THOUGHT EXPERIMENTS IN EINSTEIN'S WORK

John Norton

I. INTRODUCTION: CHARACTERIZATION OF THOUGHT EXPERIMENTS

MY strategy in this paper will be to give a general characterization of thought experiments as they appear in modern physics, to draw some conclusions about them from this characterization and then to provide illustrations in a range of examples of thought experiments from Einstein's work.

Thought experiments are arguments which:

(i) posit hypothetical or counterfactual states of affairs, and
(ii) invoke particulars irrelevant to the generality of the conclusion.

Let me now explain the various components of this characterization: ...*arguments*...: Thought experiments in physics provide or purport to provide us information about the physical world. Since they are *thought* experiments rather than *physical* experiments, this information does not come from the reporting of new empirical data. Thus there is only one non-controversial source from which this information can come: it is elicited from information we already have by an identifiable argument, although that argument might not be laid out in detail in the statement of the thought experiment. The alternative to this view is to suppose that thought experiments provide some new and even mysterious route to knowledge of the physical world. Thus Brown (1986, pp. 12-13) argues that thought experiments are a special window through which we can grasp the universals of an Armstrong-like account of physical laws. I can see no benefit in adopting a mysterious window view of thought experiments, when all the thought experiments I shall deal with (and have seen elsewhere) in modern physics can be readily reconstructed as arguments.[1]

A very broad range of argument forms should be allowed here; in particular they should include inductive argument forms. In the thought experiments of modern physics, the premisses of the arguments are generally held within one or other physical theory. Thus a thought experiments might be a purely deductive derivation within some physical theory. But not all such thought experiments are. As we shall see later, some do involve inductive reasoning or the introduction as a premiss of a general philosophical principle not contained within any specific physical theory.

...hypothetical or counterfactual states of affairs...: This condition gives thought experiments their thought-like character. For if they did not posit such states of affairs they would not be thought experiments; they would be the description of a real experiment or state of affairs. Thus we cannot base a thought experiment on the supposition that projectiles on the earth's surface have roughly parabolic trajectories; for they really do have roughly parabolic trajectories and, so far, we have only given an accurate description of the way things really are. But we could begin a thought experiment by supposing *counterfactually* that projectiles followed trajectories which are the arcs of circles. Or we could incorporate parabolic trajectories in a thought experiment if in addition we suppose *counterfactually* that lightweight projectiles are not impeded by air resistance or if we consider a *hypothetical* projectile of some specified size, shape and composition.

...particulars irrelevant to the generality of the conclusion...: The presence of these particulars is what makes thought experiments experimentlike. Thus, in one version of the thought experiment in which Einstein sought to demonstrate that the effects of acceleration mimic those of gravitation, he asked us to imagine a physicist-observer who has been drugged and reawakens closed up inside a box (Einstein, 1912, pp. 1254-55). That there is an observer, that the observer is a physicist, that the physicist has been drugged, that he is enclosed within a box—all these are particulars which are irrelevant to the generality of the conclusion which Einstein seeks to draw. Without particulars such as these, however, thought experiments would not have their experimental appearance.

The above characterization provides necessary conditions for something to be a thought experiment. It must be an argument, if its conclusion is to be credible; it must (i) posit hypothetical or counterfactual states of affairs to warrant the "thought" label; and it must (ii) involve particulars irrelevant to the generality of the conclusion to warrant the "experiment" label. But this characterization does not provide sufficient conditions for thought experiments. For example, any argument which is not a thought experiment could be made to satisfy both conditions (i) and (ii) by appending to its list of premisses, sleeping premisses of the appropriate form which play no further role at all in the argument. The resulting modified argument clearly need not be a thought experiment. Further, to recover sufficient conditions for a thought experiment from the characterization, the nature of the particulars in (ii) would have to be specified more closely. They must be of a type sufficient to guarantee the appropriate experimental character to the argument.

Fortunately, for my purposes, there is no need to pursue the question of sufficient conditions. The original characterization already yields a number of interesting conclusions on which the remainder of the paper will focus.

(1) *Reconstruction of thought experiments.* Since it is not always obvious from their presentation that thought experiments are arguments, the analysis and appraisal of a thought experiment will involve reconstructing it explicitly as an argument. In particular, thought experiments are to be appraised by just those criteria used to judge the goodness of an argument. A good thought experiment is a good argument; a bad thought experiment is a bad argument.[2]

(2) *The elimination thesis.* Thought experiments are arguments which contain particulars irrelevant to the generality of the conclusion. Thus any conclusion reached by a good thought experiment will also be demonstrable by an argument which does not contain these particulars and therefore is not a thought experiment. Whilst thought experiments may be eliminable in principle, it does not follow that this elimination will be easy. In fact one can almost guarantee that it will not. For thought experiments are usually introduced when the straight argument would be difficult to develop. For example a direct derivation of the desired result within a theory may be possible but extremely cumbersome. (For an example of this, see section 2.) Or the vehicle of the thought experiment might make easier the introduction of certain philosophical principles or facilitate certain inductive moves. (For examples of both, see sections 3 and 4.)

(3) *Classification of thought experiments.* The conclusion of a thought experiment must be free of the particulars involved in condition (ii). Thought experiments can be classified according to the means through which these particulars are removed.

Type 1: The particular-free conclusion follows deductively from the premisses. Thought experiments of this type are typically *reductio* arguments. The particulars might be involved in a counter-example to a universally quantified assertion through which contradiction the conclusion follows. (For an example, see section 2.)

Type 2: The conclusion is freed of the particulars by an inductive step. This step might involve the assertion that the case involving the particulars is "typical" or that the particulars are "inessential," so that the result derived holds in other cases as well. (For example, see sections 3 and 4.)

II. THERMODYNAMIC THOUGHT EXPERIMENTS

Thermodynamics lends itself to some of the most effective of all thought experiments. This is because the three laws of thermodynamics are such that they can readily be stated as assertions of impossibilities.

First law: It is impossible to design a perpetual motion machine of the first kind, that is, a machine whose sole effect is to produce more energy than it consumes.

Second law: It is impossible to design a perpetual motion machine of the second kind, that is, a machine whose sole effect is to transfer heat from a colder heat reservoir to a hotter reservoir.

Third law: It is impossible for any finite process to yield a temperature of absolute zero.

An easy way to derive consequences from these assertions of impossibility is by a *reductio* argument. To derive a theorem X, take one or more of the above laws as premisses. Assume –X, the negation of X. Show that –X allows the design of a machine prohibited by the premisses. Then conclude X.

These *reductio* arguments almost automatically become thought experiments. –X is assumed hypothetically, or even counterfactually if the outcome of the derivation has already been foreshadowed. The designing of the machine requires the positing of further hypothetical states of affairs, as well as particulars which give the argument its experimental character. There is generally a very strong incentive to include many such particulars to simplify the machine's design so that the contradiction with the laws of thermodynamics is easy to see. Thus in a Carnot engine, it is convenient to use an ideal gas as the working fluid, rather than one of unspecified properties; and one assumes that the piston moves frictionlessly, rather than with some unspecified frictional relation. Moreover, since the argument is a *reductio* argument, one has nothing to lose by including as many such particulars as is convenient. For once the contradiction is established, they drop from sight and do not appear in the conclusion X. Thus these thought experiments are of type 1, the conclusion being freed of the particulars by deductive argumentation.

My example of a thermodynamic thought experiment comes from Einstein's 1909 work on the quantum theory of black body radiation. The thought experiment is the most complicated of all those I shall consider, but I urge the reader to persevere since it is one of the most rewarding of its kind.

At that time, three distinct theories were involved in the analysis of black body radiation: classical electrodynamics, thermodynamics and statistical mechanics. The best known application of statistical mechanics was Boltzmann's kinetic theory of gases, in which the macroscopic behavior of bodies of gases was analysed in terms of the most likely behavior of a large collection of rapidly moving molecules. In particular, the energy and pressure of a volume of gas in thermal equilibrium with its container would both be fluctuating about their mean values, although these fluctuations would not usually be macroscopically observable. Black body radiation is the heat radiation emitted by a perfectly black body. At the turn of the century, this radiation was treated as a wave phenomenon within classical electrodynamics. In 1905 Einstein argued that black body radiation sometimes behaved thermodynamically as though it were not constituted of electromagnetic waves, but as though it consisted of mutually independent quanta of energy. By 1909 Einstein was prepared to argue that we could not choose whether we were prepared to treat black body radiation as a wave phenomenon or as a particulate phenomenon. We needed *both* points of view.[3]

One of Einstein's arguments for this wave-particle duality of black body radiation was a thought experiment (Einstein 1909a, pp. 189-90 and 1909b, p. 823). He imagined a cavity containing an ideal gas, black body radiation and a mirror which is free to move only perpendicular to its surface. The entire system is in thermal equilibrium. The gas exerts a fluctuating pressure on the mirror. Similarly the radiation exerts a pressure on the mirror. From the beginning Einstein had treated black body radiation by statistical mechanical methods. So he assumed that this radiation pressure would be fluctuating as well. As a result the mirror executes a jiggling motion, which is just a form of Brownian motion; therefore the mirror's momentum will fluctuate about its mean value. Given the empirically known distribution of energy over the spectrum of black body radiation, Einstein could calculate the size of these fluctuations and in particular that part due to the radiation. The resulting expression was the sum of two terms. One was clearly due to the wave-like properties of radiation; the other due to its particulate properties. Thus adopting either a wave or a particulate view would enable one to recover just half of the correct expression. One had to adopt the view that radiation had both wave and particle aspects simultaneously.

I will not try to recount Einstein's calculations, since they are quite involved. Rather I will analyse just the first part of the thought experiment, since remarkably it can be analysed entirely qualitatively without loss of rigor. Further, the first part of the thought experiment is a clear example of a thought experiment of type 1 using a *reductio* argument, whereas the complete thought experiment is not. Einstein assumed that black body radiation can be treated by the methods of statistical mechanics and that its radiation pressure does fluctuate like that of a gas. It is by no means obvious that we are forced to this assumption, especially given that then there was no satisfactory account of the constitution of black body radiation. Indeed we might well want to deny it in an attempt to avoid the torments of wave-particle duality. What Einstein establishes is that if one treats gases kinetically, then one is forced to conclude that there are fluctuations in black body radiation pressure. Otherwise one would have an uncontrolled transfer of heat energy from the gas to the radiation in violation of the second law of thermodynamics. It follows that black body radiation must be treated by statistical mechanical methods. I reconstruct the thought experiment as an argument. (See Figure 1 appended at the end of this essay.)

To be shown, X: There are random fluctuations in black body radiation pressure.

(1) Premisses: (a) second law of thermodynamics; (b) kinetic theory of gases; (c) the phenomenon of radiative damping, in which an isotropic radiation field applies a damping force to slow the motion of any totally mirrored object moving in it. (This effect results from a greater amount of radiation being reflected by the leading face of the object than the trailing face.)[4]

Proof by *reductio*. (2) Assume –X.

(3) Consider the above cavity with mirror, ideal gas and black body radiation.

(4) From (1c) and the fact the black body radiation is isotropic, it follows that there will be a radiative damping force on any motion of the mirror, which will slow that motion and transfer the kinetic energy of the mirror to the radiation.

(5) From the equipartition theorem of (1b), it follows that the mirror will have a mean kinetic energy of $kT/2$ due to its thermal equilibrium with the gas. (k is Boltzmann's constant and T the temperature.) Any reduction in this equipartition energy of the mirror will lead to a transfer of heat energy from the gas to the mirror to restore the equilibrium.[5]

(6) From (4) and (5) it follows that there will be an uncontrolled transfer of heat energy from the gas to the kinetic energy of the mirror and then to the radiation. This is a spontaneous transfer of heat from a colder region (the gas) to a hotter region (the radiation), which violates the second law of thermodynamics.

(7) Conclude: the assumption –*X* is false. *X* is true.

Note finally that the existence of fluctuations in the radiation pressure opens the necessary route for the return of energy from the radiation to the gas so that a thermal equilibrium can be established.

I find it striking if not startling that Einstein could devise a thought experiment which could force his conclusion so rapidly and easily. It is important to see just how general Einstein's conclusion was. We need to know essentially nothing about the constitution of black body radiation and the conclusion is still forced upon us. Whatever black body radiation is, it need only provide the effect of radiation damping and the argument goes through. At a time when the constitution of radiation was proving to be a very great puzzle, such an argument is especially appropriate.

Einstein's conclusion follows from his premisses. Thus we could find another argument which is not a thought experiment but which still takes us from the premisses to the conclusion (elimination thesis). The argument may well even be a *reductio* argument, but not one of an experimental character. But I think it is clear that such an alternative argument would be very difficult to find because of the great complexity of the case. The advantage of the thought experiment is clear. It makes the analysis much easier by enabling us to deal with a very simple situation in which the interaction of the various effects can be followed fairly easily. Correspondingly we can see the strength of thought experiments of type 1, for the generality of the conclusion is in no way impaired by the fact that the thought experiment deals only with a quite specific case.

III. SPECIAL AND GENERAL RELATIVITY

A common feature of Einstein's thought experiments in the realm of special and general relativity is that they tend to exploit a particular philosophical viewpoint concerning the relationship between theory and observation. The viewpoint has its origins in the positivist rejection of metaphysics and the primacy of observation over theoretical constructs. My statement of it is:

> *Verifiability heuristic for theory construction (version 1)*: A theory should not use theoretical terms which have no observational support.

The difficulty with this statement of the heuristic is that it is not obvious just what observational support is. How are we to identify that a theory does use theoretical terms without observational support? How are we to find which terms are at fault? In the third of the cases examined below, Einstein answers these questions in a way that is unacceptably strict and the thought experiment suffers for it. In the other two cases, Einstein found an ingenious way of identifying theories which violate the heuristic and the terms involved, without the need to give an account of observational support. In some cases, he found that such a theory could be induced to violate a particular verifiability principle; that is, it would insist that certain states of affairs are distinct, even though no possible observation could verify the difference. This narrower version of the heuristic is:

> *Verifiability heuristic for theory construction (version 2)*: States of affairs which are not observationally distinct should not be distinguished by the theory.

The observationally unsupported terms which are to be eliminated will be just those that lead the theory to insist that the observationally indistinguishable states of affairs are distinct.

Einstein tended not to state explicitly these heuristics in the thought experiments. In fact this was one of things that made the thought experiments so convincing. They automatically focus attention on observables, highlighting their importance, and thus make the implicit introduction of the heuristics possible. When Einstein has to introduce these heuristics explicitly, his arguments lose some of their impact. On the other hand, in reconstructing the thought experiments as arguments, we are forced to expose the heuristics, thus allowing a better informed appraisal of the thought experiment. Similarly the first two thought experiments to be reviewed in this section are of type 2 and contain one or more inductive steps. Once again in reconstructing the thought experiments as an argument, we must introduce these steps explicitly and thus can appraise the thought experiment more accurately.

3.1 The Magnet and the Conductor

Einstein's best known thought experiment appears in the opening paragraph of his celebrated 1905 special relativity paper (Einstein, 1905, p. 37). The thought experiment deals with the phenomenon of electromagnetic induction within the very successful classical electrodynamics of Maxwell and Lorentz. That theory posited a luminiferous aether which determined an absolute state of rest and it seemed impossible for the theory to do without it. For example the theory entailed that the speed of light *in vacuo* was a constant, 186,000 miles per second, and thus was independent of the speed of the emitter. How could this result be intelligible without an absolute state of rest against which to measure this speed?

What Einstein sought to establish with the thought experiment was that electrodynamics *should* seek to do without this absolute state of rest. He considered a magnet and conductor in relative motion. See Figure 2 (appended at the end of this essay). In the first case, the magnet is at (absolute) rest and the conductor moving. The motion of the conductor through the magnetic field produces an electromotive force in the conductor which generates a measurable current. The second case involves an identical magnet and conductor which is also identical in set-up as to the relative positions and the relative velocities of both. But in this case it is the conductor which is at (absolute) rest and the magnet moving. The classical theory's account of this second case is quite different to the first. The motion of the magnet leads to a changing magnetic field which in turn induces an electric field. This electric field produces a current in the conductor of the same size as the first case. The presence of the induced electric field "with a certain definite energy," to quote Einstein, clearly distinguishes the second case from the first according to the theory. However as far as the observables are concerned—that is, the measurable current in the conductor—the two cases are indistinguishable. Thus the observables are sensitive only to relative velocities, whereas the theory is sensitive to absolute velocities as well. Einstein concluded:

> Examples of this sort, together with the unsuccessful attempts to discover any motion of the earth relatively to the "light medium," suggest that the phenomena of electrodynamics as well as of mechanics possess no properties corresponding to the idea of absolute rest.

It is now easy to reconstruct the thought experiment as an argument. In abbreviated form it is:

(1) In the case of electromagnetic induction, through its positing of an absolute state of rest, classical electrodynamics distinguishes states of affairs which are not observationally distinct.

Therefore in this case, classical electrodynamics violates

(2) Verifiability heuristic for theory construction (version 2): therefore

(3) Absolute velocities should be eliminated from the theoretical account of electromagnetic induction.

(4) Inductive step: This example is typical since (a) there are other examples of this type and (b) there is a history of unsuccessful attempts to detect this state of rest by optical experiments. Therefore

(5) Absolute velocities should be eliminated from electrodynamics.

3.2 The Einstein Elevator

Special relativity had re-established the relativity of *uniform* motion. According to it all inertial frames of reference were intrinsically indistinguishable and we were free to regard any as "at rest." Einstein next sought to extend this relativity of motion to accelerated motion. What made this task difficult was the fact that accelerating frames of reference were readily distinguishable observationally from inertial frames of reference by the presence of inertial forces. These forces, for example, had been used by Newton to distinguish absolute rotation from relative rotation in his bucket thought experiment.

What struck Einstein in 1907 was the fact that the inertial field of a uniformly accelerating frame of reference influenced the motion of test bodies in exactly the same way as a homogeneous gravitational field. He postulated (i) the exact physical equivalence of the uniformly accelerating frame with inertial field and the corresponding unaccelerated frame with a homogeneous gravitational field, so that (ii) one could now regard the accelerated frame as unaccelerated. This entire postulate is the principle of equivalence. Here I shall be concerned with Einstein's argument for (i). Whether Einstein succeeded in establishing (ii) remains controversial.[6] (i) leads to some surprising results. In special relativity in a uniformly accelerating frame, resting clocks run at different speeds and light follows a curved trajectory. Thus it follows from (i) that gravitational fields must have the same effects on resting clocks and light propagation.

Einstein justified (i) via the following thought experiment (Einstein, 1917, pp. 66-70; Einstein and Infeld, 1938, pp. 230-35). He imagined an opaque chest in a region of space remote from gravitational source masses. There is a rope attached to the chest and some "being" pulls on the chest so that it accelerates uniformly. An observer inside the chest will see that all free bodies fall with equal acceleration in the chest; that is their motion is indistinguishable from the case in which the chest is at rest in a homogeneous gravitational field. Einstein then posits the complete equivalence of the two cases (which I understand to mean simply that they are the same case). Reconstructed more explicitly as an argument, we have:

(1) In an opaque chest, an observer will see free bodies move identically in case the box is uniformly accelerated in gravitation free space and in case the box is at rest in a homogeneous gravitational field.

(2) Inductive step: (a) the case is typical and will hold for all observable phenomena and (b) the presence of the chest and observer are inessential to the equivalence. Therefore

(3) A uniformly accelerating frame in gravitation free space and a frame at rest in a homogeneous gravitational field are observationally identical, but theoretically distinguished, which contradicts

(4) Verifiability heuristic for theory construction (version 2). Therefore

(5) A uniformly accelerating frame in gravitation free space and a frame at rest in a homogeneous gravitational field are the same thing (which becomes a postulate of a new theory).

There are several points to be made. First, the inductive step (2) is actually quite problematic, but this is effectively masked by the thought experiment format. The extension from the motion of bodies in free fall to arbitrary processes is quite a leap, especially in view of the bizarre consequences that follow. Even allowing that the opaqueness of the chest is inessential is challengeable. One way in which Einstein's contemporaries sought to refute his argument was to insist that gravitational fields always have sources, but inertial fields never do. Thus the observational equivalence of the two cases could depend essentially on the observer not being able to check whether the field in question has source masses or not.

Second, the thought experiment is usually known as the Einstein *elevator* experiment and involves the case of an elevator in free fall above the earth's surface, in which, it is argued, the gravitational field of the earth vanishes locally. Einstein most commonly employed a chest accelerating in a region of gravitation free space. He had a good reason for preferring the latter; as he frequently stressed, it is impossible to transform away completely a gravitational field such as the earth's through acceleration. Complete interchangeability arises only in special cases, such as uniform rectilinear acceleration and homogeneous gravitational fields. For further discussion see Norton (1985). Finally, in two of its best known presentations (Einstein, 1911, p. 99-100; 1916, pp. 113-14), Einstein did not present the above argument in thought experiment form. But the argument is clearly the same argument stripped of the experimental particulars. This illustrates the elimination thesis in a very concrete way.

3.3 The Two Fluid Bodies

One of Einstein's major motives in seeking his general theory of relativity was the belief that both classical mechanics and special relativity suffered from an "*epistemological defect.*" The two fluid bodies thought experiment was intended to reveal this defect and show the need for an extension of the principle of relativity to accelerated motion (Einstein, 1916, pp. 112-13).

Einstein imagined two fluid bodies of the same size and nature in a region of space remote from other bodies and also well removed from one another. The only observable motion is a constant relative rotation of each sphere with respect to the other about the imaginary axis which joins them. In spite of the symmetry of the set-up described so far, Einstein asks us to imagine that one of them is a sphere and the second an ellipsoid of revolution. What could cause this asymmetry? Both classical mechanics and special relativity allow the above set-up. Both explain the asymmetry by saying that the ellipsoidal body is rotating in an inertial frame and is thus deformed by centrifugal forces, whereas the spherical body is not. This, Einstein proclaims, is entirely unsatisfactory. Inertial frames (he calls them "spaces of Galileo") are "merely *factitious* causes, and not a thing that can be observed," for as he explained a few sentences earlier:

> No answer [to the question of the cause of the asymmetry] can be admitted as epistemologically satisfactory, unless the reason given is an *observable fact of experience*. The law of causality has not the significance of a statement as to the world of experience, except when *observable facts* ultimately appear as causes and effects. (Einstein's emphasis)

(Lest Einstein's position seem outrageous, I should point that it does seem to allow unobservable fields as *intermediate* causes which carry the effects of the ultimate causes, the observable source bodies of the fields.) Einstein continues his argument by saying that the source of the asymmetry must be sought outside the system in the only observables available, the other bodies of the universe. (This conclusion is now commonly called "Mach's principle.") Finally it follows that in an epistemologically acceptable theory, no frame is privileged *a priori*. Reconstructing Einstein's argument we have:

(1) Classical mechanics and special relativity posit an unobservable cause, an inertial frame of reference, to account for the asymmetry between the two bodies of the set-up. This violates

(2) Strict version of the verifiability heuristic for theory construction: No cause is "epistemologically satisfactory, unless [it] is an *observable fact of experience*." Therefore

(3) Classical mechanics and special relativity are epistemologically unsound.

(4) The only ultimate causes compatible with (2) are bodies, which are to explain the asymmetry in a new theory.

(5) Unlike classical mechanics and special relativity, the new theory must not set aside any frame of reference as privileged *a priori*.

There are several major points of difficulty in this argument, which are made more apparent by the reconstruction but which I do not wish to analyse here. (What justifies the extremely harsh (2)? Why are bodies the only candidate for (4)? How do we arrive at (5)?) Rather the interesting point is the success of the introduction of the philosophical principle in (2). When stated boldly, the principle is very extreme. It is far stricter than version 1 of the heuristic above, for it does not just require observational support for a cause, but that it be observable itself. Nevertheless Einstein's argument is much celebrated, as the expansive literature on Mach's principle shows. I can only imagine that (2) has been allowed to slip by, since its extremism is disguised by a thought experiment format, which forces the reader's attention to observables.

IV. THE EINSTEIN-BOHR DEBATE

There is an irony in Einstein's use of the verifiability heuristic in thought experiments. That heuristic was used very effectively *against* him by another master of the thought experiment, Niels Bohr, in their famous debate over quantum theory, initiated in the late 1920s. The story is told, for example, by Bohr in Bohr (1949).

The "new" quantum theory of that time, developed by Heisenberg, Schroedinger and others, represented a particle, such as an electron, as a wave. In general that wave will be distributed in space, so the theory cannot specify a definite position for the particle. Nevertheless the particle will sometimes behave as if it has a definite position. Consider the wave representing an electron approaching a scintillation screen. The wave will intersect the screen at many different locations. However the flash on the screen arising from the collision of the electron will be localized to a definite point. In general, even given a complete specification of the wave function, the quantum theory will be unable to say precisely which point that will be; the best it can do is to offer a probability distribution, which gives information on the likelihood of the flash arising at the different points of the screen. This uncertainty extends to other quantities. The theory groups quantities into complementary pairs, such as the pair position/momentum or the pair time/energy. There is an inverse relationship between the uncertainty in the specifications of the values of each member of these pairs. For example, if there is little uncertainty in the position of a particle, then there will be very great uncertainty in its momentum and *vice versa*. A precise statement of this result is known as the Heisenberg uncertainty principle.

Einstein took this indeterminancy to indicate that a wave function could not be a complete description of a single particle. Rather it represented an ensemble of particles and the probabilistic predictions derived from it about position, for example, deal with the likelihood of one or other particle within the ensemble having a given position. Bohr, on the other hand, insisted on the completeness of the description of the state of a single particle by its wave function and that this state of affairs signified a fundamental break with classical concepts.

Einstein and Bohr used a sequence of thought experiments to play out this difference of opinions and I wish to examine the early (that is, pre EPR) phase of their debate (Bohr, 1949, pp. 211-30). Through them Einstein sought to show that one could have precise knowledge of the magnitude of complementary quantities, in violation of the uncertainly principle. The most famous of these is Einstein's clock-in-a-box thought experiment, which purported to enable exact determination of both the energy of a particle and the time at which it had it. Here, however, I want to consider the simplest of the thought experiments, which deal with momentum/position determination, since they are already sufficient to make my point about the use of inductive steps and the verifiability heuristic in the debate.[7]

Einstein's view was:

> The wave function does not give a complete description of the state of a single particle.

Bohr denied this view and held:

> A particle can manifest either a definite position or a definite momentum, but not both.

Einstein sought to establish his view by designing imaginary experiments in which both the position and momentum of a particle are measured exactly. What Bohr showed was that, in each case, the experiment could only establish one or the other but not both. For example, consider vertical height above a laboratory bench for the position of the particle and the complementary vertical component of momentum. The height above the bench of a particle moving across the laboratory can be determined by passing it through a horizontal slit in a vertical plate bolted firmly to the bench. See Figure 3 (appended at the end of this essay). In passing through the slit, however, the direction of motion of the particle will be changed in an undetermined way. This is due to a diffraction effect associated with the wavelike properties of the particle. See Figure 4 (appended at the end of this essay). This deflection of the particle corresponds to an undetermined change in its vertical momentum, that is an exchange of momentum with the plate of undeterminable size. Thus if we measure the position of the particle, we do not know its momentum. If we wish to know the size of the change in momentum of the particle, we need to use a different experimental apparatus, such as the one illustrated in Figure 5 (appended at the end of this essay). The vertical plate is suspended by delicate springs. Any exchange of momentum between a particle passing through the slit and the plate will be measurable from the resulting change in momentum of the plate. However the very arrangement that enables determination of this momentum interchange now makes the slit moveable, so that precise determination of its position is no longer straightforward. Bohr points out (Bohr, 1949, p. 220) that any reading of the pointer against the scale to determine the plate's position (by, for example, shining a light beam on it) will lead to an uncontrolled exchange of momentum with the plate, compromising the accuracy of the momentum control.[8]

Reconstructed as arguments, these thought experiments become:

Einstein's argument:

(E1) There is an experiment that can measure the exact position of a particle and the momentum it has at that position. This entails

(E2) The wave function does not give a complete description of the state of a single particle.

Bohr's argument:

(B1) In the case of position and momentum determinations using a plate and slit as above, one can only establish one of the position and momentum of the particle, the other being uncontrollably altered and thus undeterminable.

(B2) Inductive step: (B1) is typical. All possible experiments can only yield one of the position and momentum of a particle, the other becoming undeterminable. This entails:

(B3) (E1) is false. Thus Einstein's argument does not establish (E2).

(B4) To insist that (E2) is true whilst (E1) is false violates the *Verifiability heuristic for theory construction (version 1)*, for one has no direct experimental access to whatever further properties the particle may have over those allowed by its wave function.

Reconstructing the thought experiments in explicit argument form clearly reveals the inductive step and the appeal to the verifiability heuristic for theory construction. Notice once again that the vehicle of the thought experiment makes implicit introduction of the verifiability heuristic quite natural, since it focuses our attention onto observables.

V. CONCLUSION

The main conclusion I wish to urge is that the thought experiments of modern physics are simply arguments and not some kind of mysterious new window onto the physical world. Thus we analyse and appraise thought experiments by reconstructing them explicitly as arguments and testing them against just those standards which we apply to arguments of other forms. It is hard to see why we should ever want to think otherwise of such thought experiments as the thermodynamic one analysed in section 2. Such thought experiments are clearly no more than derivations within a given theory. The examples of section 3 and 4 were more complex, however, for they involved inductive moves and philosophical principles not usually a part of any physical theory and its derivations. Indeed I claimed that the introduction of these inductive moves and philosophical principles were greatly facilitated by the thought experiment format of the argument. Nonetheless I cannot see any reason for according these moves and principles special status just because they appear within a thought experiment. If the inductive move is justified or the principle warranted then that should be apparent when the thought experiment is reconstructed as an argument. If the justification is not apparent on reconstruction, what grounds can we offer for retaining it?[9]

NOTES

1. Of course I do not rule out the possibility that thought experiments can allow us to gain access to universals as Brown argues. However I urge that they may do so only insofar as these universals can be accessed via argumentation. Thus a thought experiment which purports to access a universal should on my view still be analysed as an argument.

2. In principle, there could be a case in which a thought experiment could not be reconstructed *explicitly* as an argument, because the thought experiment invokes some acceptable, inductive move, to which we only assent because of the suggestiveness of the thought experiment format. The deficiency here lies in our lack of understanding of the relevant inductive move. Presumably if we understood it

better, the reconstruction to explicit argument form would be possible. I know of no examples of thought experiments of this type.

3. For a brief survey of Einstein's work here and for a treatment of the thought experiment outlined below, see Klein (1980).

4. A potential problem here is that the radiative damping effect was at that time derived from classical electrodynamics, a theory which is inconsistent with the quantum treatment of radiation. As far as the argument is concerned, the consistency problem can be solved by simply positing the effect independently or showing that the effect is derivable from a subtheory of classical electrodynamics which is not inconsistent with the quantum treatment of radiation. I adopt the latter approach for a similar case in the old quantum theory of black body radiation in Norton (1987). The same subtheory could be used in this case as well.

5. This application of the equipartition theorem is truly ingenious. The theorem is usually applied to the molecules of a gas in thermal equilibrium and states that each molecule has the mean energy $kT/2$ for each degree of freedom of its motion. What Einstein realized is that nothing prevents us from treating the mirror as a rather large molecule (with one degree of freedom) in the gas and that it must also be subject to the equipartition theorem.

6. See Norton (1985) for an analysis of Einstein's principle of equivalence.

7. I also simplify the analysis a little by dealing only with the extreme case in which one or other of momentum or position is determined exactly and the other quantity is then completely undetermined. In the event, Bohr dealt with the more general reciprocal relation between the sizes of the uncertainties in position and momentum. If the uncertainty in one was large, the uncertainty in the other would be small, according to Heisenberg's relation that (uncertainty in momentum)x(uncertainty in position) is of the order of Planck's constant h.

8. Precisely what Bohr intends with this remark is not clear to me. We have seen that the simultaneous determination of the position and momentum of the particle in turn depends on whether we can determine simultaneously the position and momentum of the plate. So I presume Bohr intends to point out with his remark that we face an analogous problem in the latter determination as in the former. This triggers a fatal infinite regress which can only be halted by begging the question, that is, by *assuming* that there is something whose momentum and position (or energy/time *etc.*) can both be known precisely in violation of the uncertainty principle.

9. I am grateful to John Earman, Al Janis, Jim Lennox, and Tim Maudlin for helpful discussion and comments.

REFERENCES

Bohr, N. (1949), "Discussion with Einstein on Epistemological Problems in Atomic Physics," pp. 201-41 in P. A. Schilpp (ed.) *Albert Einstein: Philosopher Scientist*, 2nd ed. (New York: Tudor, 1951).

Brown, J. R. (1986), "Thought Experiments since the Scientific Revolution," *Int. Studies in the Philosophy of Science*, vol. 1, pp. 1-15.

Einstein, A. (1905), "On the Electrodynamics of Moving Bodies," in *The Principle of Relativity* (Dover).

Einstein, A. (1909a), "Zum gegenwaertigen Stand des Strahlungsproblems," *Physikalische Zeitschrift 10*, pp. 185-93.

Einstein, A. (1909b), "Ueber die Entwicklung unserer Anschauungen ueber das Wesen und die Konstitution des Strahlung," *Physikalische Zeitschrift 10*, pp. 817-25.

Einstein, A. (1911), "On the Influence of Gravitation on the Propagation of Light," in *The Principle of Relativity* (Dover).

Einstein, A. (1912), "Zum gegenwaertigen Stande des Gravitationsproblems," *Physikalische Zeitschrift 14*, pp. 1249-66.

Einstein, A. (1916), "The Foundation of the General Theory of Relativity," in *The Principle of Relativity* (Dover).

Einstein, A. (1917), *Relativity: The Special and the General Theory* (London: Methuen, 1977).

Einstein, A. and Infeld, L. (1938), *The Evolution of Physics* (Cambridge: Cambridge Univ. Press).

Klein, M. (1980), "No Firm Foundation: Einstein and the Early Quantum Theory," pp. 161-85 in H. Woolf (ed.) *Some Strangeness of the Proportion: A Centennial Symposium to Celebrate the Achievements of Albert Einstein (Addison-Wesley).*

Norton, J. (1985), "What was Einstein's Principle of Equivalence," *Studies in History and Philosophy of Science vol. 16*, pp. 203-46.

Norton, J., "The Logical Inconsistency of the Old Quantum Theory of Black Body Radiation," *Philosophy of Science*, 54, pp. 327-50.

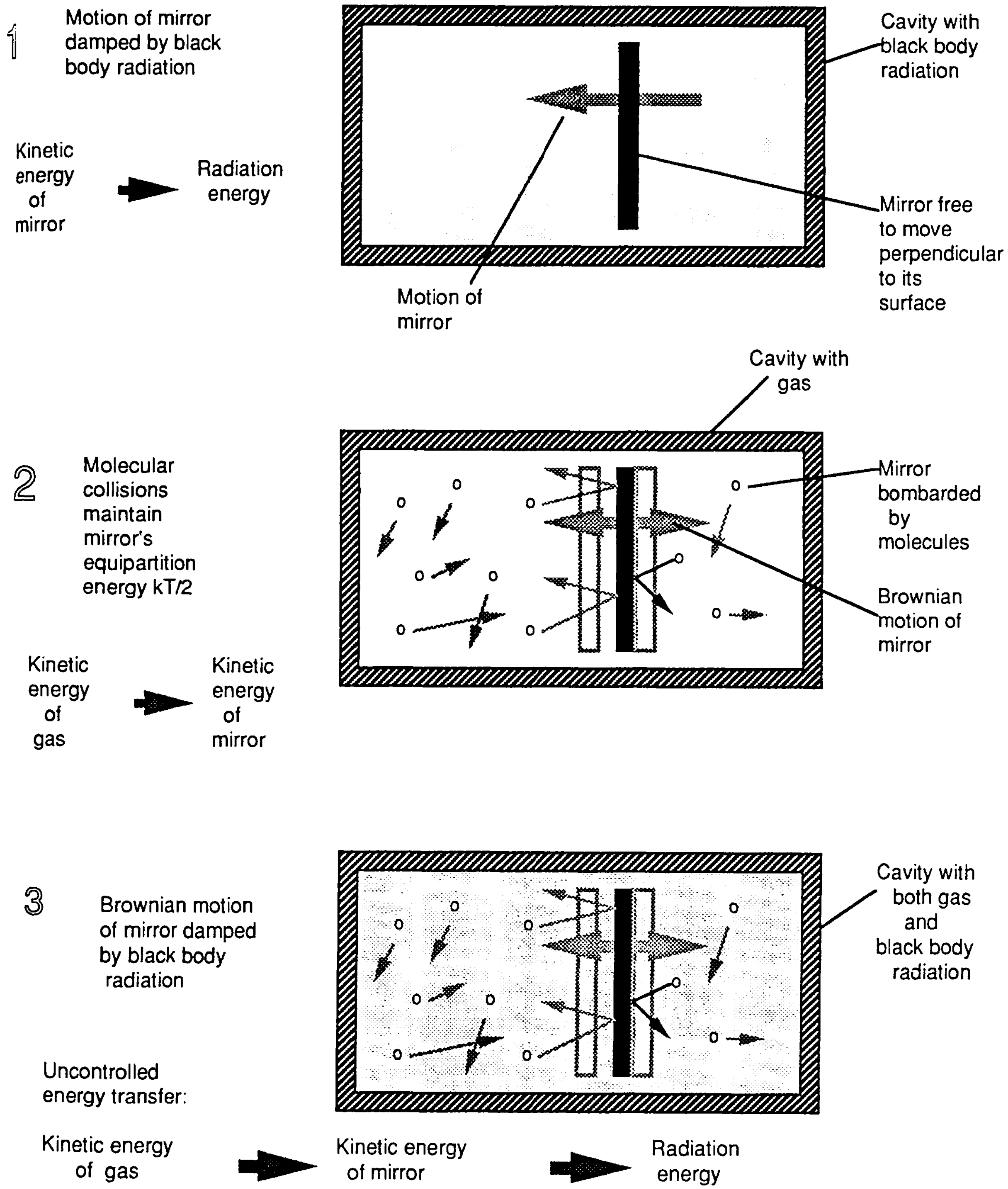
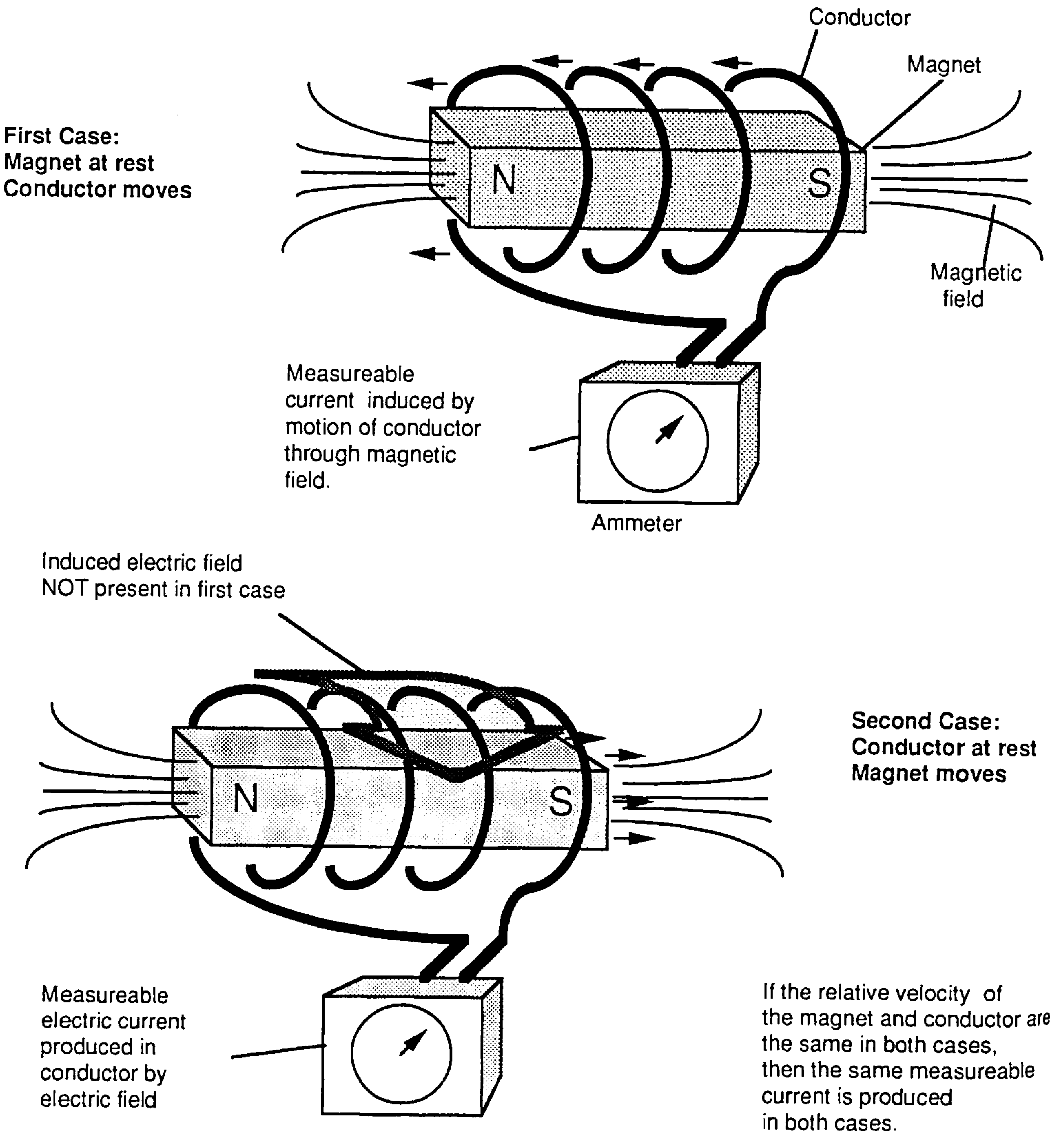


FIGURE 1: No Fluctuations in Black Body Radiation Pressure leads to Violation of Second Law of Thermodynamics



Measureable current induced by motion of conductor through magnetic field.

Induced electric field NOT present in first case

Measureable electric current produced in conductor by electric field

If the relative velocity of the magnet and conductor are the same in both cases, then the same measureable current is produced in both cases.

FIGURE 2: The Magnet and Conductor Thought Experiment

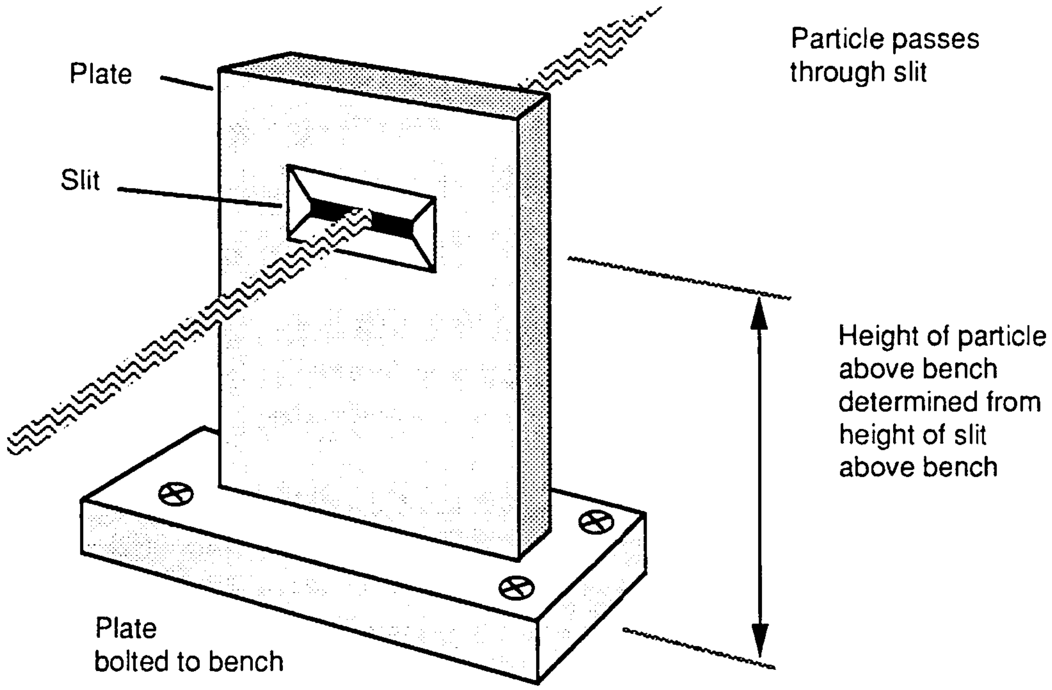


FIGURE 3: Apparatus for Determining Position of Particle.

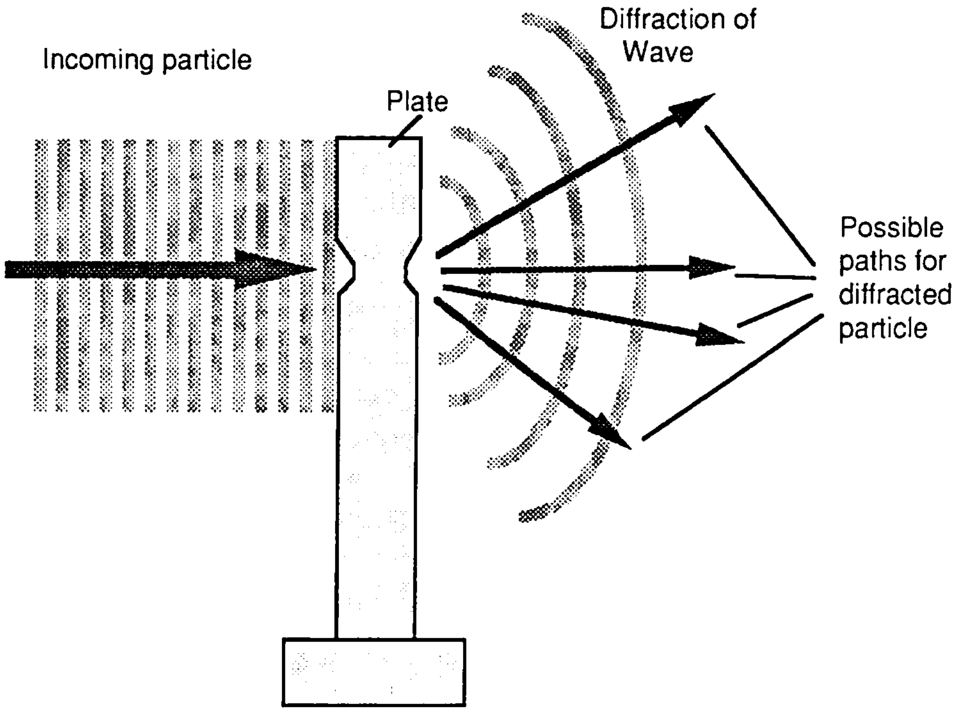


FIGURE 4: Diffraction of Particle by Slit

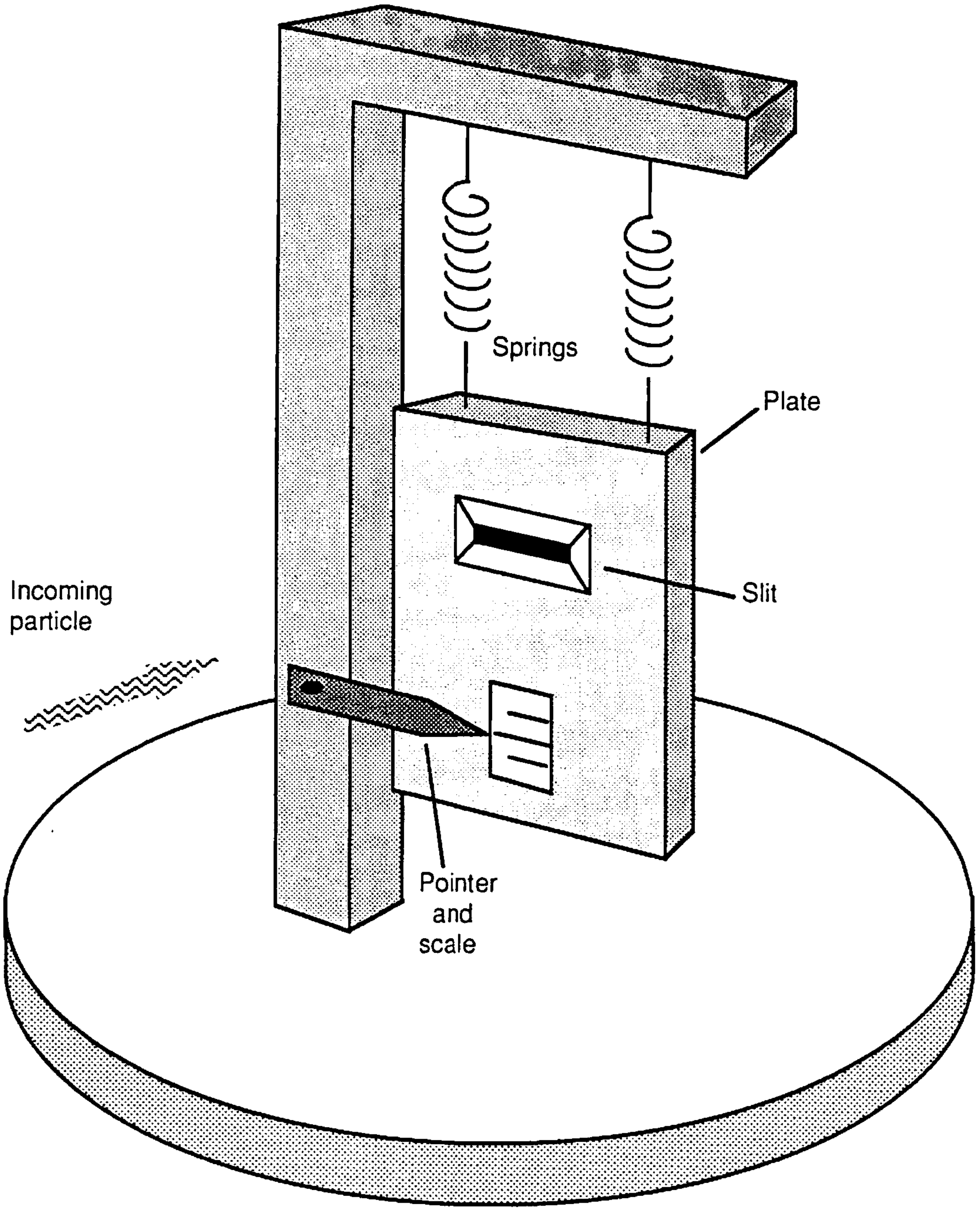


FIGURE 5: Apparatus for Controlling Momentum Exchange between
Particle and Plate with Slit