

THE TRADITION OF THOUGHT EXPERIMENTS IN EPISTEMOLOGY

Rolf George

I

ERNST Mach cites as one of the classic thought experiments in physics the demonstration, given by Galileo, that the distance traversed by a freely falling body cannot be proportional to its weight. For, if we fasten with a string a 10 lb. weight to a 100 lb. weight, it would seem that this contraption ought to fall more slowly than the 100 lb. weight alone, being held back by the lighter body. But, plainly, the length of the string cannot be a factor in this. Reducing it, therefore, to zero, we get the lighter weight to attach directly to the heavier. This item should now fall both faster and slower than the 100 lb. weight, showing the Aristotelian view to be untenable.

Whether this style of thinking, which presumably obtained some currency in the physics of the time, had influence on epistemological speculation I do not know. It is certain, however, that in the eighteenth century a new style of epistemology became current, which is perhaps best described as epistemological thought experimentation. The interest in criteriological issues begins to wane; no longer is there, except in Hume, much interest in the defense or refutation of scepticism, which is, instead, summarily dismissed, by Kant, for instance, as a scandal for philosophy. In its place we find imaginative reconstructions of the cognitive process. Assumptions are made about initial representational states of perceivers. It is then argued that, given these initial states, certain mental processes must occur that lead to the representations of objects and events with which we are familiar. In the end, lessons are drawn from these constructions. These lessons are what we are wont to call "philosophical Positions": Transcendental Idealism, Hume's Scepticism, Common Sense Philosophy. It is evident that these lessons could not have been drawn if there had been no constructions or thought experiments to draw them from.

I now want to survey, rather briefly, some of the features of these thought experiments, focussing on arguments that were offered in connection with the three concerns I just mentioned. First, some argument or evidence is usually offered that the initial epistemic state from which construction begins is of a certain sort. There are arguments, secondly, that purport to shed light on the mental processes that generate the constructions, and, thirdly, argument

is often offered on the construction as a whole. This is what I have just called the "lesson."

II

Most epistemological experimenters have assumed that it is not obvious what the initial or primitive epistemic states are like. If one takes this line some sort of argument is needed to ascertain them, and here we find a whole grab-bag of diverse considerations. The reason for this is that ordinary adult perception is of familiar objects and, by supposition, always presupposes some processing beyond the initial phase.

1. For the sake of completeness, I mention first the argumentative minimalists. Kant thought that "through long practice" we become aware of the role of sensation in experience (CPR B1 f.), though as his argument progresses we realize that sensations must be what Aristotle called the proper objects of the senses and that, for example, there can be no sensations of shape, a point on which I have to say some more. But there is no independent argument that shows us what they are. We learn incidentally that they are such things as "red," "warm," "hard," and the like, and not "square" or "earlier." In this department he probably relied on the work of others, or on the state of the art, as he saw it, in mental science.

Much earlier than Kant, Malebranche had also assumed that the cognitive process begins with sensations, which are non-intentional mental occurrences, much like feelings, and are specific to the senses (Malebranche, I, 151f.). This non-intentional character remains a theme in sensationism: the initial phase of the cognitive process is to be thought on the analogy of such things as pain and feelings of warmth, rather than on the analogy of images or concepts. Malebranche appears to offer no argument for this, but he set a problem that engaged many others, namely to identify the more central processes of the mind that are responsible for creating, in some way, the perceptions of things. He himself suggests that judgment is involved, but that is as far as he goes.

2. More interesting is the style of thought offered by Condillac. He imagines a perceiver, a mind encased in heavy marble so that nothing unwanted can get in. He then imagines that only one sensory pathway is opened at a time, beginning with the sense of smell. If such a perceiver had a rose under its nose, then, Condillac thought, it would report its sensory taking as "I am rose odor" (Condillac, p. 3). This way of putting the matter is to suggest that there would be no attribution of a property to a perceived object, but only an attribution of a kind of feeling to a self. Condillac was perhaps not fully aware of the problems in this that were subsequently revealed by Kant, but he did not in any obvious way fall into the trap of taking "I am rose odor" to signify attribution to an abiding self.

The most profoundly influential step was taken when Condillac con-

templated the takings that would result from the sense of sight taken in isolation. He thought (and I am not sure that he had predecessors in this) that only light and colors would then be perceived, the reports being "I am red," "I am green," and so on (Condillac, p. 61). Some further thought must have motivated the exclusion of shape. I conjecture that Condillac simply assumed that the initial perceptual state must be wholly subjective, whereas only phenomenal *objects* have shape. There may also have been considerations of parsimony. If he could show, as he thought he could, that the mind *constructs* shape, then he need not suppose that it was given.

Kant's motivation for denying sensations of shape was of a different sort. Two thoughts come to mind: Since, as he thought, the mind cannot take apart what it has not put together, and since shapes are complex, they must be the product of "synthesis," more central functions of the mind. Also, we should remind ourselves that his account of the foundations of geometry would not work if there were sensations of shape: the mind must *draw* the lines it sees according to its own laws, which it then retrieves under the guise of geometric theorems.

3. The theme of imagining a mind with less than all the senses or with only one sense is a common device for discovering the nature of sensory taking. This often seems to be based on the assumption that perception of ordinary objects is based on the synergy of several senses, as it is in Berkeley and Reid. Both consider at some length what it is that we could take in by sight, if we had neither locomotion, nor sense of touch. Berkeley thought that distance from the observer could then not be observed, but he seems to have held that visual takings have not only color, but also shape. We should note this as a major difference between the "idea philosophy" and sensationism. Reid gives a complex account of the geometry of purely visual objects, which results from the central projection, with the eye as vantage point, of shapes upon the inside of a large sphere. This geometry has such features as that straight lines will meet twice, and the like (Reid 1763, VI/IX). The passages are part of a pervasive concern of Reid's to show that sensations are exceptionally elusive, and in many cases inaccessible to consciousness. His point here is that a mind equipped only with stationary vision would see things in this way, but we do not. Visual takings, like all other sensations, are necessary conditions for perception, but "immediately disappear and are forgot" (Reid 1785, II, 17; p. 315) and, indeed, have features that even the minutest attention does not disclose.

4. On the assumption that we must have a sufficiency of sensational information for the construction of familiar objects, it seems possible to conduct experiments that point to unperceivable "sensations." For instance, if successful monocular distance perception is possible to, say 15 feet (even of unfamiliar objects in an unstructured setting), we must, on this view, attribute this to sensations corresponding to the accommodation of the lense of the eye. But this cannot be felt. In a similar spirit, Herbart

and Lotze postulated other "topogenic" sensations which accompany the pivoting of the eye. There are, of course, many other experiments that would seem to disclose the panoply of sensations, among them perhaps even such things as the attempts, by Cheselden and others, surgically to provide vision to a congenitally blind person. Since in such a case the synergy of the senses would have had no opportunity to assert itself, we would learn what it is that we see when we just see.

It should surprise no one that, given these various approaches, different lists of sensations are extant. In Kant I found these: Taste of wine, color, sound, warmth, red, black, sweet, hard, pain, and the shudder one feels when reading certain portions of Haller's poem about eternity. Reid and others postulate topogenic sensations, Pius IX reported having sensations of infallibility, and von Neumann and Morgenstern both had sensations of preference, and indeed claim that everyone has them.

5. Carnap, in the *Logical Structure of the World*, assumes that there are elementary experiences, which are global sensational states containing all momentary sensory input. On these he defines a relation "recollection of similarity" which then allows him, through a procedure he calls quasi-analysis, to construct sets and sets of sets so that, at a certain level he arrives at entities that he identifies with sensations in the more traditional sense: sounds, colors, and so on. In this case the choice of basis is obviously and avowedly determined by the system itself: it is the simplest, given the constructive machinery that is postulated.

6. The preceding attempts at identifying initial cognitive states have one thing in common: it is their tentativeness. The lists of sensations are incomplete; we only get examples. There is some doubt about their exact nature. No certainty attaches to sensations. These epistemic systems are not foundational in the sense that the initial states are also indubitable, incorrigible or most accessible. Indeed, it is really a category mistake to ask if they are. These are not the kinds of predicate that attach to them.

The mistake to bring considerations of incorrigibility into the matter was left to the twentieth century. At least some eighteenth century philosophers, Kant and Condillac among them, would have had sympathy for the view that physical objects are in some sense constructions out of sense data. But they would have regarded with incomprehension the claim that sense data are incorrigible and indubitable sensory takings, arrived at by disregarding all those features in a perceptual situation which allow the least doubt. Twentieth century phenomenalism is, for this reason, a completely different enterprise from the thought experiments I began to describe above.

III

The thought experiment now proceeds to the next task, which is to identify,

if this is possible, the procedures and capacities needed to progress from the initial cognitive states to the perception of objects. The conjectures we find here (which are of course often given out to be certainties) vary according to the base that was assumed.

1. Berkeley took the visual sense to be paradigmatic. It was probably a consequence of this that he took the initial takings to be already highly structured, particularly for the sense of sight, but probably also for touch and hearing. There is therefore less to do when it comes to describing how the mind comes to perceive things, or things at a distance. His moves are familiar enough. "Distance from the observer," for example, is a construction involving sight, touch, and kinesthetic sensation: There is first a visual field with individuable visual objects in it. Upon certain sensations of movement (later to be interpreted as movement toward the object) a certain touch sensation is obtained and becomes associated with the visual object. The subjective time lapse between the first sighting and eventual touching becomes the measure of distance, and the distances of visual objects are thus learned through "custom" and require the three items mentioned (*New Theory of Vision*, sects. 44ff.). Berkeley in fact says that this is what distance is, a point to which I shall return later. A distance estimate is thus at bottom a forecast of what will happen to the sense of touch after appropriate motions of the body. The distance estimate, to be possible at all, must rely on visual and other sensational cues, which Berkeley takes a fair amount of trouble to describe, and, being a forecast, it is impossible without experience. A perceiver who has gone through this process sufficiently often, will begin to "see" things at a distance, and to be three dimensional. This is no longer naive seeing, but visual perception informed by the experience of associated (or associable) touch sensations. There are other constructions, of position, of size, and so on, all somewhat vague and turning on "custom" and the synergy of the senses. Like many others, Berkeley only gives some constructional high points and is interested chiefly in the lessons he can draw from them. Distance, for example, is important to him since he must defuse the criticism that idealism is incompatible with the plain fact that we see things outside us. This challenge is answered by explaining, through a thought experiment, how a being that is given only sensations (or ideas, as he would have it), but has some additional capacities, can ascertain distances to objects. This becomes an argument in support of idealism through a definitional move, namely, that to say that an object is at a distance is just the same as to say that a touch sensation can be had after certain sensations of movement are experienced. (Now of course there are a lot of unanswered questions: the sun is at a distance from the moon, and we from both of them. Berkeley rather sweepingly deals with this by saying that this means that if we were in a certain position in space, they would subtend a certain angle. But it is not my aim here to point to unsolved problems, of which there are legion.)

2. Let me now turn to Condillac. How could a being that has by sight only

color sensations come to see shapes? That is, how can a being whose visual takings have no internal structure come to obtain that result? The answer commonly given to this is that it must *scan*. According to Burke, some of his contemporaries thought that "there is but one point of any object painted on the eye in such manner as to be perceived at once; but by moving the eye we gather up with great celerity, the several parts of the object, so as to form one uniform place" (Burke 1759, p. 259).

Condillac held just this view, except that he must have realized that if the scanning were arrested, if the perceiver took a snap shot, no point would be seen. For to see a point is to see it against a background, and this is impossible, by assumption. It must be supposed, then, that the point wholly fills the mind. It is, in fact, a color sensation, which subsequently may be interpreted as the perception of a point. A series of color sensations might then be interpreted by a Condillacian perceiver as the perception of a line. But it could not do this, Condillac observes, unless there is memory. If each new sensation crowds out the preceding one without trace, then no line, no shape can be perceived. The perceiver must therefore know that it has run through a series of sensations. But even this is not enough. Unless it also judges that these perceptions form one object, it will still not have seen a line. Condillac says this:

The...eye cannot grasp the whole of the simplest shape...until it has noticed successively all its parts. It must make a judgment on each part, and another judgment on the whole of them taken together. I must say: here is one side, here a second, and here a third; here, then, is a space bounded by three sides, and from this results this triangle. (Condillac, p. 68.)

Nor is this the whole story. Our thought experiment tells us that a perceiver equipped with all these things, even if they worked the way Condillac thought they did, might never come to find a line. The eye might scan randomly, given us disorganized sequences of color sensations. Even if it were to look at a physical triangle, it might not find it but stray back and forth across the three lines. It is to Condillac's great credit that he saw this problem. The hand, he thought, can follow contours by feeling them, and the eye, following the hand, can then learn to do the same. But it is not clear how this can happen, since the eye cannot see the hand, and the hand itself, we must suppose, is also in need of instruction. (Kant, I believe, was also aware of this problem, and thought to solve it by means of the Schematism—but that is a very risky interpretation of the Schematism chapter that I do not want to push here.)

3. Before I go on to the matter of lessons drawn from these thought experiments, I want to pay some attention to Kant. I suppose, and have previously argued, that he followed Condillac both in the adoption of the sensationist basis and in certain other ways. For instance, Condillac's three steps in the perception of shape: uptake, memory and judgment, are just the parts of the threefold synthesis of the Transcendental Deduction in the

first edition of CPR. Let me then pose a problem as I think it must have occurred to Kant.

A Condillacian perceiver, at the point where we have left Condillac, cannot have much of a conception of time. There is experience of succession, to be sure, but no sense of duration, since a single sensation, involving no change, if it is all there is in a mind, cannot be perceived to last a long or short time. Condillac comments on this, and Kant agrees with him. There is, secondly, the problem that the temporal order of perceived objects cannot just be the order of sensations. We have already seen that the perception of the triangle involved succession even though the triangle itself is not judged to have any succession in it. In addition, from the viewpoint of the finished construction, the sequence of sensations must appear simply to be a partially ordered set: there will appear to be distinct sequences (each of which might be held in the mind by itself) whose order with respect to each other remains undetermined, given the powers and performances so far attributed to the perceiver. There is, thirdly, the problem of the simultaneity of objects that cannot be brought into one image. Such a simultaneity cannot even be perceived in the way in which the simultaneity of the sides of the triangle was. Kant calls duration, continuity (which implies connectedness) and simultaneity the three "modes of time." Without them there is "mere succession." How can we move from a perception of mere succession to an appreciation of time in the fuller sense? We can see that Kant's problem is not the problem of time perception as it was posed, for instance, by Husserl, who was concerned to discover how we can even have a sense of succession, of a local proto time, which for Kant did not appear problematic. Kant's problem was how we get from there to time in the fuller sense.

Let us concentrate on the problem of connectedness. We want the perceiver to appreciate time as a series, when it is aware only of subseries that are as long as it can hold together. We must think of these series to be not just of sensations, but at least of such things as shapes, perhaps bodies, at any rate more fully formed objects than contemplated by Condillac. We must also suppose that it can remember having had such series, but that it is in no position to determine their temporal order with respect to each other. Here is what Kant thinks it can do.

Suppose it remembers that there was sunlight on a stone, and that the stone then warmed up. This is stored in the form of a "Judgment of Perception." If this happens repeatedly, a Humean expectation might be formed, so that the perception of the sunlight brings about a belief that the stone will warm up. Kant certainly believed this, but in itself it is not enough to perform the desired construction. What we need, in addition, is an assumption of necessity, which means no more and no less than this, that perceived regular successions also hold when they are not perceived as successions. Suppose I perceive sunlight on a stone (at the end of some series

of perceptions *A*), then turn away and perceive some other series *B*, then return to the stone and perceive it to be warm. Let this be the beginning of a third series *C*, after the end of which I engage in yet a fourth *D*. At this point I look back and recall having had *A*, *B*, *C*, and *D*. Perhaps I still recall that *D* was last, but what order was there between *A*, *B*, *C*? The solution is obtained by drawing on what I know about regular succession. I know that warmth follows sunlight, and thus place *C* after *A*. Kant's theory of time is a causal theory of time in the sense that the connectedness of time is established by drawing on knowledge of regular succession. I take this to be the argument of the *Second Analogy*. Through repeated application of this procedure, with a growing repertoire of perceptions of succession, we are, Kant thinks, able to form a series. This is almost a truism, since items that do not fit can be discarded as illusory. Further, the effect of this construction is that causal connections are *necessary*, in the sense of holding at all times. This does not mean, of course, that attributions of necessity are infallible, but it does mean that any regularity used to join two subseries is at that time judged to be necessary (though a retraction may no doubt sometimes be in order). It also follows, after a suitable definition of what an event is (an irreversible subjective series) that every event has a cause. Because to be a real event, it must occur in real time, and real time is constructed through causality. The contents of real time are called objects of experience, and, in their totality, simply "experience." Judgments about them are judgments of experience: the judgment "There was sunlight on the stone and then it became warm," when deemed usable to establish connectedness can be promoted from a judgment of perception to a judgment of experience, and can then be rendered as "The sunlight caused the stone to become warm," which is really only a change in modality, not the assertion of additional content.

There are then various arguments which mean to determine just what processes must occur so that the cognitive process can advance, given that it has started from such and such an initial state. We find here appeals to association and custom together with the world's most famous transcendental argument, and appeals to various principles. The list could be much extended since all manner of conjectures and arguments were offered concerning the workings of the black box that processes sensations, generating perception. Reid and the Common Sense school alone showed a remarkable reluctance to speculate about these matters, taking them to be largely impenetrable. But even Reid allowed himself to attribute certain operations to custom, others to nature.

It is most important not to consider these arguments merely in isolation, but in the context and service of the kind of epistemological thought experiment that Carnap later called the "epistemically adequate constructional system."

IV

Since these epistemological thought experiments focus so much upon the

processes through which we come to know or perceive certain things, objects of various kinds, they have lead their designers to peculiar views on the ontological status of what is perceived or cognized.

Let us return to Berkeley's account of distance:

I believe whoever will look narrowly into his own thoughts, and examine what he means by saying, he sees this or that thing at a distance, will agree with me, that what he sees only suggests to his understanding, that after having passed a certain distance, to be measured by the motion of his body, which is perceivable by touch, he shall come to perceive such and such tangible ideas which have been usually connected with such and such visible ideas. (*New Theory of Vision*, sect. 45.)

Previously we have considered only the process by which the perceiver comes to appreciate distance. Here now we are told that *to say x is at such and such a distance means a certain thing*. We are, in other words, given a contextual definition of "x is at distance y."

Consider now Kant: An object, he says, "is that in the concept of which the manifold of a given intuition is united" (CPR B 137). Phrasings of this sort occur several times. They are best understood as also contextual definitions, or schemata thereto. This needs a little explanation. Let me recall the last step in the threefold synthesis, in the version of Condillac. It was an interpretation placed on a series of sensations in the form of the judgment: "This is a line." In the absence of such a judgment, no line could be seen or enter consciousness in any form. Now "line" is the concept in which the sensations, the manifold of that particular intuition, are united. Generally, Kant seems to say that x is perceived, thought of, or that reference to x is made, if the concept "x" is used to unite a certain manifold. (In Condillac's example, the concept was "triangle.") There is a temptation to conclude that "x" could not have been obtained by abstraction, since the mental capacity which we call the concept "x" is here a prerequisite of reference to x. But that would be rash. It would lead to the unwelcome conclusion that such empirical concepts as "apple," must be present before any apple is ever seen.

This is not how Kant pictures it. It is his view, rather, that there must be *some* concepts, perhaps "thing" that must precede the perception of any object whatever. But once several things have been constructed, other concepts can be abstracted, and become available as concepts for future synthesis.

V

We can now see how various philosophical labels have come to be attached to the philosophical systems, for instance, of Berkeley and of Kant. They have been described as verificationist, reductivist, phenomenalist, idealist, constructivist, and probably some others. All of these seem to me to be reasonable ascriptions. To such questions as "What is distance" Berkeley should make the response roughly: "Let's reformulate this a little, if

you don't mind; instead of answering *that* question, I will tell you how one finds out that a thing is at distance n , for any n , and that really gives the meaning of 'is at a distance.'" Because of his thought experiment, Berkeley knows how to answer that question. This attests to the verificationist strain: in response to the question what an x is we are told how one finds out that a thing x 'es. It is just the same in Kant.

It is in this way that ontological lessons arise from the thought experiments. But these lessons can be of two quite distinct sorts. The perceiver (and that means not only the subject of the thought experiment, but the designer of it as well, since he pictures himself as in the same situation) can go the definitional route, as above, and say that the objects, the furniture, of the world, just *are* the products of his synthesis—what these are can be expressed more or less cautiously. If this line is taken, then the question for the origin and cause of the sensations that started the cognitive process is played down. One strategy, in particular, is to take a strong verificationist line in which that question makes no sense—causation and related concepts being defined only for the experienced context.

The other line is to reject this currency devaluation. The very same thought experiments that generate transcendental idealism also support scepticism. For if we discover, through minute scrutiny, how the mind comes to believe, for example, that there are unperceived things, we have discovered how a certain input would bring the perceiver to that conclusion (that is, just how his mind works), whether there really were unperceived things or not. (In the same situation the idealist would claim that "unperceived thing" has no meaning outside the construction). Consequently, the sceptic argues, we still don't know whether there really are such things. I take this to be Hume's line.

Both, philosophers like Kant and philosophers like Hume, need epistemological thought experiments, and both of them worked on them. It has sometimes been argued that when they were engaged in this enterprise they were really doing psychology, and not philosophy. I won't respond to this, and, in any case, it does not bother me. But it is difficult for me to conceive of these systems without the thought experiments. I am inclined to think that not much would be left of them.

It thus seems that some of the grander conclusions of the philosophers we have considered, their idealism, or scepticism, are comments or glosses on, or lessons taken from, constructions that we may reasonably describe as thought experiments. But I want to make a further claim, namely, that thought experimentation, or constructivism, forms a discernible tradition which extends at least from Malebranche, Berkeley, Condillac, Kant and many intermediaries to Carnap, Goodman and perhaps further to cognitive science and some AI research. To be sure, traditions of this sort (I mean to include such "movements" as Empiricism, Rationalism and the like) are historians' artifacts; but they are not, for that reason, unfounded fiction. I

believe, in particular, that much can be learned about the history of modern epistemology if we take a large part of it to have formed the tradition of which I have given here the merest sketch.

BIBLIOGRAPHY

- George Berkeley, *Essay towards A New Theory of Vision* (London, 1709; many subsequent editions).
- Edmund Burke, *A Philosophical Enquiry into the Origin of our Ideas of the Sublime and the Beautiful* (London, 1759; repr. N.Y. 1971).
- Rudolf Carnap, *The Logical Structure of the World* (Berkeley: University of California Press, 1969).
- N. Malebranche, *Recherche de la vérité* (Paris, 1674; New edition, Paris, 1962).
- Thomas Reid, *An Inquiry into the Human Mind* (Edinburgh, 1764).
- Thomas Reid, *On the Intellectual Powers of Man*, 1785, ed. Sir William Hamilton (Edinburgh, 1880).