the sounds—or for that matter, between themselves and other selves? If so, how so?

## QUINTON'S TWO-SPACE MYTH

[S]uppose that your dream-life underwent a remarkable change. Suppose that on going to bed at home and falling asleep, you found yourself to all appearances waking up in a hut raised on poles at the edge of a lake. A dusky woman, whom you realize to be your wife, tells you to go out and catch some fish. The dream continues with the apparent length of an ordinary human day, replete with an appropriate and causally coherent variety of tropical incident. At last you climb up the rope ladder to your hut and fall asleep. At once you find yourself awaking at home, to the world of normal responsibilities and expectations. The next night life by the side of the tropical lake continues in a coherent and natural way from the point at which it left off. Your wife says "You were very restless last night. What were you dreaming about?" and you find yourself giving her a condensed version of your English day. And so it goes on. Injuries given in England leave scars in England, insults given at the lakeside complicate lakeside personal relations. One day in England, after a heavy lunch, you fall asleep in your armchair and dream of yourself, or find yourself, waking up in the middle of the night beside the lake. Things get too much for you at the lakeside, your wife has departed with all the cooking-pots, and you suspect that she is urging the villagers to sacrifice you to the moon. So you fall on your fish-spear and from that moment on your English slumbers are disturbed no more than in the old pre-lakeside days.

Is such a two-space reality conceivable? That is, is it conceivable that we could live in two different but real spaces?

Source: Anthony Quinton. "Spaces and Times." *Philosophy* 37 (1962): 130-147. 141.

Quinton is investigating with this thought experiment the commonly held notion that space and time are unitary—that is, that we take (and according to Kant, are *compelled* to take) "real spatial extents and temporal durations to be part of the one space and the one time" (139). Wondering whether there are any conceivable circumstances in which it would be reasonable to revise that notion, Quinton presents his Two-Space Myth and shows that it *is* conceivable that we could live in two different but real spaces. Quinton argues that the lake life is just as coherent as the life in England, and it could be just as public (the lake villagers confirm your experiences there, a confirmation as reliable as that provided by your neighbors in England about your experiences there) or it could remain private ("in this case everyone would inhabit two real spaces, one common to all and one peculiar to each" [143]).

But, Quinton anticipates the objection, the lake place is *not real*—no one, in fact, can locate it. So? "Why," asks Quinton, "do we have this ontological wastepaper basket for the imaginary?" (144). Is it because, he wonders, there are no consequences in our imaginary world and we thus don't have to take it seriously? But there *are* consequences in the lake world and you *do* take it seriously (you fall on your fish spear, remember?). Interpreted this way, Quinton says, reality doesn't need to be located in a (single) physical space.

Can the same be said about time? Can we conceive two coherent experiences such that the people *within* each experience are temporally related but there is no temporal relation *between* the two experiences? This doesn't seem possible, Quinton says: "If an experience is mine it is memorable, and if it is memorable it is temporally connected to my present state" (146); in other words, unless you remember the experiences of the one world while in the other, there's no reason to say you are in both worlds, but if you do remember the experiences of the one while in the other, then the worlds aren't in two separate times. So, Quinton concludes, while our concepts of experience need not be spatial (see "Strawson's No-Space World" for agreement on this point), it does need to be temporal (see "Shoemaker's Time-Freezing World" for disagreement on this point).

## LOCKE'S VOLUNTARY PRISONER

[S]uppose a man be carried, whilst fast asleep, into a room where is a person he longs to see and speak with; and be there locked fast in, beyond his power to get out: he awakes, and is glad to find himself in so desirable company, which he stays willingly in, i.e. prefers his stay to going away. I ask, is not this stay voluntary?

Source: John Locke. *An Essay Concerning Human Understanding*. Book 2, Chapter 21, Section 10. 1690. As collated and annotated by Alexander Campbell Fraser. New York: Dover, 1959. Volume 1. 317.

It is perhaps easy to think that voluntary action is evidence of freedom. This thought experiment shows otherwise. Locke claims that the man's staying in the room *is* voluntary (he stays willingly), but it is *not free*: "So far as a man has power to think or not to think, to move or not to move, according to the preference or direction of his own mind, so far is a man *free*" (315). Choosing to do something makes the action voluntary, but unless you could actually do otherwise (the man can't leave), the action is not free. (But since we can do only one thing, we can never really know whether we really *could* have done otherwise. Right?)

Locke claims, therefore, that questions about "free will" don't make sense—"freedom" and "will" are two different things: will is the capacity to think of various actions and choose whichever is preferable, whereas freedom is the capacity to actually do as one wills. So the question isn't whether *the will* is free, says Locke, but whether *a person* is free.

Which should bear on moral responsibility—the voluntariness or the freedom? Is it that a person is morally responsible for doing X as long as she does X because she *chooses* to do X—*whether or not* she could have done otherwise? Or is it that a person is morally responsible for doing X only if he could have done otherwise (in which case determinism is incompatible with moral responsibility—in a determined world, we can't do other than what we do, so we can't be held morally responsible for our actions)?

# GOLDMAN'S BOOK OF LIFE

While browsing through the library one day, I noticed an old dusty tome, quite large, entitled "Alvin I. Goldman." I take it from the shelf and start reading. In great detail, it describes my life as a little boy. It always gibes with my memory and sometimes even revives my memory of forgotten events. I realize that this purports to be a book of my life, and I resolve to test it. Turning to the section with today's date on it, I find the following entry for 2:36 P.M. "He discovers me on the shelf. He takes me down and starts reading me. . . ." I look at the clock and see that it is 3:03. It is quite plausible, I say to myself, that I found the book about half an hour ago. I turn now to the entry for 3:03. It reads: "He is reading me. He is reading me. He is reading me." I continue looking at the book in this place, meanwhile thinking how remarkable the book is. The entry reads: "He continues to look at me, meanwhile thinking how remarkable I am."

I decide to defeat the book by looking at a future entry. I turn to an entry 18 minutes hence. It says: "He is reading this sentence." Aha, I say to myself, all I need do is refrain from reading that sentence 18 minutes from now. I check the clock. To ensure that I won't read that sentence, I close the book. My mind wanders; the book has revived a buried memory and I reminisce about it. I decide to reread the book there and relive the experience. That's safe, I tell myself, because it is an earlier part of the book. I read that passage and become lost in reverie and rekindled emotion. Time passes. Suddenly I start. Oh yes, I intended to refute the book. But what was the time of the listed action?, I ask myself. It was 3:19, wasn't it? But it's 3:21 now, which means I have already refuted the book. Let me check and make sure. I inspect the book at the entry for 3:17. Hmm, that seems to be the wrong place for there it says I'm in a reverie. I

Source: Alvin I. Goldman. "Actions, Predictions, and Books of Life." *American Philosophical Quarterly* 5.3 (1968): 135-151. 143-144.

skip a couple of pages and suddenly my eyes alight on the sentence: "He is reading this sentence." But it's an entry for 3:21, I notice! So I made a mistake. The action I had intended to refute was to occur at 3:21, not 3:19. I look at the clock, and it is still 3:21. I have not refuted the book after all.

---

Would Goldman ever be able to falsify the predictions made in his "book of life"? If not, does that prove the world, and our lives, are determined?

---

Goldman continues his thought experiment a little further, describing two more predicted events that he considers falsifying, but he finds that he has good reasons (currently existing reasons in the one case, new and unanticipated reasons in the other case) to do as predicted, and so he does. It would seem, then, that his answer to the first question is "no"—in the world he has constructed, which is a determined world, he will not be able to falsify the predictions. As for the second question, Goldman's intent is not to show that our lives are determined, but rather that determinism is *compatible* with our lives as we experience them—that is, as having voluntary behavior such as deliberation, choice, and decision.

But so what? Do we want to know our choices are *compatible* with the real world, or do we want to know they have *causal* force (rather than being merely ineffectual illusions)? Goldman seems to suggest the latter is the case (as well as the former), claiming that although our action is determined, or causally necessitated, "one of the antecedent conditions which necessitate it is [our] deliberation" (150). Still, doesn't something seem "wrong" about "deliberating" over a decision that's inevitable?

# TAYLOR'S INGENIOUS PHYSIOLOGIST

[W]e can suppose that an ingenious physiologist can induce in me any volition he pleases, simply by pushing various buttons on an instrument to which, let us suppose, I am attached by numerous wires. All the volitions I have in that situation are, accordingly, precisely the ones he gives me. By pushing one button, he evokes in me the volition to raise my hand; and my hand, being unimpeded, rises in response to that volition. By pushing another, he induces the volition in me to kick, and my foot, being unimpeded, kicks in response to that volition. We can even suppose that the physiologist puts a rifle in my hands, aims it at some passer-by, and then, by pushing the proper button, evokes in me the volition to squeeze my finger against the trigger, whereupon the passer-by falls dead of a bullet wound.

---

Am I free?

Source: Richard Taylor. *Metaphysics*. 2nd edition. Englewood Cliffs, NJ: Prentice-Hall, 1974. 50.

If everything in the universe is determined, it would seem that we can't have free will. (Though it may be important to define determinism—for example, to say that everything is determined, or caused, by preexisting conditions is not necessarily to say that at any given time only one action is possible.) “Compatibilism” (a view claiming that free will is compatible with determinism) provides a “solution” to this “problem” by defining freedom as the absence of obstacles that prevent one from doing something and/or forces that compel one to do something. Thus, one can be free (free of obstacles and forces) even in a causally determined world—free to act according to one's volitions or desires.

Taylor's thought experiment is intended to challenge compatibilism: one may be free to act according to one's desires (that is, one is neither prevented nor compelled), but as long as one's desires are caused by something (as indeed they must be, according to determinism), then one is hardly free. But, one might respond, that “something” may be our own selves (and not some ingenious physiologist): *we* are responsible for our desires because of past choices (that make us who and what we are) and/or because of our reasoning about our options. What if you had *asked* the physiologist to “cause” those desires? Consider a person who hires a hypnotist to implant the desire to go outside. Is he not acting according to free will? (When he hires the hypnotist *and* when he later goes outside?) And yet, what makes us choose as we do in the past? What makes us reason as we do? Can't who and what we are (including the capacity to change who and what we are) be attributed solely to our genetic makeup and the events that happened to us—both of which are external causes, as “compelling” as the ingenious physiologist? Consider Gardner's turtles: “[Imagine] a mechanical turtle that crawls across the floor in obedience to internal mechanisms. It moves here and there, seemingly at random. Contrast this with a toy turtle that a child pulls with a string. The toy is compelled by outside forces to move as it does, whereas the mechanical turtle is under no extraneous compulsion” (Martin Gardner, *The Whys of a Philosophical Scrivener*, 105). Is either one free?

Note that if the world is *not* determined, we also can't have free will, for in a world in which events are not caused, our will would have no effect at all on our actions.

But does it have to be “all or nothing”? Can't we say that some events are caused and some not? But then, which are which? Perhaps events are caused by a constellation of preexisting conditions, and perhaps in the case of human behavior, our will is one of them. So is it that our will *influences* but does not completely cause our behavior?

## PALEY'S WATCH

In crossing a heath, suppose I pitched my foot against a *stone*, and were asked how the stone came to be there; I might possibly answer that for anything I knew to the contrary, it had lain there forever; nor would it perhaps be very easy to show the absurdity of this answer. But suppose I had found a *watch* upon the ground, and it should be inquired how the watch happened to be in that place.

Source: William Paley. *Natural Theology, or Evidences of the Existence and Attributes of the Deity Collected from the Appearances of Nature*. 1802. As reprinted in *A Modern Introduction to Philosophy: Readings from Classical and Contemporary Sources*. 3rd edition. Paul Edwards and Arthur Pap, eds. New York: The Free Press, 1973. 419–434. 419.

Paley's response is that "the watch must have had a maker" (420) because "its several parts are framed and put together for a purpose [that being to tell time]" (419). He then reasons that since the natural world shows not only as much but more design toward a purpose, it too must have had a maker. This "argument from design" for the existence of a creator god is actually, then, an argument by analogy: the watch is to the watchmaker as the natural world is to the creator god.

But is the analogy sound? First, *is* the natural world as "framed and put together" as a watch? One can point to several instances that suggest not. Paley would respond that he needs only one instance of design—and he focuses on the human eye—in order to conclude that there is indeed a designer. One might then point out that the human eye isn't very well designed; for example, it's useless unless there's light. But, Paley would respond, imperfections in design are relevant to the *attributes* of a creator (such imperfections might suggest, for example, an unimaginative or inept designer); he is establishing only the *existence* of a creator.

Second, *do* the parts of the natural world work together "for some purpose"? One might respond that the purpose of much of the natural world, ourselves included, is not as evident as the purpose of the watch. Paley might respond that it doesn't matter whether we understand how the parts work together—it matters only that they are designed to do so. But if we don't know what the purpose of the natural world is, how can we say it *is* designed for some purpose?

Even if the parts of the natural world do fit together, achieving some purpose, is a creator god the only explanation possible? Perhaps the world is that way by chance. Paley would say the watch—and by analogy, the natural world—is too complicated, too organized, to have been the result of chance—a pimple might be the result of chance, but not an eye! (*Is* a stone so different from a watch?)

Perhaps the world was always that way. Paley would say that appealing to some infinite regress still leaves design unaccounted for.

Or perhaps, as evolutionary theory suggests, the parts fit together because those that didn't fit together (didn't adapt to their environment) didn't survive. The evolutionary theory does seem to challenge Paley's argument, but it need not challenge his conclusion: advocates of theistic evolution would argue that a god designed the developmental processes that led to the world we have (rather than, as Paley claims, designing the world as is).

# HICK'S RESURRECTED PEOPLE

**F**irst picture: Suppose that at some learned gathering in this country, one of the company were suddenly and inexplicably to disappear, and that at the same moment, an exact replica of him were suddenly and inexplicably to appear at some comparable meeting in Australia. The person who appears in Australia is exactly similar, as to both bodily and mental characteristics, with the person who disappears in America. There is continuity of memory, complete similarity of bodily features, including even fingerprints, hair and eye coloration, and stomach contents, and also of beliefs, habits, and mental propensities. In fact there is everything that would lead us to identify the one who appeared with the one who disappeared, except continuity of occupancy of space. . . .

Second picture: Now let us suppose that the event in America is not a sudden and inexplicable disappearance, and indeed not a disappearance at all, but a sudden death. Only, at the moment when the individual dies, a replica of him as he was at the moment before his death, complete with memory up to that instant, appears in Australia. . . .

Third picture: My third supposal is that the replica, complete with memory, etc. appears, not in Australia, but as a resurrection replica in a different world altogether, a resurrection world inhabited by resurrected persons. This world occupies its own space, distinct from the space with which we are now familiar. . . .

---

Can we not imagine this?

Source: John Hick. "Theology and Verification." *Theology Today* 17.1 (1960): 12-31. 22, 23.

**T**his thought experiment is presented in the context of a discussion about whether or not the existence of the Christian god is, in principle, verifiable. That is to say, can we at least *imagine* some experience that would prove that such a god exists? "Life after death"—that is, "continued conscious existence after bodily death" (16)—is such an experience, claims Hick. However, others claim that such a concept of immortality is unintelligible: the self cannot exist without the physical body. Hick's thought experiment is designed to show that the idea of life after death *is* intelligible, that we *can* imagine, without contradiction, continued conscious existence after bodily death. (And so the existence of the Christian god is thus, in principle, verifiable).

But, one might ask, considering the third picture, how will the person know he has really died? Maybe he just fell asleep and then woke up—so it's not life after death after all. Hick adds to his picture the possibility that the person will meet in the resurrection world people he knows to have died.

Even with that addition, does Hick's experiment demonstrate what he thinks it demonstrates? Perhaps immortality *is* intelligible, and perhaps such immortality is in accord with the concept of the Christian god. But is it in accord *only* with a Christian god? Perhaps life after death verifies some other god or just that death as we know it isn't the end many of us think it is. To this, Hick merely adds another possibility to his picture, the possibility that the person will in some way meet with the Christian god in the resurrection world.

But how will the person *know* that he has met with that god—how can a mere human *recognize* a transcendent being with qualities that so exceed human experience? Hick's reply is that the Christian god reveals himself to us through Jesus Christ, so if the person were to meet Jesus Christ in the resurrection world, that would suffice to verify the Christian god's existence (though, as Hick concedes, that existence would be verified only for that person). Is that too big an "if"?

And is verification *in principle* of any significant value?

## LOCKE'S INVERTED SPECTRUM

[Suppose] by the different structure of our organs it were so ordered, that *the same object should produce in several men's minds different ideas* at the same time; for example, if the idea that a violet produced in one man's mind by his eyes were the same that a marigold produced in another man's and vice versa.

---

Could this be known to be so?

Source: John Locke. *An Essay Concerning Human Understanding*. Book 2, Chapter 32, Section 15. 1690. As collated and annotated by Alexander Campbell Fraser. New York: Dover, 1959. Volume 1. 520.

Locke's response is that we could not know if this were so "because one man's mind could not pass into another man's body, to perceive what appearances were produced by those organs" (520). The possibility that another person's experience of the same thing may be different from your own (for example, that his or her color spectrum may be inverted relative to yours), and the impossibility of knowing this, underscores the absolute subjectivity of experience: we can experience only "the view from here" (see "Nagel's Bat"). One of the implications of this is the limitation it puts on establishing truth: there is simply no way to prove or disprove subjective experiences.

Another implication is that if we can't know the contents of other minds, can we even know there *are* other minds? (This is called the "other minds" problem. See "Kirk and Squires's Zombies.")

Contemporary developments of this thought experiment postulate not an *intersubjective* difference, but an *intrasubjective* one: suppose that a person's own spectrum has changed—as a result of inverting glasses, neurological rewiring, or transport to a planet with yellow skies and red grass. Such suppositions are intended to challenge the view that mental states are equivalent to functional states or behaviors. That such a person may long for the way the colors used to be shows that her mental state has changed (for example, she now experiences something different when she sees red) even though her functional state has not (for example, after a period of adaptation, she can still stop when she sees a "red" light)—thus, one can conclude, there are two independent states involved. Or not—if her actual subjective experience reverted (and it wasn't just that she had adapted to the inverted subjective experience), then mental states may be equivalent to functional states after all.



## NAGEL'S BAT

I assume we all believe that bats have experience. After all, they are mammals, and there is no more doubt that they have experience than that mice or pigeons or whales have experience. . . .

. . . Now we know that most bats (the microchiroptera, to be precise) perceive the external world primarily by sonar, or echolocation, detecting the reflections, from objects within range, of their own rapid, subtly modulated, high-frequency shrieks. Their brains are designed to correlate the outgoing impulses with the subsequent echoes, and the information thus acquired enables bats to make precise discriminations of distance, size, shape, motion, and texture comparable to those we make by vision. . . .

. . . [But] I want to know what it is like for a *bat* to be a bat.

Source: Thomas Nagel. "What Is It like to Be a Bat?" *Philosophical Review* 83.4 (1974): 435-450. 438, 439.

Can we say? Can we know what it's like to be a bat? Nagel's response is that we can't say—we can't know what it's like *for a bat to be a bat*: "[B]at sonar, though clearly a form of perception, is not similar in its operation to any sense that we possess, and there is no reason to suppose that it is subjectively like anything we can experience or imagine. This appears to create difficulties for the notion of what it is like to be a bat" (438). We might be able to say what it would be like for *us* to be a bat, but not what it is like *for the bat* to be a bat. "Even if I could by gradual degrees be transformed into a bat," says Nagel, "nothing in my present constitution enables me to imagine what the experiences of such a future stage of myself thus metamorphosed would be like" (439). In fact, Nagel suggests, while this would most certainly be true as well of any extra-terrestrial life form we may meet, it is also true of other human beings—we may not even be able to say what it's like for another person to be that person (unless he or she is sufficiently similar to ourselves).

Nagel is investigating here the relation between mind and body, which is, he says, particularly difficult because of consciousness: "[T]he fact that an organism has conscious experience *at all* means, basically, that there is something it is like to *be* that organism" (436)—he calls that something the subjective character of experience. "What it's like" is accessible only from one point of view, the viewpoint of the subject (see "Locke's Inverted Spectrum"). Therefore, since the subjective experience can't be accessed by anyone outside the subject, inferences from observable physical behavior (body) to mental states (mind) seem questionable. One might wonder, then, since we can't know the nature of others' experiences, can we at least know they have them? Can we at least know there's a "they"—that there *are* other minds?

Is Nagel correct? Can we never imagine something that is totally *outside*, totally *beyond*, our own experience? (Can we describe the taste of chocolate to someone who has never tasted it?)

# BLOCK'S CHINESE NATION

Imagine a body externally like a human body, say yours, but internally quite different. The neurons from sensory organs are connected to a bank of lights in a hollow cavity in the head. A set of buttons connects to the motor-output neurons. Inside the cavity resides a group of little men. Each has a very simple task: to implement a "square" of a reasonably adequate machine table that describes you. On one wall is a bulletin board on which is posted a state card, i.e., a card that bears a symbol designating one of the states specified in the machine table. Here is what the little men do: Suppose the posted card has a "G" on it. This alerts the little men who implement G squares—"G-men" they call themselves. Suppose the light representing input  $I_{17}$  goes on. One of the G-men has the following as his sole task: when the card reads "G" and the  $I_{17}$  light goes on, he presses output button  $O_{191}$  and changes the state card to "M." This G-man is called upon to exercise his task only rarely. In spite of the low level of intelligence required of each little man, the system as a whole manages to simulate you because the functional organization they have been trained to realize is yours. . . .

Suppose we convert the government of China to functionalism, and we convince its officials that it would enormously enhance their international prestige to realize a human mind for an hour. We provide each of the billion people in China (I chose China because it has a billion inhabitants) with a specially designed two-way radio that connects them in the appropriate way to other persons and to the artificial body mentioned in the previous example. We replace the little men with a radio transmitter and receiver connected to the input and output neurons. Instead of a bul-

Source: Ned Block. "Troubles with Functionalism." *Minnesota Studies in the Philosophy of Science* 9 (1978): 261-325. 278-280.

letin board, we arrange to have letters displayed on a series of satellites placed so that they can be seen from anywhere in China. Surely such a system is not physically impossible. It could be functionally equivalent to you for a short time, say an hour.

Does this system have mental states?

Those who claim that mental states can be explained by the nonmental states of function or organization would say the Chinese nation *does* have mental states; since it is functionally equivalent to you, it would have the mental states you have. Block disagrees, however: the Chinese nation would *not* have mental states, because neurologically and psychologically it is unlike you (and you have mental states). So do mental states (minds) depend on neurology (brains)? What exactly is meant here by "depend"?

One may agree with Block intuitively, because it is hard to imagine the Chinese nation so described as being conscious, but surely it is equally hard to imagine the brain being conscious.

Perhaps the Chinese nation *does* have mental states—they're just not like the mental states we have. But how could we know this?

And would whatever conclusion that is reached be true for both non-qualitative mental states (for example, thoughts, desires, intentions) as well as qualitative mental states (for example, the way pain feels and the way red seems)? That is, do the different kinds of mental states have different relationships to either physical/material states or functional/organizational states?

## RORTY'S ANTIPODEANS

Far away, on the other side of our galaxy, there was a planet on which lived beings like ourselves—featherless bipeds who built houses and bombs, and wrote poems and computer programs. These beings did not know that they had minds. They had notions like “wanting to” and “intending to” and “believing that” and “feeling terrible” and “feeling marvelous.” But they had no notion that these signified *mental* states—states of a peculiar and distinct sort—quite different from “sitting down,” “having a cold,” and “being sexually aroused.” . . . [T]hey did not explain the difference between persons and non-persons by such notions as “mind,” “consciousness,” “spirit,” or anything of the sort. . . .

In most respects, then, the language, life, technology, and philosophy of this race were much like ours. But there was one important difference. Neurology and biochemistry had been the first disciplines in which technological breakthroughs had been achieved, and a large part of the conversation of these people concerned the state of their nerves. When their infants veered toward hot stoves, mothers cried out, “He’ll stimulate his C-fibers!” . . . Their knowledge of physiology was such that each well-formed sentence in the language which anybody bothered to form could easily be correlated with a readily identifiable neural state.

---

What can we say about the Antipodeans with respect to mental phenomena?

Source: Richard Rorty. *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press, 1979. 70–71.

Rorty hopes to show with his Antipodeans that we can do quite well without conceptualizing a “mind”; our notion of “mental” and “physical” is a mistaken inheritance from the seventeenth century. Reference to neural states could sufficiently replace all our talk about mental states. This is not to deny the existence of things like pain, for what the Antipodeans report when they say “Stimulated C-fibers!” is what we report when we say “Pain!” Nor is this to deny that “pain” “hurts”—the Antipodeans avoid stimulated C-fibers just as we avoid pain.

One might question, however, whether reference to neural states could cover not only mental states such as pain but also mental states such as thoughts, beliefs, intentions, desires, and so on. Rorty seems to suggest it could when he says the following of his Antipodeans (71–72):

Sometimes they would say things like “It looked like an elephant, but then it struck me that elephants don’t occur on this continent, so I realized that it must be a mastodon.” But they would also sometimes say, in just the same circumstances, things like “I had G-412 together with F-11, but then I had S-147, so I realized that it must be a mastodon.”

*Could* knowledge of physiology be such that every sentence we wanted to form could be correlated with a neural state?

## PUTNAM'S BRAIN IN A VAT

[1] Imagine that a human being (you can imagine this to be yourself) has been subjected to an operation by an evil scientist. The person's brain (your brain) has been removed from the body and placed in a vat of nutrients which keeps the brain alive. The nerve endings have been connected to a super-scientific computer which causes the person whose brain it is to have the illusion that everything is perfectly normal. There seem to be people, objects, the sky, etc., but really all the person (you) is experiencing is the result of electronic impulses travelling from the computer to the nerve endings. The computer is so clever that if the person tries to raise his hand, the feedback from the computer will cause him to "see" and "feel" the hand being raised. Moreover, by varying the program, the evil scientist can cause the victim to "experience" (or hallucinate) any situation or environment the evil scientist wishes. He can also obliterate the memory of the brain operation, so that the victim will seem to himself to have always been in this environment. It can even seem to the victim that he is sitting and reading these very words about the amusing but quite absurd supposition that there is an evil scientist who removes people's brains from their bodies and places them in a vat of nutrients which keep the brains alive.

---

Could we be brains in a vat?

Source: Hilary Putnam. *Reason, Truth, and History*. Cambridge, UK: Cambridge University Press, 1981. 5–6.

Though the possibility of being a brain in a vat is often used to make an epistemological point (see "Descartes's Evil Demon"), Putnam uses it to explore the relationship between the mind and the external world. Though the brain in the vat may be conscious and intelligent, the words it thinks do not—cannot—refer to what our words refer to; the words of the brain in the vat necessarily refer to the images generated by the vat machinery, not to the actual external objects we call, for example, trees. But the images generated by the vat machinery are similar, indeed identical, to, say, trees—so how can we say the brain isn't therefore referring to real trees? Consider, responds Putnam, an ant whose path turns out to have traced a perfect image of Winston Churchill: surely we won't say the ant has drawn a picture of Winston Churchill—similarity is not enough for us to say that something represents something else.

"So," continues Putnam, "if we really are brains in a vat, then the sentence 'We are brains in a vat' says something false (if it says anything)" (15)—for the words would not be referring to real things. The supposition that we are brains in a vat is, thus, self-refuting (one whose truth implies its own falsity—see "The Liar Paradox"): "In short, if we are brains in a vat, then 'We are brains in a vat' is false" (15). (So we can't possibly be brains in a vat.)

Putnam's larger point is about the preconditions of reference and hence thought: "[O]ne cannot refer to certain kinds of things, e.g., *trees*, if one has no causal interaction at all with them, or with things in terms of which they can be described" (16–17). But can't we refer to unicorns? Yes, but we have causal interaction with horses and two-horned goats. Okay, doesn't what happens to the brain in the vat—the back-and-forth experience of neural signals—count as "causal interaction"? (So maybe we *are* brains in a vat. . . .)

## REID'S BRAVE OFFICER

Suppose a brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard [a flag] from the enemy in his first campaign, and to have been made a general in advanced life: suppose also, which must be admitted to be possible, that, when he took the standard, he was conscious of his having been flogged at school, and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging.

---

Is the general the same person as the boy?

Source: Thomas Reid. *Essays on the Intellectual Powers of Man*. 1785. As edited by A. D. Woozley. London: Macmillan, 1941. 213.

This thought experiment is intended to illustrate a weakness in Locke's theory that personal identity depends on our consciousness or memory of our thoughts and actions and can be extended backwards only as far as that consciousness or memory goes. If that were so, Reid says, then the officer is the same person as the boy, and the general is the same person as the officer, but the general is *not* the same person as the boy. And yet logic indicates that the general *is* the same person as the boy (if  $A = B$  and  $B = C$ , then  $A = C$ ). Reid therefore rejects Locke's view (and accepts the logic). But is the logic applicable in this case? Does "=" mean the same as "is the same person as"?

Reid suggests, instead, that the succession from A to B to C is sufficient for identity: "My thoughts, and actions, and feelings change every moment—they have no continued, but a successive existence; but that *self* of I to which they belong is permanent, and has the same relation to all the succeeding thoughts, actions, and feelings, which I call mine" (203). (See "Parfit's Teletransporter.") Perhaps, then, part of the problem is just our sloppy way of talking—do we really mean that the officer *is* the same person as the boy or that the officer is the person the boy has *become*?

Reid points out further problems with Locke's view: it confounds consciousness with memory (are they the same?), and it confounds personal identity with evidence of personal identity (can't you have the one without the other?). Memory is *evidence that* I am who I was, says Reid; it is not what *makes* me who I was (remembering that you did something doesn't make you to have done it).

Pointing to our ever changing consciousness, Reid also asks, in further critique of Locke's view, "Is it not strange that the sameness or identity of a person should consist in a thing which is continually changing and is not any two minutes the same?" (214).

Lastly, Reid says, if our personal identity consists in consciousness, then "as our consciousness sometimes ceases to exist, as in sound sleep, our personal identity must cease with it. Mr. Locke allows that the same thing cannot have two beginnings of existence; so . . . our identity would be irrecoverably gone every time we cease to think, if it was but for a moment" (216). And that is, Reid implies, absurd.

So if it's neither consciousness nor memory, what is it that makes you the same person tomorrow—or ten years from tomorrow—as you are today? (Or are you someone else every time you wake up?) (And do you have a problem with that?)

## SHOEMAKER'S BROWNSON

So let us imagine the following. . . . One day, to begin our story, a surgeon discovers that an assistant has made a horrible mistake. Two men, a Mr. Brown and a Mr. Robinson, had been operated on for brain tumors, and brain extractions had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One of these men immediately dies, but the other, the one with Robinson's body and Brown's brain, eventually regains consciousness. Let us call the latter "Brownson." Upon regaining consciousness, Brownson exhibits great shock and surprise at the appearance of his body. Then, upon seeing Brown's body, he exclaims incredulously, "That's me lying there!" Pointing to himself, he says, "This isn't my body; the one over there is!" When asked his name he automatically replies, "Brown." He recognizes Brown's wife and family (whom Robinson had never met), and is able to describe in detail events in Brown's life, always describing them as events in his own life. Of Robinson's past life, he evidences no knowledge at all. Over a period of time, he is observed to display all of the personality traits, mannerisms, interests, likes and dislikes, and so on that had previously characterized Brown, and to act and talk in ways completely alien to the old Robinson.

What would we say if such a thing happened?

Source: Sydney Shoemaker. *Self-Knowledge and Self-Identity*. Ithaca, NY: Cornell University Press, 1963. 23–24.

Since, Shoemaker claims, we would say that Brownson is actually Brown in Robinson's body, we are not using the body as our criterion for identity. (Note the use of "in"—indeed note that we say "I *have* a body" rather than "I *am* a body"—which suggests that one's self is somehow separate from one's body.) (Or maybe we're just speaking wrongly. Perhaps the whole notion of "I" is an unfortunate and mistaken by-product of our language.) (Can you imagine a language without "I"? Can you imagine the people who would have such a language?)

But the brain is part of the body, one might respond. To this, Shoemaker points out that we're not really using the brain as our criterion either: "If upon regaining consciousness, Brownson were to act and talk just as Robinson had always done in the past" (24), then we'd say it's Robinson even though he has Brown's brain. So we're actually using psychological features—personality and memory of past events. And since we think such psychological features are causally related to the brain, we figure Brownson will have the personality features and memories of Brown rather than of Robinson.

But does the rest of one's body have *no* causal relationship to one's psychological features? Not even developmentally? That is, if one's body happens to be robust, might one not develop more confidence than if one's body happened to be otherwise? And so might not Brown with Robinson's body become, sooner or later, a different person than he was when he had his own body? (How long would Brownson remain Brown if Brown had been white skinned and Robinson had been black skinned, or if Brown had been female and Robinson had been male?)

## PARFIT'S FISSION

My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people believes that he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me. And he has a body that is very like mine. . . .

. . . What happens to me?

Source: Derek Parfit. *Reasons and Persons*. Oxford, UK: Oxford University Press, 1984. 254–255.

Describing first a case like Shoemaker's Brownson (see "Shoemaker's Brownson"), Parfit suggests that "you go where your brain goes"—the resulting person is Brown. Then, since it is actually possible to survive with only one functioning hemisphere (consider stroke victims), Parfit reasons also that "you go where only half your brain goes"—so if half of your brain is destroyed and the other half transplanted into another body, the resulting person is indeed you. But what if the other half is not destroyed? This is the case Parfit considers here (attributing it to David Wiggins, who modifies, in *Identity and Spatio-Temporal Continuity* (1967), Shoemaker's Brownson, postulating that Brown's brain is split and the two halves housed in two different bodies).

Parfit considers four possibilities. The first is that he does not survive. But, he reasons, since he would survive if his whole brain had been successfully transplanted and since people do survive with half their brain injured, this can't be the case. The second and third possibilities are that he survives as one or the other of the two resulting people. But if the halves are identical, why would he survive as only one—and which one? The fourth possibility is that he survives as both of the two resulting people. But one person can't be two people. (Why not? See "Parfit's Teletransporter"—what if the original you isn't lost in the replication process?)

Or, Parfit suggests, perhaps he does survive the operation and "its effect is to give me two bodies, and a divided mind" (256). In fact, people with the connection between the two hemispheres of their brains severed *do* have a divided mind, two *separate* spheres of consciousness. But, Parfit says, this "solution" involves "a great distortion in our concept of a person" (256).

Parfit then suggests that the question "Shall I be one of these two people, or the other, or neither?" is an empty question, and he argues for giving up altogether the language of identity. Identity is an all-or-nothing thing, but the things that are really important to us (like psychological connectedness) are matters of degree.

The more important, and perhaps more appropriate, question, he says, is one of survival—and he can say he survives without having to say he is one of those people. But who is it who wants to survive? (And if "who" is not important, *why* do "you" want to survive?)

## AYER'S ROBINSON CRUSOE

Imagine a Robinson Crusoe left alone on his island while still an infant, having not yet learned to speak. Let him, like Romulus and Remus, be nurtured by a wolf, or some other animal, until he can fend for himself; and so let him grow to manhood. He will certainly be able to recognize many things upon the island, in the sense that he adapts his behaviour to them. Is it inconceivable that he should also name them?

Source: A. J. Ayer. "Can There Be a Private Language?" *Proceedings of the Aristotelian Society* supplementary vol. 28 (1954): 63–76, 70.

Ayer believes it is *not* inconceivable that his Robinson Crusoe would name things, and he offers this thought experiment as evidence against Wittgenstein that one *can* have a private language (see "Wittgenstein's 'S'"). Wittgenstein might reply that, given our faulty memories, Ayer's Crusoe could never be sure he was following his own rules. But isn't there a difference between using a language incorrectly or inconsistently and not being able to use it at all?

Furthermore, Ayer concedes that there is a problem with endowing signs with meaning, but "it is no less of a problem in the case where the object for which the sign is supposed to stand is public than in the case where it is private" (68–69). So a private language is neither more nor less impossible than a public language.

Indeed, especially given the lack of objective standards, one could insist there *must* be a private language—if only because the subjective experience described by words can't possibly be publicly understood. That is, insofar as we use words to describe our subjective experience, which is known only to ourselves, we *necessarily* use a private language. But can that be called a language? How exactly should we define "language"?



## QUINE'S GAVAGAI

[C]onsider the linguist who, unaided by an interpreter, is out to penetrate and translate a language hitherto unknown. All the objective data he has to go on are the forces that he sees impinging on the native's surfaces and the observable behavior, vocal and otherwise, of the native. . . .

. . . A rabbit scurries by, the native says "Gavagai," and the linguist notes down the sentence "Rabbit" (or "Lo, a rabbit") as tentative translation, subject to testing in further cases. . . . For, suppose the native language includes . . . "Animal," "White," and "Rabbit". . . . How then is the linguist to perceive that the native would have been willing to assent to ["Animal"] in all the situations where he happened to volunteer ["Rabbit"], and in some but perhaps not all of the situations where he happened to volunteer ["White"]? Only by taking the initiative and querying combinations of native sentences and stimulus situations so as to narrow down his guesses to his eventual satisfaction.

So we have the linguist asking "Gavagai?" in each of various stimulatory situations, and noting each time whether the native assents, dissents, or neither. But how is he to recognize native assent and dissent when he sees or hears them? . . . [S]uppose that in asking "Gavagai?" and the like, in the conspicuous presence of rabbits and the like, he has elicited the responses "Evet" and "Yok" often enough to surmise that they may correspond to "Yes" and "No," but has no notion which is which.

---

Can the linguist ever come to know what "Gavagai" means?

Source: Willard Van Orman Quine. *Word and Object*. Cambridge, MA: MIT Press, 1960. 28, 29.

With this thought experiment, Quine is investigating the relationship between words and the objects they represent. He suggests that, at best, we can come to know not what "gavagai" means but only what prompts or stimulates the utterance of "gavagai." And even then, the best we can do is achieve an approximation.

One difficulty, Quine points out, is the role of prior collateral information: "[The native] may assent on the occasion of nothing better than an ill-glimpsed movement in the grass, because of his earlier observation, unknown to the linguist, of rabbits near the spot. Since the linguist would not on his own information be prompted by that same poor glimpse to assent to 'Rabbit?', we have here a discrepancy between the present stimulus meaning of 'Gavagai' for the informant and that of 'Rabbit' for the linguist. . . . [Or] there may be a local rabbit-fly, unknown to the linguist, and recognizable some way off by its long wings and erratic movements; and seeing such a fly in the neighborhood of an ill-glimpsed animal could help a native to recognize the latter as a rabbit" (37). Another difficulty, Quine says, is that "the native may dissent from 'Gavagai' in plain sight of the rabbit's ears, because the rabbit is in no position for shooting; he has misjudged the linguist's motive for asking 'Gavagai?'" (39). Furthermore, suggests Quine, maybe the objects to which "gavagai" applies are not rabbits after all, but "mere stages, or brief temporal segments, of rabbits" (51), or "undetached parts of rabbits" (52), or "the fusion . . . of all rabbits, . . . that single though discontinuous portion of the spatiotemporal world that consists of rabbits" (52), or the "recurring universal [of] rabbithood" (52)—for in all cases, what would prompt "Gavagai" would also prompt "Rabbit": "Point to a rabbit and you have pointed to a stage of a rabbit, to an integral part of a rabbit, to the rabbit fusion, and to where rabbithood is manifested" (52–53).

To determine what exactly "gavagai" means, the linguist would have to ask, while pointing, questions like "Is that one gavagai or two?"—that is, would have to know other words of the native's language. "The whole apparatus is interdependent" (53), says Quine—meaning can be determined only within the context of other meanings, within the context of the whole language.

Is the situation as "hopeless" as Quine suggests? (If we can't establish meaning in such a benign case as "gavagai" and "rabbit," where one can at least point to something, what about words that refer to abstract relationships? Imagine the linguist trying to establish the meaning of something like "Neutrinos lack mass.") Isn't learning one's first language as a child a similar case of "radical translation"—and don't we achieve more than "approximation"? (Then again, is more than approximation really necessary? See "Wittgenstein's Games.")

# PUTNAM'S TWIN EARTH

[S]uppose that somewhere in the galaxy there is a planet we shall call Twin Earth . . . [that], apart from the differences we shall specify, . . . is *exactly* like Earth. . . .

One of the peculiarities of Twin Earth is that the liquid called “water” is not H<sub>2</sub>O but a different liquid whose chemical formula . . . I shall abbreviate . . . as XYZ. I shall suppose that XYZ is indistinguishable from water at normal temperatures and pressures. In particular, it tastes like water and it quenches thirst like water. Also, I shall suppose that the oceans and lakes and seas of Twin Earth contain XYZ and not water, that it rains XYZ on Twin Earth and not water, etc.

If a spaceship from Earth ever visits Twin Earth, then the supposition at first will be that “water” has the same meaning on Earth and on Twin Earth. This supposition will be corrected when it is discovered that “water” on Twin Earth is XYZ, and the Earthian spaceship will report somewhat as follows: “On Twin Earth the word ‘water’ means XYZ.” . . .

Symmetrically, if a spaceship from Twin Earth ever visits Earth, then the supposition at first will be that the word “water” has the same meaning on Twin Earth and on Earth. This supposition will be corrected when it is discovered that “water” on Earth is H<sub>2</sub>O, and the Twin Earthian spaceship will report: “On Earth the word ‘water’ means H<sub>2</sub>O.”

Now let us roll the time back to about 1750. At that time chemistry was not developed on either Earth or Twin Earth. The typical Earthian . . . did not know water consisted of hydrogen and oxygen, and the typical Twin Earthian . . . did not know “water” consisted of XYZ. Let Oscar<sub>1</sub> be

Source: Hilary Putnam. “The Meaning of Meaning.” In *Minnesota Studies in the Philosophy of Science*. Vol. 7: *Language, Mind, and Knowledge*. Keith Gunderson, ed. Minneapolis: University of Minnesota Press, 1975. 131–193. 139–141.

such a typical Earthian . . . and let Oscar<sub>2</sub> be his counterpart on Twin Earth. You may suppose that there is no belief that Oscar<sub>1</sub> had about water that Oscar<sub>2</sub> did not have about “water” [and] that Oscar<sub>1</sub> and Oscar<sub>2</sub> were exact duplicates in appearance, feelings, thoughts, interior monologue, etc. Yet the extension [the set of things the term is true of] of the term “water” was just as much H<sub>2</sub>O on Earth in 1750 as in 1950; and the extension of the term “water” was just as much XYZ on Twin Earth in 1750 as in 1950.

Will Oscar<sub>1</sub> and Oscar<sub>2</sub> understand the term “water” to have the same meaning?

The broad question Putnam is trying to answer is “What is the meaning of meaning?” A standard view of “meaning” is that “knowing the meaning of a term is just a matter of being in a certain psychological state” (135), so if two people understand a word differently, they must be in different psychological states (mental states). Putnam hopes that his Twin Earth thought experiment shows otherwise: Oscar<sub>1</sub> and Oscar<sub>2</sub> are physically identical, so when they think “water,” they are in the same psychological or mental state; *however*, when they think “water,” they are *not* thinking the same thing—one is thinking about H<sub>2</sub>O and the other is thinking about XYZ. Thus, Putnam concludes, meaning is *not* “just in the head.”

Rather, Putnam argues, meaning is determined by the external environment—the truth of the matter. In this regard, his view is decidedly realist (see “Putnam’s Brain in a Vat”). And yet, because we often don’t really know the truth of the matter (most of us couldn’t tell whether the wet stuff we’re talking about is H<sub>2</sub>O or XYZ), Putnam says meaning is also determined by the sociolinguistic conventions or practices of a community. In this regard, his view is relativist. Is this a problem?

Furthermore, if neither Oscar<sub>1</sub> nor Oscar<sub>2</sub> knows chemistry, *don’t* they mean the same thing when they say “water” (something like “clear tasteless thirst-quenching liquid”)? Is there a difference between “mean” as in “intend” and “mean” as in “refer to”?

Lastly, does Putnam’s point apply only to words for natural or material objects? What about, for example, words such as “red” and “pain”?

## MOLYNEUX'S BLIND MAN

Suppose a man *born* blind, and now adult, and taught by his *touch* to distinguish between a cube and a sphere of the same metal, and nighly of the same bigness, so as to tell, when he felt one and the other, which is the cube, which the sphere. Suppose then the cube and sphere placed on a table, and the blind man be made to see: *quaere*, whether by his sight, before he touched them, he could now distinguish and tell which is the globe, which the cube?

Source: William Molyneux. In a letter dated 1693 to John Locke, quoted by Locke in the second edition (1694) of *An Essay Concerning Human Understanding*. Book 2, Chapter 9, Section 8. As collated and annotated by Alexander Campbell Fraser. New York: Dover, 1959. Volume 1. 186–187.

Molyneux predicts that the man could *not* distinguish between the globe and the cube by sight alone because he hasn't had the necessary experience; he hasn't learned how visual perceptions relate to physical realities. Molyneux believes his thought experiment disproves the existence of innate ideas (see "Plato's Equal Portions of Wood and Stone") that are argued to exist by rationalists (usually on the basis of universal agreement on certain principles); rationalists would say the man *would* be able to recognize and distinguish the globe and cube, by matching what he now sees with the ideas he has of them in his mind—ideas he has always had, independent of his experience through life. Empiricists such as Molyneux and Locke, however, say we are not *born* with such ideas, with such knowledge about the physical world; rather, when we are born, as Locke says, our minds are a *tabula rasa* (a blank tablet) and we *acquire* knowledge through sensory experience and the subsequent reasoning of association and abstraction. Is Molyneux correct in his prediction—and its implication?

Contemporary philosopher Janet Levin modifies Molyneux's experiment (in "Could Love Be like a Heatwave?"), postulating that if the man had learned, while blind, geometric facts about three-dimensional figures and had heard statements about such figures made by sighted people, and then, when newly sighted, had been shown other geometrical figures and told what they were, he *would* be able to distinguish between the globe and cube. If that is so, what are the implications for how we know what we know?

# MILL'S CHAOTIC WORLD

Were we to suppose . . . that the present order of the universe were brought to an end, and that a chaos succeeded in which there was no fixed succession of events, and the past gave no assurance of the future; if a human being were miraculously kept alive to witness this change, he surely would soon cease to believe in any uniformity, the uniformity itself no longer existing. If this be admitted, the belief in uniformity either is not an instinct, or it is an instinct conquerable, like all other instincts, by acquired knowledge.

Source: John Stuart Mill, *A System of Logic*. Book 3, Chapter 21, Section 1. 1843. J. M. Robson, ed. Toronto: University of Toronto Press, 1973. 565–566.

And since, according to Mill, it is from the many “uniformities of sequence” in our world that we have generalized the universality of cause and effect (see “Hume’s Constant Conjunction”), he is suggesting with this thought experiment not merely that a belief in uniformity is not an instinct (or an innate idea—see “Plato’s Equal Portions of Wood and Stone”), but that the law of causality is not an instinct or innate idea. Rather, Mill argues, it is a habit of thought, an induction (a generalization based on particulars), formed by our experience of the world.

Can other supposed innate ideas be disproved in a similar fashion? (See “Kant’s *A Priori* Space.”)

## GOODMAN'S GRUE

Suppose that all emeralds examined before a certain time  $t$  are green. At time  $t$ , then, our observations support the hypothesis that all emeralds are green; and this is in accord with our definition of confirmation. Our evidence statements assert that emerald  $a$  is green, that emerald  $b$  is green, and so on; and each confirms the general hypothesis that all emeralds are green. So far, so good.

Now let me introduce another predicate less familiar than "green." It is the predicate "grue" and it applies to all things examined before  $t$  [that] are green [and all things examined after  $t$  that] are blue. Then at time  $t$  we have, for each evidence statement asserting that a given emerald is green, a parallel evidence statement asserting that that emerald is grue. And the statements that emerald  $a$  is grue, that emerald  $b$  is grue, and so on, will each confirm the general hypothesis that all emeralds are grue. Thus according to our definition, the prediction that all emeralds subsequently examined will be green and the prediction that all will be grue are alike confirmed by evidence statements describing the same observations. But if an emerald subsequently examined is grue, it is blue and hence not green. Thus although we are well aware which of the two incompatible predictions is genuinely confirmed, they are equally well confirmed according to our present definition.

Source: Nelson Goodman. *Fact, Fiction, and Forecast*. 4th ed. Cambridge, MA: Harvard University Press, 1983. 73.

Philosophers have traditionally distinguished between deduction (reasoning from generals to particulars) and induction (reasoning from particulars to generals). The latter, typically involving predictions, is problematic because there is no *logical* reason (nor empirical data, because an unperceived future event is involved) to justify such claims: that the sun has always risen in the past imposes no logical necessity that it rise tomorrow—we merely assume regularity about the facts involved (see "Hume's Constant Conjunction").

However, Goodman points out, deductive claims are considered valid as long as they correctly follow the rules (that is, it doesn't matter whether or not the conclusion is in accord with the facts—truth is required for soundness, not for validity). So, he asks, what rules would justify inductive claims? "That a given piece of copper conducts electricity increases the credibility of statements asserting that other pieces of copper conduct electricity, and thus confirms the hypothesis that all copper conducts electricity," Goodman notes. "But," he continues, "the fact that a given man now in this room is a third son does not increase the credibility of statements asserting that other men now in this room are third sons, and so does not confirm the hypothesis that all men now in this room are third sons" (73). He therefore suggests that only inductive claims that are *lawlike* statements (such as "Copper conducts electricity") can be confirmed by (particular) past instances.

However, as his thought experiment about "grue" shows, past instances may confirm two incompatible statements: before time  $t$ , each emerald we find is green and grue, particulars which confirm equally well the general statements that "Emeralds are green" and "Emeralds are grue"; but since after time  $t$ , all grue emeralds are blue, we seem to have confirmed both "All emeralds are green" and "All emeralds are blue." What went wrong?

Goodman's response is that a definition of "lawlike" is needed—"All emeralds are grue" is apparently *not* a lawlike statement. So for what sorts of statements *is* it valid to reason from particular instances to general claims—that is, what *is* a "lawlike" statement? (What is it—if anything—about the statement "All emeralds are grue" that makes it problematic?)

## PLATO'S RING OF GYGES

The story is that [Gyges] was a shepherd in the service of the ruler of Lydia. There was a violent rainstorm and an earthquake which broke open the ground and created a chasm at the place where he was tending sheep. Seeing this and marveling, he went down into it. He saw . . . a corpse which seemed of more than human stature, wearing nothing but a ring of gold on its finger. This ring the shepherd put on and came out. . . . As he was sitting among the others he happened to twist the hoop of the ring towards himself, to the inside of his hand, and as he did this he became invisible to those sitting near him and they went on talking as if he had gone. He marvelled at this and, fingering the ring, he turned the hoop outward again and became visible. Perceiving this he tested whether the ring had this power and so it happened: if he turned the hoop inwards he became invisible, but was visible when he turned it outwards. When he realized this, he at once arranged to become one of the messengers to the king. He went, committed adultery with the king's wife, attacked the king with her help, killed him, and took over the kingdom.

Now if there were two such rings, one worn by the just man, the other by the unjust, no one, as these people think, would be so incorruptible that he would stay on the path of justice or bring himself to keep away from other people's property and not touch it, when he could with impunity take whatever he wanted from the market, go into houses and have sexual relations with anyone he wanted, kill anyone, free all those he wished from prison, and do the other things which would make him like a god among men. His actions would be in no way different from those of the other and they would both follow the same path.

Source: Plato. *The Republic*, Book II, 380–370 BCE. G. M. A. Grube, trans. Indianapolis: Hackett, 1974. As reprinted in *Moral Philosophy: Selected Readings*. George Sher, ed. San Diego: Harcourt Brace Jovanovich, 1987. 235–243. 237.

This thought experiment is part of a larger discussion about justice and the value of being just (why should we do the right thing?), part of a larger discussion still about what kind of society, and what kind of government, is best. Responding to the claim that success in this world comes to those who are *unjust*, Plato's character "Socrates" (who speaks for Plato himself) says that injustice leads to hatred and fighting, while justice results in harmony and a working together of the various parts—both within society as a whole and within the individual.

"Glaucou" then asks Socrates to imagine the scenario he has described. That the just man, if invisible, would act as badly as the unjust man proves, says Glaucou (who is assuming people will do what they believe to be in their own best interest unless compelled otherwise), that being just is *not* in our own best interest: "Every man believes that injustice is much more profitable to himself than justice" (237).

But *would* all of us do whatever we wanted if we knew we wouldn't be caught? If so, does that prove that being good is not good for us? (In which case, why *should* we do the right thing?) Or does it just prove that we don't know, or don't act according to, what's good for us?

# MOORE'S TWO WORLDS

Let us imagine one world exceedingly beautiful. Imagine it as beautiful as you can; put into it whatever on this earth you most admire—mountains, rivers, the sea; trees, and sunsets, stars, and moon. Imagine these all combined in the most exquisite proportions, so that no one thing jars against another, but each contributes to increase the beauty of the whole. And then imagine the ugliest world you can possibly conceive. Imagine it simply one heap of filth, containing everything that is most disgusting to us, for whatever reason, and the whole, as far as may be, without one redeeming feature. . . . [No one] ever has or ever, by any possibility, *can*, live in either, can ever see and enjoy the beauty of the one or hate the foulness of the other. . . . [E]ven so, supposing them quite apart from any possible contemplation by human beings, . . . is it irrational to hold that it is better that the beautiful world should exist than the one which is ugly? Would it not be well, in any case, to do what we could to produce it rather than the other?

Source: G. E. Moore. *Principia Ethica*. 1903. Cambridge, UK: Cambridge University Press, 1959. 83–84.

Moore postulates his two worlds in order to challenge the claim that things are good only in relation to human existence. If, as Moore suggests, the beautiful world he describes must be considered a greater good than the ugly world—despite the fact that neither world will be seen by anyone—then “we shall have to include in our ultimate end something beyond the limits of human existence” (84). (Such as?) By implication, hedonism, the view that (human) happiness or pleasure is the sole good, must be rejected.

A criticism of Moore’s thought experiment is that it is logically impossible: if the beautiful world is indeed beautiful, it must have been seen by someone, if only an imaginary someone—otherwise, how can it be called beautiful? In other words, the concept of beauty necessarily entails, because it is defined by, human presence. (Must hedonism, therefore, be accepted?)

# NOZICK'S EXPERIENCE MACHINE

Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. . . . If you are worried about missing out on desirable experiences, we can suppose that business enterprises have researched thoroughly the lives of many others. You can pick and choose from their large library or smorgasbord of such experiences, selecting your life's experiences for, say, the next two years. After two years have passed, you will have ten minutes or ten hours out of the tank, to select the experiences of your *next* two years. Of course, while in the tank you won't know that you're there; you'll think it's all actually happening. . . . Would you plug in?

Source: Robert Nozick. *Anarchy, State, and Utopia*. New York: Basic Books, 1974. 42–43.

The question Nozick is asking is “What else can matter to us, other than how our lives feel from the inside?” (43). Nozick considers three possible answers. Perhaps what matters is the desire to do, rather than just experience, certain things; the experience machine doesn't seem to allow those desires. (But what if the machine could—what if it could enable not only the experience but also the desire for the experience?) Or perhaps what matters is the desire to be a certain sort of person; a blob in the experience machine tank can't be said to be, for example, courageous, kind, intelligent, witty, or loving (43). (Well, suggests Nozick, “imagine a transformation machine which transforms us into whatever sort of person we'd like to be” [44].) Or perhaps what matters is the possibility of some sort of transcendent experience; the experience machine is limited to providing only experiences conceived by humans. Nozick concludes that we would *not* plug in because “what we desire is to live (an active verb) ourselves, in contact with reality” (45).

Nozick's thought experiment is part of a larger discussion about the moral limits to what we may do to each other. With it, he questions the hedonistic view (which considers only one's experiences, one's pleasure and pain, in determining such limits) and the derivative utilitarian view (which advocates that we should do that which promotes the greatest good, measured in terms of pleasure, for the greatest number). Since something besides our experience (of pleasure) matters to us, Nozick suggests that something else should be considered when determining what actions are morally permissible. (What might this “something else” be?)



## DEWEY'S FINELY WROUGHT OBJECT

Suppose . . . that a finely wrought object, one whose texture and proportions are highly pleasing in perception, has been believed to be a product of some primitive people. Then there is discovered evidence that proves it to be an accidental natural product. . . .

---

Is it a work of art?

Source: John Dewey. *Art as Experience*. New York: Capricorn Books, 1934. 48.

Dewey claims that even though the object is precisely what it was before the discovery, immediately upon such discovery, "it ceases to be a work of art" (48) and "belongs in a museum of natural history, not in a museum of art" (48). Why? His answer is that to qualify as art, a work must be "framed for enjoyed receptive perception" (48). Mere technique or virtuosity on the part of the artist is not enough, but neither is simply being perceived. What is important, says Dewey, is the *relation* between "doing and undergoing" (48), the *connection* between production and reception: "The doing or making is artistic when the perceived result is of such a nature that *its* qualities *as perceived* have controlled the question of production" (48). Furthermore, just as the artist must create with the perception always in mind, the beholder must perceive with the creation in mind.

So what if that connection were intended but failed to occur—what if, for example, a listener didn't hear what it was the musician intended to be heard? Is that just bad art? Failed art? Nonart?

And what if no connection were intended—what if someone painted a painting, then wrapped it up immediately and sent it into outer space never to be seen by anyone?