On Models and Theories
with applications to economics

Richard L. Epstein[1]

What is a model or theory? What constitutes a good model or theory?

These are questions at the heart of the nature and practice of science. Yet no satisfactory answers have been proposed that relate well to the actual work of scientists or that are generally accepted. Nowhere is this more apparent than in the study of economics, where these issues have typically arisen under the question of how realistic a model must be in order to be acceptable or useful.

In this paper I will present ideas to help resolve these questions. To motivate the solutions, I will first look at nine models and theories from both ordinary life and science that illustrate key points. Then I will discuss how experiments can be understood to confirm or disconfirm a theory, and how disconfirming experiments lead us to modify theories. I will then show how these answers can resolve the large debates in economics about how realistic a model should be and the relation of models in economics to prediction and practice.

Analogies

We can better understand the reasoning involved in using models by first looking at reasoning by analogy.

A comparison becomes *reasoning by analogy* when a claim is being argued for: On one side of the comparison we draw a conclusion, so on the other side we are justified in drawing a similar conclusion.

The difficulty in reasoning by analogy is to make clear what we mean by "a similar conclusion" and to justify that inference in terms of the comparison. The justification calls for some general claim under which the two sides of the comparison fall. Often analogies are so sketchy, with only the comparison offered, that their main value is to stimulate us to search for such a general claim. It must be one that relies on the similarities of the two sides of the comparison, and for which the differences between the sides of the comparison don't matter.

Analogies, then, involve abstraction from experience: a process of paying attention to some of our experience and ignoring other parts, the "differences."

---

[1] This paper was stimulated by a presentation on models in economics by Stephen Epstein at the 3rd Conference on Reasoning and Logic of the Advanced Reasoning Forum in Berkeley, July 2001, and later discussions with Peter Eggenberger, Troels Krøyer, Stephen Epstein, Carolyn Kernberger, Fred Kroon, and Peter Adams. A version of this paper was given at the 4th Conference on Reasoning and Logic of ARF in João Pessoa, Brazil, December 2002, in 2004 to the Economics Department of the University of New Mexico, and also in 2004 to the Biology Seminar of New Mexico Tech, where helpful suggestions concerning the material were made.

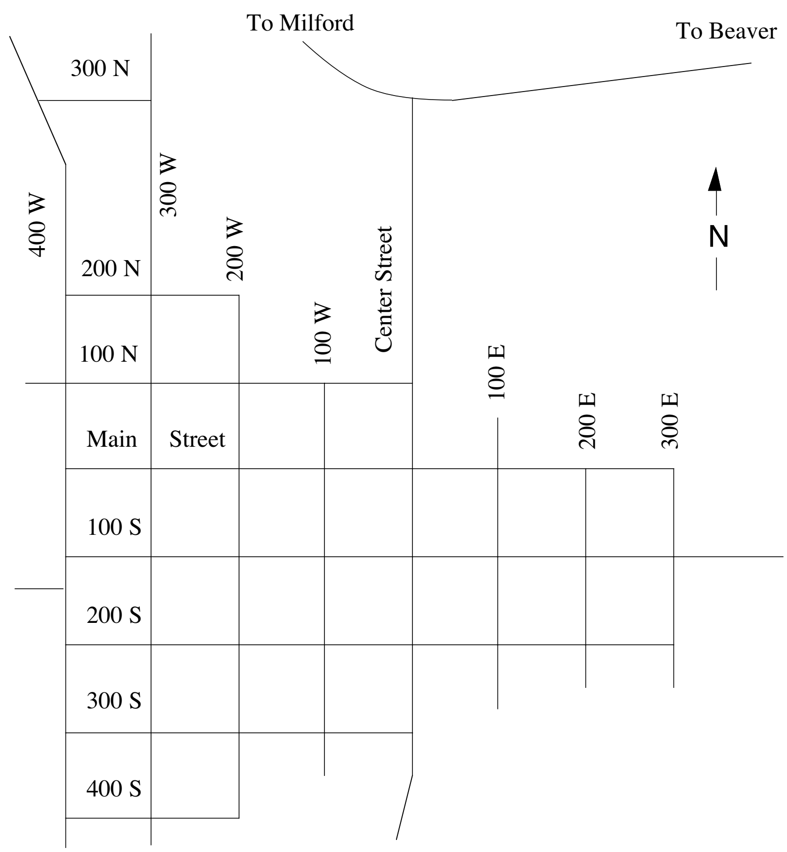
Bulletin of Advanced Reasoning and Knowledge 2 (2004) 79–100.

Epstein, 2002A, discusses analogies more and shows with a series of examples that the description above is how we use analogies in our daily lives. In the next section we will see that some models are clearly meant to be used for reasoning by analogy, while other models can be understood similarly: We ignore much and then can draw conclusions when the differences don't matter. This is the process of *reasoning by abstraction*.

Examples of models and theories

1. A map of Minersville

Here is an accurate map of Minersville, Utah:



We can see from this map that the streets are evenly spaced. For example, there is the same distance between 100 N and 200 N as between 100 E and 200 E. To the East, the last street is 300 E. There's no paved road going north beyond Main St. on 200 E.

That is, from this map we can deduce claims about Minersville, even if we've never been there. But there's much we can't deduce: Are there hills in Minersville? Are there many trees? How wide are the streets? How far apart are the streets? Where are houses located?

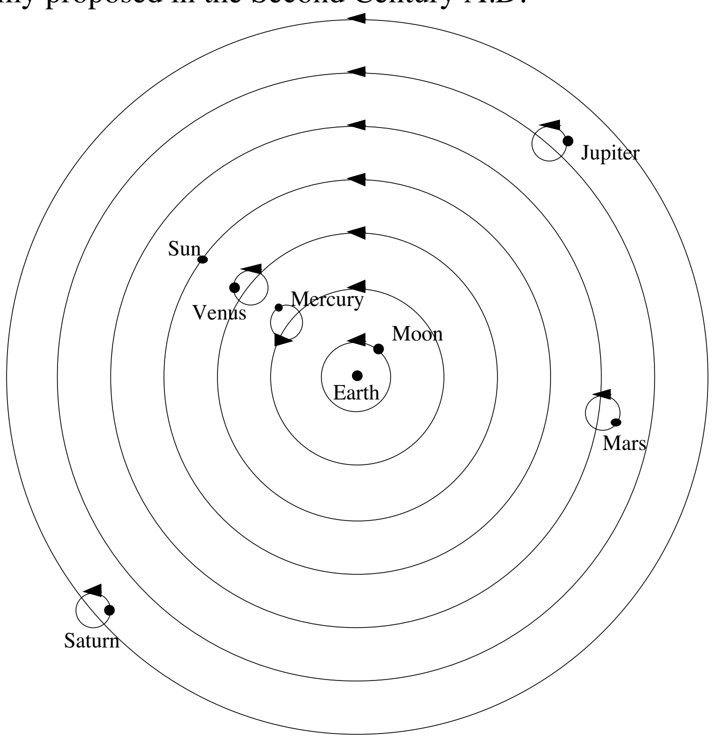
Reasoning about Minersville from this map is reasoning by analogy. The map is similar to Minersville in the relative position of streets and their orientation to North. The differences between the map and Minersville aren't important when we infer that the north end of 200 W is at 200 N. The map is accurate for what it pays attention to, but is not informative about what it ignores.

In contrast, a scale model of a city or a landscape abstracts less (or models more) from the actual terrain: height and often placements of rivers and trees are shown. The map of Minersville abstracts more from the actual terrain than a scale model of that town would, that is, it ignores more.

To use this model is to reason by analogy: We can draw conclusions when appropriate similarities are invoked and the differences don't matter. The general principle, in this example, is not stated explicitly. The discussion above suggests how we might formulate one, but it hardly seems worth the effort. We can "see" when someone has used a map well or badly.[2]

2. Models of the solar system

Here is a sketch of the model of the universe that the Egyptian astronomer Ptolemy proposed in the Second Century A.D.

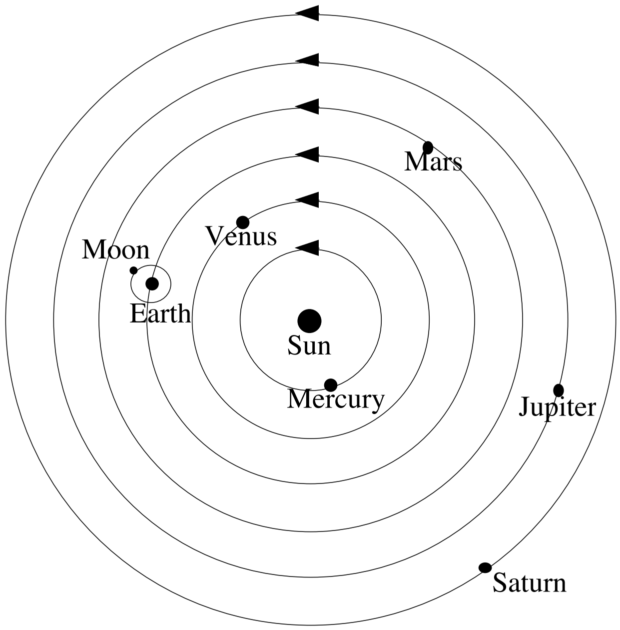


Ptolemy's model

This map of the universe is meant to show the relative positions of the planets, sun, and moon, and the ways they move. We can't deduce anything about, say, the size of the planets, the distances between them, nor the speeds at which they move, because this model ignores those. According to this model, each of the moon, sun, and planets revolves around the Earth in a circular orbit, all moving in the same direction. Along that orbit, each planet also revolves in a smaller circle, called an "epicycle." The sun, Earth, and Venus are always supposed to be in a line as shown in the picture. Ptolemy made a lot more claims about the planets, Earth, and sun that were to be used in making predictions, but for our purposes this sketch will do.

[2] Yet when I asked a friend from Colombia to read a map for me when we were driving, she was incapable, since she had no experience reading road maps.

Ptolemy's model accorded pretty well with observations of the movements of the planets and was the generally accepted way to understand the universe for many centuries. But in 1543 the Polish astronomer Copernicus published a book with a different model of the universe.



Copernicus' model

This sketch, too, abstracts a lot from what is being modeled. The sun is shown to be larger than the planets, but that's all we see about their relative sizes. We can't tell from the picture whether the orbits are all on the same plane or on different planes. We do see that the planets all revolve in the same direction, and that the Earth, sun, and Venus do not always stay lined up.

Ptolemy accounted for the motion of the sun, planets, and stars in the sky by saying they revolved around the Earth every 24 hours. Copernicus accounted for those motions by saying that the Earth revolved around its own axis every 24 hours. How could someone in the late 16th Century decide between these two models? Both were in accord with the observations that had been made.

In the early 1600s the telescope was invented, and in 1610 Galileo built his own telescope with a magnification of about 33 times, using it to study the skies. One of his students suggested an experiment that might distinguish between the Ptolemaic and Copernican models. Venus was too far from the Earth to be seen as anything other than a spot of light. But according to Ptolemy's model, viewed from the earth, at most only a small crescent-shaped part of Venus will ever be illuminated by the sun when Venus is at the extreme distance from the sun in its epicycle. From Copernicus' model, however, we can deduce that from the earth Venus should go through all the phases of illumination, just like the moon: full, half, crescent, dark, and back again. Galileo looked at Venus through his telescope for a period of time and saw that it exhibited all phases of illumination, and this he took to be proof that Copernicus' model was correct.

Not a lot of other people were convinced, however. Telescopes were rare and not very reliable: they introduced optical illusions such as halos from the imperfections in the glass and the mounting. Why should astronomers have trusted Galileo's observations?

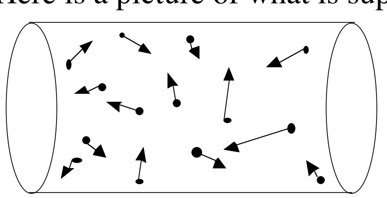
It was more due to Newton that something like Copernicus' model of the universe was finally accepted. Newton deduced from his laws of motion that the orbits of the Earth, sun, and the planets would have to be ellipses, not circles. And the distances between them would have to be much greater than supposed. Using Newton's laws, Edmond Halley predicted correctly the return of a comet that had been observed in 1682. Telescopes were better, with fewer optical illusions, and common enough that most astronomers could use one, so better and better observations of the planets and stars could be made. Those observations could be deduced from the Copernican-Newtonian model, while new epicycles had to be invented to account for them in the Ptolemaic model.

Note that each model is supposed to be similar to the universe in only a few respects, ones that would have an effect on how we could see the objects in the universe from the earth. Differences, such as whether Venus is rocky or gaseous, are not supposed to matter for those observations. If the model is correct, then reasoning by analogy—very precise analogy—certain claims can be deduced.

3. The kinetic theory of gases

> This theory is based on the following postulates, or assumptions.
> 1. Gases are composed of a large number of particles that behave like hard, spherical objects in a state of constant, random motion.
> 2. The particles move in a straight line until they collide with another particle or the walls of the container.
> 3. The particles are much smaller than the distance between the particles. Most of the volume of a gas is therefore empty space.
> 4. There is no force of attraction between gas particles or between the particles and the walls of the container.
> 5. Collisions between gas particles or collisions with the walls of the container are perfectly elastic. Energy can be transferred from one particle to another during a collision, but the total kinetic energy of the particles after the collision is the same as it was before the collision.
> 6. The average kinetic energy of a collection of gas particles depends on the temperature of the gas and nothing else.
>
> J. Spencer, G. Bodner, and L. Rickard, *Chemistry*, 1998

Here is a picture of what is supposed to be going on in a gas in a closed container.



The molecules of gas are represented as dots, as if they were hard spherical balls. The length of the line emanating from a particle models the particle's speed; the arrow models the direction the particle is moving. The kinetic energy of a particle is defined in terms of its mass and velocity; the model defines what is meant for a collision to be elastic.[3]

What are we to make of these assumptions of the kinetic theory of gases? Some are certainly false: Molecules of gas are not generally spherical and are certainly not solid; the collisions between molecules and the walls of a container or each other are not perfectly elastic; there is some gravitational attraction between the particles and each other and also with the container.

However, seen as reasoning by analogy, these assumptions can be invoked to make predictions. We are comparing molecules of gases in a closed container to hard spherical objects in a state of constant random motion.

But no such hard spherical objects actually exist. So is the comparison, the analogy, to an imaginary situation, to something called "ideal gas"?

We can view the model as proceeding by abstraction: To the extent that we can ignore how the shape of molecules of gases is not spherical, and ignore the physical attraction between molecules, and . . . , we can draw conclusions that may be applicable to actual gases. To the extent that the differences between actual gases and the abstractions don't matter, we can draw conclusions. But how can we tell if the differences matter?

The model suggests that the pressure of a gas results from the collisions between the gas particles and the walls of the container. So if the container is made smaller for the same amount of gas, the pressure should increase; and vice-versa, if the container is made larger, the pressure should be less. So the pressure should be proportional to the inverse of the volume of the gas. That is, the model suggests a claim about the relationship of pressure to volume in a gas. Experiments can be performed, varying the pressure or volume, and they are close to being in accord with that claim.

Other laws are also suggested by the model: Pressure is proportional to the temperature of the gas, where the temperature is taken to be the average kinetic energy of the gas. The volume of the gas should be proportional to the temperature. The amount of gas should be proportional to the pressure. All of these are confirmed by experiment.

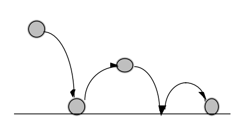
Those experiments confirming predictions from the model do not mean the model is more accurate than we thought. Collisions still aren't really elastic; molecules aren't really hard spherical balls. The kinetic theory of gases is a map, useful where the differences don't matter.

4. Euclidean geometry

Euclidean plane geometry speaks of points and lines: a point is location without dimension, a line is extension without breadth. No such objects exist in our experience. But Euclidean geometry is remarkably useful in measuring and calculating distances and positions in our daily lives.

Points are abstractions of very small dots made by a pencil or other

[3] In contrast, here is a picture of what happens in an inelastic collision between a rubber ball and floor. Each time the ball hits the ground, some of its kinetic energy is lost either through being transferred to the floor or in compressing the ball.



implement. Lines are abstractions of physical lines, either drawn or sighted. So long as the differences don't matter, that is, so long as the size of the points and the lines are very small relative to what is being measured or plotted, we can rely on whatever conclusions we draw.

No one asks (anymore) whether the axioms of Euclidean geometry are true. Rather, when the differences don't matter, we can calculate and predict using Euclidean geometry. When the differences do matter, as in calculating paths of airplanes circling the globe, Euclidean plane geometry does not apply, and another model, geometry for spherical surfaces, is invoked.

Euclidean geometry is a deductive theory: A conclusion drawn from the axioms is accepted only if the inference is valid. It is a purely mathematical theory, which, taken as mathematics, would appear to have no application since the objects of which it speaks do not exist. But taken as a model it has applications in the usual sense, reasoning to conclusions when the differences don't matter.

Some people, however, invoke points and lines as actual things, abstract, non-sensible things, which then provide the other side of the analogy. To reason solely by abstraction without positing the other side for an analogy seems to impede our ability to imagine clearly. There is no harm in such imaginings, so long as we are clear that the abstract things of which we speak are nothing more than what is left when we ignore much of our experience; they are meant to fill out our picture.[4]

5. The acceleration of falling objects

Galileo argued that falling objects accelerate as they fall: They begin falling slowly and fall faster and faster the farther they fall. He didn't need any mathematics to demonstrate that: A heavy stone dropped from 6 meters will drive a stake into the ground much farther than if it is dropped from 6 centimeters.

Galileo also said the reason that a feather falls more slowly than an iron ball when dropped is because of the resistance of air. He argued that at a given location on the earth and in the absence of air resistance, all objects should fall with the same acceleration. He claimed further that the distance traveled by a falling object is proportional to the square of the time it travels. Today, from many measurements, the equation is given by:

(*) $d = 9.80 \text{ meters}/_{\text{sec}^2} \cdot t^2$ t measured in seconds

For example, if you drop a ball from the Empire State Building, after 4 seconds it should travel: $d = 9.80 \text{ m}/_{\text{sec}^2} \cdot (4 \text{ sec})^2 = 156.8 \text{ m}$.

The equation (*) is a model by abstraction: We ignore air resistance and the shape of the object, considering only the object's mass and center of gravity. If the differences don't matter, then a calculation from the equation, which is really

[4] Another view of Euclidean geometry is that the axioms implicitly define the subject matter. But since there is nothing in experience that precisely fits those axioms, the subject matter, again, must be abstract things.

a deduction, will hold. But often the differences do matter. Air resistance can slow down an object: If you drop a hat from an airplane, it will reach a maximum velocity when the force of the air resistance equals the force of acceleration.

With this model there is no visual representation of that part of experience that is being described. There is no point-to-point conceptual comparison, nor are we modeling a static situation. The model is couched in the language of mathematics; equations can be models, too.

6. Newton's laws of motion and Einstein's theory of relativity

Newton's laws of motion are taught in every elementary physics course and are used by physicists. Yet modern physics has replaced Newton's theories with Einstein's and quantum mechanics. Newton's laws, physicists tell us, are false.

But can't we say that Newton's laws are correct relative to the quality of measurements involved, even though Newton's laws can't be derived from quantum mechanics? Or perhaps they can if a premise is added that we ignore certain small effects? Yet how is that part of a theory?

A theory is a schematic representation of some part of the world. We draw conclusions from the representation (we calculate or deduce). The conclusion is said to apply to the world. The reasoning is legitimate so long as the differences between the representation and what is being represented don't matter. Newton's laws of motion are "just like" how moderately large objects interact at moderately low speeds; we can use those laws to make calculations so long as the differences don't matter. Some of the assumptions of that theory are used as conditions to tell us when the theory is meant to be applied.

7. Ether as the medium of propagation of light waves

In the 19th Century light was understood as waves. In analogy with waves in water or sound waves in the air, a medium was postulated for the propagation of light waves: the ether. Using that assumption, many predictions could be made about the path of light in terms of its wave behavior. Attempts were made to isolate or further verify the existence of the ether. But the predictions that were made turned out to be false. When a better theory was postulated by Einstein, one which assumed no ether and gave as good or better predictions in all cases where the ether assumption did, the theory of ether was abandoned.

8. Classical propositional logic

Suppose we ignore everything about claims except whether they are true or false and how they are built up as compounds using "and", "or", and "not". We can use the symbols $\wedge$, $\vee$, and $\neg$ to stand for these abstracted versions of the connectives. Then the following tables summarize the usual classical understanding of how the truth-value of a compound relates to the truth-value of its parts, where A, B stand for claims, and T stands for "true," F for "false."

A	B	A ∧ B
T	T	T
T	F	F
F	T	F
F	F	F

A	B	A ∨ B
T	T	T
T	F	T
F	T	T
F	F	F

A	¬A
T	F
F	T

There is also a table for "if . . . then . . .", but these tables are enough for our discussion. Using them repeatedly we can calculate the truth-value of any compound claim built using "and", "or", and "not" if we know the truth-values of its parts. For example, "Either Ralph is a dog or Howie is not a cat, and George is a duck, but Birta is a dog" would correspond to $(A \vee \neg B) \wedge (C \wedge D)$. Using these determinations, we can model whether one claim follows as conclusion from a finite set of other claims: It must be impossible for all the premises to be true and the conclusion false at the same time, which is the case based solely on the form of the propositions (relative to these connectives) if there is no assignment of truth-values to the premises that makes all of them true and the conclusion false.

This model or theory of reasoning is quite different from the other models we've seen. Though it is in some part descriptive and begins with abstractions from ordinary language, it is not the worse for finding that people do not reason as it describes. Though it may be thought of as describing nonsensible, abstract propositions of which it makes correct predictions, as a model of reasoning its role is *prescriptive*. It says that this is the correct way to reason.

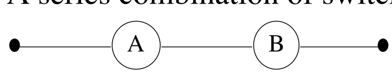
Or rather, it says that this is the correct way to reason so long as the differences don't matter. But if the differences do matter, say if we wish to suspend judgment on some claims and see where that leads us, then this model won't be useful for drawing conclusions about how we should reason.

Typically in presenting this model, logicians do not point out the general principle on which it is based: The model is applicable when all that is of concern in reasoning is whether a claim is true or false and how it is built up in terms of these connectives. If we focus on the result of the abstracting and invoke abstract propositions as the subject matter of logic, then it is hard to see the general principle or even to conceive of the model as an abstraction from experience.

9. Electrical switches

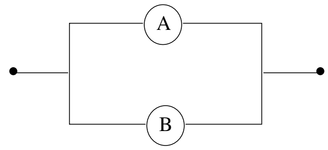
We can model circuits of electrical switches using the same tables as for classical logic. Instead of "true" and "false," T and F stand for "on" and "off," where these are the only two possibilities for the switches.

A series combination of switches A and B is:



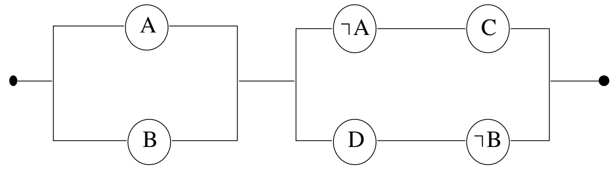
The picture is meant to convey that there is a wire, the line, connecting the switches to two terminals, the dots. Current flows through a switch if it is on, and not through it if it is off. Current will flow through a series combination exactly when both switches are on, just as the table for ∧ tells us.

A parallel combination of switches A and B is:



Current will flow through a parallel combination exactly when one or the other or both switches are on, just as the table for $\vee$ tells us.

If we let $\neg$A stand for a switch that is off when A is on, and on when A is off, then we can model any circuit built from these kinds of switches using the tables for $\wedge$, $\vee$, and $\neg$. For example, $(A \vee B) \wedge ((\neg A \wedge C) \vee (D \wedge \neg B))$ models:



Depending on how we view the mathematics, the analogy is between the symbols and the way we manipulate them on the one hand and the circuits on the other, or it is between reasoning with claims and switching circuits. Either way, we can draw conclusions about the flow of electricity in very complicated circuits using this model, so long as we can ignore the differences: the wire has length and diameter and electrical resistance; switches aren't just on or off, but have some short period where the current is neither fully on or off when they are switched; and so on. This model is descriptive, not prescriptive.

Models, theories, and truth

We have seen models of static situations (the map) and of processes (acceleration of falling objects, gases in a container). We have seen examples of models that are entirely visual, intended as point-to-point comparisons, and of models formulated entirely in terms of mathematical equations. We have seen models where the assumptions of the model are entirely implicit (the map) and models where the assumptions are quite explicit (Newton's laws of motion).

In all the examples either the reasoning is clearly reasoning by analogy or can be seen as proceeding by abstraction much as in reasoning by analogy. We do not ask whether the assumptions of a theory or model are true, even if that was the intention of the person who created the theory. Rather, we ask whether we can use it in the given situation: Do the similarities that are being invoked hold and do the differences not matter? Even in the case of Newton's laws of motion, where it would seem that what is at stake is whether the assumptions are true, we continue to use the model when we know that the assumptions are false in those cases where, as in any analogy, the differences don't matter.[5] In only one example

[5] C. Kittel, W. D. Knight, and M. A. Ruderman, 1965, pp. 55–56 say in their textbook:

> *Newton's third law*. Whenever two bodies interact, the force F_{12} on the second body (2) due to the first body (1) is equal and opposite to the force F_{21} on the first (1) due to the second (2): $F_{12} = -F_{21}$. There are inherent limitations to the validity of the third law: we believe . . . that all signals or forces have a finite propagation velocity. The third law, however, states that F_{12} is equal and opposite to F_{21} when both are measured

did it seem that what was at issue was whether a particular assumption of the theory was actually true of the world (the ether).

Laws in science are false when we consider them as representing all aspects of some particular part of our experience.[6] The key claim in every analogy is false in the same way. When we say that one side of an analogy is "just like" the other, that's false. What is true is that they are "like" one another in some key respects that allow us to deduce claims for the one from deducing claims for the other. In the same way we can use the assumptions of the theory so long as the differences—what we ignore—don't matter.

The exact conditions for an analogy or abstraction to qualify as a model or a theory seem to be quite informal. The term *model* seems to be more readily applied to what can be visualized or made concrete, while the term *theory* seems to be applied to analogies that are fairly formal and have very explicitly stated assumptions or general principles. But in many cases it is as appropriate to call an example a theory as to call it a model, and there seems to be no definite distinction between those terms.[7]

Theories and confirmation

From theories we can make predictions, and we say that when a prediction turns out to be true it confirms (to some degree) the theory. But this is not the same as confirming an explanation. As discussed in Epstein, 2002B, confirming an explanation is understood as showing (to some degree) that the claims doing the explaining are true, whereas we have seen that it rarely makes sense to say that the claims that make up a theory—the assumptions of the theory—are true or false.

We cannot say that verifications of the relation of pressure, temperature, and

---

at the same time. This requirement is inconsistent with the finite time interval required for one particle to feel the force the second particle is exerting. In atomic collisions the third law is therefore not always a good approximation. In automobile collisions it is quite a good approximation, because the duration of the collision is long in comparison with the time it takes a light signal to traverse a crumpled automobile.

[6] Michael Scriven, 1961 and 1962, discusses this issue. In Scriven, 1962, p. 212 he says:

> The examples of physical laws with which we are all familiar are distinguished by one feature of particular interest for the traditional analyses [of explanations]—they are virtually all known to be in error. Nor is the error trifling, nor is an amended law available which corrects for all the error. The important feature of laws cannot be their literal truth.

Nancy Cartwright, 1981, p. 381 says:

> What is important to realize is that if the theory is to have considerable explanatory power, most of its fundamental claims will not state truths, and that this will in general include the bulk of our most highly prized laws and equations.

[7] Compare Jack Birner, 2002 (discussed on pp. 109–117 of this volume of BARK):

> It is quite common to use "theory" for a set of abstract propositions and "model" for a set of propositions on a lower level of abstraction. . . . [T]he question of whether or not something is called a theory or a model is mostly a matter of convention. What matters is how theories or models of different degrees of idealization are related, i.e. their relative levels of abstraction. Therefore, I will use "theory" and "model" interchangeably.

volume in a gas confirm that molecules are hard little balls and that all collisions are completely elastic. We cannot say that a carpenter's square fitting exactly into a wooden triangle that is 5cm x 4cm x 3cm confirms the theorem of Pythagoras. Nor can we say that finding a tree at the corner of 100 W and 100 S in Minersville disconfirms the model given by the map of Minersville.

Except in rare instances where we think (usually temporarily) that we have hit upon a truth of the universe to use as an assumption in a theory, we do not think that the assumptions of a theory are true or false. Theories are best understood as analogies or abstractions, and they can only be true in the sense of correctly representing that part of the world they are meant to represent in a particular situation. In brief, we can only say of a theory such as Euclidean plane geometry or the kinetic theory of gases that it is *applicable* or not in a particular situation we are investigating, where a "situation" is just some part of the world we can describe using claims.

To say that a theory is applicable is to say that though there are differences between the world and what the assumptions of the theory state, those differences don't matter for the conclusions we wish to draw. Often we can decide if a theory is applicable only by attempting to apply it. We use the theory to draw conclusions in particular instances, claiming that the differences don't matter. If the conclusions—the predictions—turn out to be true (enough), then we have some confidence that we are right. If a prediction turns out false, then the model is not applicable there. We do not say that Euclidean plane geometry is false because it cannot be used to calculate the path of an airplane on the globe; we say that Euclidean plane geometry is inapplicable for calculating on globes.

When we make predictions and they are true, we confirm a range of application of a model. When we make predictions and they are false, we disconfirm a range of application, that is, we find limits for the range of application of a model. More information about where the model can be applied and where it cannot may lead, often with great effort, to our describing more precisely the range of application of a model. In that case, the claims describing the range of application can be added to the theory. We often use mathematics as a language for making the art of analogy precise. For example, for Newton's laws of motion we can give limits on the size and speed of objects for which the theory is applicable. But in many cases it is difficult to state precisely the range of application. In other cases, such as the map of Minersville, it hardly seems worthwhile to state explicitly the range of application.

We want our theories to be as widely applicable as possible. Eventually we hope to find theories whose range of application can be precisely and clearly stated, where we can say that the theory is applicable whenever this is the case, where we are justified in saying that the theory is true.

But even then we would not be justified in saying that a particular claim that is used as an assumption of the theory is true. Rather, the claim is true in those situations in which the theory as a whole is applicable.[8]

Many other terms are used to describe theories: a theory is valid, a theory is true, a theory holds, a theory works for, ... I can find no other sense to these than to assimilate them to the question of whether a particular situation, or class of situations, to which we wish to apply a theory is within the range of the theory.

Modifying theories in the light of the evidence

How do we determine whether a theory is good? How do we determine whether one theory is better than another? As we have seen, the criteria cannot in general include whether the assumptions of the theory are true, or, as is sometimes said, "realistic." Here is what Milton Friedman says:

> In so far as a theory can be said to have "assumptions" at all, and insofar as their "realism" can be judged independently of predictions, the relation between the significance of a theory and the "realism" of its "assumptions" is almost the opposite of that suggested by the view under criticism. Truly important and significant hypotheses will be found to have "assumptions" that are wildly inaccurate descriptive representations of reality, and, in general, the more significant the theory, the more unrealistic the assumptions (in this sense). The reason is simple. A hypothesis is important if it "explains" much by little, that is, if it abstracts the common and crucial elements from the mass of complex and detailed circumstances surrounding the phenomena to be explained and permits valid predictions on the basis of them alone. To be important, therefore, a hypothesis must be descriptively false in its assumptions; it takes account of, and accounts for, none of the many other attendant circumstances, since its very success shows them to be irrelevant for the phenomena to be explained.
>
> To put the point less paradoxically, the relevant question to ask about the "assumptions" of a theory is not whether they are descriptively "realistic," for they never are, but whether they are sufficiently good approximations for the purpose in hand. And this question can only be answered by seeing whether the theory works, which means it yields sufficiently accurate predictions.[9]

Friedman, then, agrees with what I have said. But in the last sentence he goes further. He, and many others following him, say that the sole criterion for judging whether a theory is good (or as he says, "valid") is whether it yields sufficiently accurate predictions. As he says,

> The only relevant test of the validity of a hypothesis is comparison of its predictions with experience.[10]

Certainly it is important to get good predictions. But if the assumptions are

[8] "Galileo said that a body subject to no forces has a constant velocity. (This is called Newton's first law of motion.) We have seen that this statement is true only in an inertial reference system—it defines an inertial system." Kittel et al., 1965, p. 61

[9] Friedman, 1953, pp. 14–15.

[10] Friedman,1953, p. 8. Compare Stephen E. Landsburg, 1993, p. 10, "Assumptions are not tested by their literal truth but by the quality of their implications."

neither true nor true for the situation being analyzed, on what basis should we base our acceptance of further predictions? A good track record in the past? But who has been keeping score? Perhaps it is just judicious uses of the theory, always supplemented with assumptions—often unstated—that result in predictions that are sufficiently accurate. What allows us to distinguish between a theory that makes good predictions whose assumptions are clearly false in the case at hand and astrology? After all, for many centuries astrology was the best theory around for divining human fate. Its predictions often came true, since they were sufficiently vague to allow that. And few people were keeping track of the predictions that turned out false.

There must be some criterion beyond the truth of predictions made using the theory that counts for whether a theory is good. Consider what we do when we discover that a prediction made using a theory is false.

When the theory of switching circuits predicts inaccurately for transistors, we look for what differences matter: What have we ignored in the case at hand that results in false predictions? Switches are not instantaneously on or off, there is electrical resistance, ... We then factor those aspects of the situation into our new theory.

When Newton's laws of motion result in inaccurate predictions for objects moving near the speed of light, we note that the theory had been assumed true for all sizes and speeds of objects and then restrict the range of application.

Even the prescriptive theory of classical propositional logic has been modified. When predictions made using it (certain inferences shown to be valid based on its assumptions) were found to be anomalous, counterintuitive, and not good prescriptions for reasoning, attention turned to what aspects of claims had been ignored. Certain differences matter, and depending on what aspects are considered significant, such as subject matter or ways in which the claim could be true, different propositional logics have been set out as models of reasoning.[11]

On the other hand, when the theory of the ether resulted in false predictions, no modification was made to the theory, for none could be made. That theory did not abstract from experience, ignoring some aspects of situations under consideration, but postulated something in addition to our experience, something we were able to show did not exist. The theory was completely abandoned.[12]

When we find that a prediction is false we have three choices:

---

[11] See Epstein, 2000.

[12] David Isles (personal communication) disagrees: "How is the ether any less of an abstraction from experience than, say, the notion of force or even the notion of energy as it appears in physics. Thus, we experience water which carries water waves, we experience less substantial air which carries sound waves, and (a slight abstraction from experience) we have a very insubstantial fluid which carries light waves." Isles is correct that we can hypothesize the existence of an ether by analogy with those other mediums of transmission, but that is distinctly different from beginning with something of experience and ignoring some of its properties. The notions of force and energy in physics are not postulated to be things or substances; see the discussion of this point in relation to the nature of cause and effect in Epstein, 2001, p. 203.

1. The theory or model can be understood as an analogy or for use as reasoning by abstraction. That is, many aspects of our experience are ignored and only some few are considered significant. Then tracing back along that path of abstraction we can try to distinguish what difference there is between our model and our experience that matters. What have we ignored that cannot in this situation be ignored?

If we cannot state precisely what difference it is that matters in some general way, then at best the false prediction sets some limit on the range of applicability of the model or theory. We cannot use the theory here—where "here" means this situation or ones that we can see are very similar.

But we are driven to find precisely the difference that matters and try to factor it into our theory. That is, we try to devise a complication of our theory in which that aspect of our experience is taken into account. As with Einstein's improvement of Newton's laws, we get a better theory that is more widely applicable and which explains why the old theory worked as well as it did and why it failed in the ways it failed. We improve the map: By adding further assumptions we can pay attention to more in our experience, and that accounts for the differences between the theories.

2. Some theories, such as the theory of the ether, are not based on abstraction but postulate entities or aspects of the world in addition to what we have from our experience or other trusted theories. In that case a false prediction, or more usually many false predictions, lead us to consider such a postulate to be false. We abandon the theory.

3. One of our examples, however, does seem to fit Friedman's prescription for deciding whether a theory is good. Modern physics says that there is no preferred frame of reference in the universe. Hence, so far as the truth of the assumptions or their applicability to the case at hand, the Ptolemaic model of the universe is as good as the Copernican. But we choose the Copernican model because it yields better predictions. We can always modify the Ptolemaic system to account for observations by adding more epicycles. That will yield true predictions in some limited cases, but will rarely work for other objects than the one for which an epicycle is posited. Moreover, the explanations we can make of the movement of planets and other objects in the solar system are clearer and simpler in the Copernican model: It produces better explanations.

The truth of the assumptions does matter for some theories: those based on claims that are not abstractions from experience. When it makes sense to talk of the truth or falsity of a particular dubious claim among the assumptions of a theory, deriving true predictions from the theory can help to confirm that claim;

a false prediction may serve to prove the dubious assumption false.

A theory that is meant as an analogy or abstraction from our experience, ignoring much of what we wish to study and focusing only on what we consider the most important aspects, comes with an (often implicit) range of application. Deriving a true prediction from such a theory confirms to some extent that range of application. A false prediction can serve to set limits on the range of application. False predictions can also lead us to prefer a simpler theory based on abstraction that yields better predictions.

True predictions are never enough to justify a theory. Indeed, the problem is that we do not "justify" a theory, nor show that it is "valid." What we do in the process of testing predictions is show how and where the theory can be applied. And for us to have confidence in that, either we must show that the claims in the theory are true or show in what situations the differences between what is represented and the abstraction of it in the theory do not matter. True (enough) predictions help in that. But equally crucial is our ability to trace the path of abstraction so that we can see what has been ignored in our reasoning and why true predictions serve to justify our ignoring those aspects of experience. Without that clear path of abstraction, all we can do is try to prove that the claims in the theory are actually true. Without that clear path or without reason to believe the claims are true, we have no more reason to trust the predictions of a theory than we have to trust the predictions of astrology.

When there is no path of abstraction: a model in economics

Milton Friedman's analysis of theories discussed above was meant to justify theories of conventional (classical, orthodox, neo-classical) economics, such as the theory of competitive equilibrium. Such theories begin with the assumption that all persons involved in the market are *rational* in the following sense:

1. People are motivated solely by the goal of maximizing utility.
2. They have fixed, transitive preferences.
3. They have available to them all the relevant information and use it.
4. They reason perfectly.

These assumptions about people are false, and have been demonstrated to be false in many behavioral studies.[13] This has led to the charge that predictions from such theories cannot be relied on. The theories are charged with having "unrealistic" premises.[14]

But, as we saw above and as Friedman stressed, false, unrealistic premises do not disqualify a theory from being very useful, that is, from giving us good reason to believe that its predictions are true. From assumptions that are abstractions from experience and which in no sense could be true, not even in particular cases as with Euclidean geometry, we can reason to conclusions in which we are justified in having great confidence. Hence, the charge by many

[13] See Richard H. Thaler, 1991, for a discussion and references.

[14] For a discussion and references see Mark Blaug, 1980.

economists that theories based on the assumption that people are rational are unrealistic is not a serious criticism of such theories.

The question for such theories is what is the range of application for reasoning by them. When we obtain false predictions, how can we use those to clarify the limits of the range of application and improve the theory? And certainly with theories based on the assumption that people are rational there are plenty of false predictions.[15]

Let us ask how we can trace back the path of abstraction for assumptions (1)–(4) to modify them to obtain better theories.

With (1) it is clear: We can assume other motives of people. Doing so complicates the models, but can yield better theories. We can also assume that people want only to satisfy some level of utility rather than maximizing.[16]

But for each of (2)–(4) there is a problem: What is being abstracted from experience to create these assumptions of the theory?

Certainly people do not have fixed transitive preferences.[17] But what is it that we are ignoring in their behavior that allows us to assume they do? At best we can say that sometimes they have fixed transitive preferences, and in those cases the theory should apply.

When do people have all the relevant information and use it in making decisions? Perhaps in very restricted situations, such as buying and selling currencies. False predictions from this assumption would limit the theories to such cases. Or models can be devised which take into account limited access to relevant information.[18]

But for (4) there is no such obvious route to limiting the range of application of the theories. It is not simply that people do not normally reason well. It is not that they do not want to reason well. Rather, most people do not have the skills to reason as well as demanded by theories based on the assumption that people are rational, as any teacher of critical thinking can attest.[19] The norms of reasoning well are prescriptive, not descriptive.

There is no abstraction from experience in postulating that people reason perfectly. To assume that people are rational is to ascribe capabilities to people that they do not have. It is not like Euclidean geometry, where we ignore much from experience; it is like postulating the existence of an ether, where we assume

---

[15] See, for example, Blaug, 1980, or Thaler, 1991, or Paul Ormerod, 1998.

[16] See Gerd Gigerenzer and Reinhard Selzen, 2001, which also discusses the history of this notion of rationality.

[17] Thaler, 1991, cites numerous psychological studies to this effect.

[18] See again Gigerenzer and Selten, 2001.

[19] There is an additional problem in applications of the assumption of rationality. It requires not just that people reason well one inference at a time. Rather, in making decisions it is assumed that people can survey all information and all consequences of certain assumptions at once. That is, people are assumed to be able to survey a completed infinite set. This contrasts with, say, Euclidean geometry, which, in applications, requires only a potentially infinite set of points: we can always find an additional point as required by the axioms.

something of the world that is in addition to what we know there is.

There is no point along the path of abstraction where we can modify theories based on the assumption that people are rational in order to take into account some further aspect of their reasoning. The moment we assume that people reason well only some times, there is no longer a model. As with theories built on the assumption that there is an ether, when we encounter false predictions we can only abandon the theory.

But sometimes theories of conventional economics do predict well. Friedman would account for that and justify the use of such theories by saying that though people do not act rationally, they act "as if" they are rational. It is worth quoting him at length to see how his justification proceeds.

> Consider the density of leaves around a tree. I suggest the hypothesis that the leaves are positioned as if each leaf deliberately sought to maximize the amount of sunlight it receives, given the position of its neighbors, as if it knew the physical laws determining the amount of sunlight that would be received in various positions and could move rapidly or instantaneously from any one position to any other desired and unoccupied position. Now some of the more obvious implications of this hypothesis are clearly consistent with experience: for example, leaves are in general denser on the south than on the north side of trees but, as the hypothesis implies, less so or not at all on the northern slope of a hill or when the south side of the trees is shaded in some other way. Is the hypothesis rendered unacceptable or invalid because, so far as we know, leaves do not "deliberate" or consciously "seek," have not been to school and learned the relevant laws of science or the mathematics to calculate the "optimum" position, and cannot move from position to position? Clearly, none of these contradictions of the hypothesis is vitally relevant; the phenomena involved are not within the "class of phenomena the hypothesis is designed to explain"; the hypothesis does not assert that leaves do these things but only that their density is the same *as if* they did. Despite the apparent falsity of the "assumptions" of the hypothesis, it has great plausibility because of the conformity of its implications with observation. We are inclined to "explain" its validity on the ground that sunlight contributes to the growth of leaves and that hence leaves will grow denser or more putative leaves survive where there is more sun, so the result achieved by purely passive adaptation to external circumstances is the same as the result that would be achieved by deliberate accommodation to them. This alternative hypothesis is more attractive than the constructed hypothesis not because its "assumptions" are more "realistic" but rather because it is part of a more general theory that applies to a wider variety of phenomena, of which the position of leaves around a tree is a special case, has more implications capable of being contradicted, and has failed to be contradicted under a wider variety of circumstances.[20]

Friedman's hypothesis about leaves seeking to maximize the amount of sunlight they receive cannot be used for reasoning by analogy. It does not begin

[20] Friedman, 1953, pp. 19–20.

by either (a) looking at a real situation and comparing it to the growth of leaves, allowing us to distinguish what are the similarities and what are the differences, or (b) abstracting from experience to state exactly what are the points of similarity that are supposed to hold, ignoring all else.

Rather, what he has posited is not an abstraction, but the addition of properties to a given situation. We are asked to suppose that leaves behave anthropomorphically with the skills of a terrific calculator. And then we are asked to ignore that as well. This doesn't make sense as a method of reasoning: Why should we have confidence that predictions made from such a hypothesis will be accurate? That some of the predictions turn out to be accurate cannot be enough, any more than they are in astrology. We need to know why they turn out accurate in order to have confidence in the theory or model.

The alternative hypothesis of passive adaptation that he presents is better, but not for the reasons he gives; rather it is better for the reason he says is not meaningful. Namely, we have better reason to accept the alternative hypothesis precisely because we can see that in this case it is reasonable to believe it is true. No clearly false assumptions incapable of fitting into reasoning by analogy or abstraction have been made.

Friedman gives another example, which leads to the rationality assumption:

> . . . the hypothesis that the billiard player made his shots *as if* he knew the complicated mathematical formulas that would give the optimum directions of travel, could estimate accurately by eye the angles, etc., . . . Our confidence in this hypothesis is not based on the belief that billiard players, even expert ones, can or do go through the process described; it derives rather from the belief that, unless in some way or other they were capable of reaching essentially the same result, they would not in fact be *expert* billiard players.
>
> It is only a short step from these examples to the economic hypothesis that under a wide range of circumstances individual firms behave *as if* they were seeking rationally to maximize their expected returns (generally if misleadingly called "profits") and had full knowledge of the data needed to succeed in this attempt; *as if*, that is, they knew the relevant cost and demand functions, calculated marginal cost and marginal revenue from all actions open to them, and pushed each line of action to the point at which the relevant marginal cost and marginal revenue were equal. Now, of course, businessmen do not actually and literally solve the system of simultaneous equations in terms of which the mathematical economist finds it convenient to express this hypothesis, any more than leaves or billiard players explicitly go through complicated mathematical calculations or falling bodies decide to create a vacuum. . . .
>
> Confidence in the maximization-of-return hypothesis is justified by evidence of a very different character [from the truth of its assumptions]. This evidence is in part similar to that adduced on behalf of the billiard-player hypothesis—unless the behavior of businessmen in some way or other approximated behavior consistent with the maximization of returns, it seems unlikely that they would remain in business for long. Let the apparent

> immediate determinant of business behavior be anything at all—habitual reaction, random chance, or whatnot. Whenever this determinant happens to lead to behavior consistent with rational and informed maximization of returns, the business will prosper and acquire resources with which to expand; whenever it does not, the business will tend to lose resources and can be kept in existence only by the addition of resources from the outside. The process of "natural selection" thus helps to validate the hypothesis—or, rather, given natural selection, acceptance of the hypothesis can be based largely on the judgment that it summarizes appropriately the conditions for survival.[21]

Just as Friedman is wrong about the leaves, he is wrong about the billiard player, and hence wrong in his justification of the maximization-of-return hypothesis. Moreover, he has hit—unwittingly—on the greater problem with both the billiard-player and the maximization-of-return hypotheses: they can only explain success, not failure of the agent. They apply to much less of the situation than is needed to make a good model: How do we explain the failure of a billiard player to make a shot? He was calculating badly? How do we explain the bankruptcy of a company? They weren't calculating correctly? A good model of either situation has to account for failure as well as success, for all the "players."

Friedman introduces his "as if" talk in order to justify economic theories based on assuming that people are rational. The problem with the rationality assumptions, though, as in his examples, is that they are not abstractions but positing of properties that are clearly false. It matters how assumptions are derived from experience in order for us to be justified in saying that they do not have to be true to get good predictions. If they are obtained by abstraction, then they need not be true: all that matters is the scope of their application. We can trace back the path of abstraction. But with hypotheses that postulate additional properties, if those properties are not true—that is, if the model is not applicable in the case at hand—we have no reason to trust the model.[22]

It will be extremely difficult to give criteria for what we mean when we say that an assumption is not an abstraction from experience but rather postulates something of the world in addition to our experience. But those who wish to base theories on assumptions that are clearly false have the burden to show in what way their theories arise by abstraction; it is for them to justify their theories.

Should we then abandon economic theories based on the assumption that people are rational? Consider a comparison. In mathematical logic certain formal systems are studied that are modifications of the formal logics meant to formalize reasoning, yet in which there can be infinitely long sentences used as propositions. Such an assumption is not an abstraction from actual reasoning, and no one argues that we should consider such a system as prescriptive of how we should reason.

---

21 Friedman, 1953, pp. 21–22.

22 There is another oddity in saying that people act as if they are rational when indeed they are not. Typically, the evidence we invoke for claiming a person is rational is how he or she acts. I discuss in an appendix of Epstein, 2001, the problem of when it is appropriate to call someone "rational" without basing that on the evidence of his or her explicit reasoning.

Yet neither do logicians say that those who study such systems are not doing logic. The systems based on infinite sentences are suggested as modifications of systems that clearly are logic in the sense of models of how to reason well. They are in the same "family," the same area of study. They have an internal coherence and beauty and illustrate many connections between other formal theories. They are, if you like, "pure" logic.

Throughout mathematics, areas of study that grew out of abstracting from experience, such as the theory of transfinite numbers, have been developed with many assumptions that cannot be seen as abstractions yet which have yielded internally coherent, beautiful theories that often link various other such theories. This is what is called the study of pure mathematics. Such theories are not criticized for having no application. Moreover, a pure mathematician is not expected to understand the applied mathematics used to model, say, ballistics or biology. Applied mathematics and pure mathematics are related though independent disciplines.

Economic theories have been developed in the last few decades that abandon the assumption that people are rational.[23] As with applied mathematics, from which they draw heavily, they are much more complicated. They rarely yield predictions of specific events so much as describe overall behavior of an economy, presenting something like the normal conditions from which causal claims could be derived.[24] They are to conventional economics much as applied mathematics is to pure mathematics. As with pure mathematics, conventional economics has a great beauty and intellectual attraction and internal coherence that make it of interest to many and which justify its study. It is up to those who practice it, however, to justify further why we should have any confidence in its predictions as applied to experience.[25] I hope to have shown here on what grounds we can judge such justifications.

Bibliography

Birner, Jack
2002 *The Cambridge Controversies in Capital Theory*
Routledge.

Blaug, Mark
1980 Economic methodology in one easy lesson
British Review of Economic Issues, May 1980.
Also in Mark Blaug, *Economic History and the History of Economics*, New York University Press, 1986.

Cartwright, Nancy
1981 The reality of causes in a world of instrumental laws
In R. Boyd, P. Gasper, and J. D. Trout, eds., *The Philosophy of Science,* MIT Press, 1991 , pp. 379–386.

[23] See Ormerod, 1998, for a non-technical introduction.

[24] See the discussion of cause and effect in Epstein, 2001.

[25] Similar comments apply to decision theory as it is customarily formulated, for it, too, assumes people to be perfect reasoners. See, for example, Michael Resnik, 1987.

Epstein, Richard L.
2000 *Propositional Logics*
2nd edition, Wadsworth.
2001 *Five Ways of Saying* "*Therefore*"
Wadsworth.
2002A *Critical Thinking*
2nd edition, Wadsworth.
2002B Arguments and explanations
Bulletin of Advanced Reasoning and Knowledge, 1: *Proceedings of the 2nd Conference on Logic and Reasoning of the Advanced Reasoning Forum, Bucharest, Romania*, *2000*, New Europe College, 2002, pp. 3–17.
(Also at <www.AdvancedReasoningForum.org>.)
Friedman, Milton
1953 The methodology of positive economics
In *Essays in Positive Economics*, University of Chicago Press, 1953, pp. 3–43.
Gigerenzer, Gerd and Reinhard Selten
2001 Rethinking rationality
In *Bounded Rationality*, Gigerenzer and Selten, eds., MIT Press, pp. 1–12.
Kittel, Charles, Walter D. Knight, and Malvin Ruderman
1965 *Mechanics* (*Berkeley Physics Course—Volume 1*)
McGraw-Hill.
Landsburg, Stephen E.
1993 *The Armchair Economist*
The Free Press.
Ormerod, Paul
1998 *Butterfly Economics*
Pantheon.
Resnik, Michael
1987 *Choices*
University of Minnesota Press.
Scriven, Michael
1961 The key property of physical laws—inaccuracy
In *Current Issues in the Philosophy of Science,* eds. Herbert Feigl and Grover Maxwell, Holt Rinehart and Winston, pp. 91–101.
1962 Explanations, predictions, and laws
In *Minnesota Studies in the Philosophy of Science,* vol. III, eds. H. Feigl and G. Maxwell, University of Minnesota Press, pp. 170–230.
Spencer, J., G. Bodner, and L. Rickard
1998 *Chemistry*
John Wiley and Sons, Inc.
Thaler, Richard H.
1991 *The Winner's Curse*
The Free Press.