Underdetermination, Black Boxes, and Measurement

Teru Miyake*†

This article introduces the notion of a kind of inference called black box measurement and argues that it is both historically and philosophically significant. Thinking about certain classic cases of underdetermination using this notion can give us a better understanding of how these cases are resolved. I take the main philosophical problem of black box measurement to be the justification of assumptions that are needed in order to make these measurements. I sketch some ways in which such enabling assumptions might be justified.

1. Introduction. The main aim of this article is to introduce the notion of what I call *black box measurement* and to argue that this is an important form of inference that has not been sufficiently investigated in the philosophical literature. This is intended to be a brief overview of the idea. I start off with a short anecdote and an example from the history of science to introduce the idea of black box measurement and its connection to the classical philosophical problem of underdetermination. I then discuss what I take to be the most significant philosophical problem with regard to black box measurement—the justification of what I call *enabling assumptions*.

*To contact the author, please write to: 14 Nanyang Drive, School of HSS #06–05, Singapore 637332; e-mail: tmiyake@ntu.edu.sg.

†Much of the material in this article comes from my dissertation. I want to thank Michael Friedman, George Smith, Pat Suppes, Helen Longino, and Tom Ryckman for advice and support during the writing of the dissertation. Financial support for the dissertation was provided by a Whiting Fellowship. The Stanford Philosophy Department provided a great place for developing my ideas, and I want to thank Quayshawn Spencer, Angela Potochnik, Alistair Isaac, Johanna Wolff, Sally Riordan, Ehud Lamm, and Joel Velasco for numerous enlightening discussions. A six-month stay at the Max Planck Institute for the History of Science in Berlin gave me time to write up my dissertation, and I thank Lorraine Daston for welcoming me into her department there. Finally, I want to thank Nanyang Technological University in Singapore, and the colleagues who made my first year in Singapore an easy one: Chenyang Li, Andres Luco, Lina Jansson, and Winnie Sung. Research in my first year there was supported by a Start-Up Grant (SUG) from NTU.

Philosophy of Science, 80 (December 2013) pp. 697–708. 0031-8248/2013/8005-0004$10.00
Copyright 2013 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/673729 Published online by Cambridge University Press

2. An Anecdote: Cube. Suppose that one day, archeologists unearth a mysterious artifact—a perfect black cube, a meter on a side, cool to the touch, made of what looks like the blackest possible steel. They decide, rather unimaginatively, to call the artifact "Cube." It is a mere curiosity at first, but scientists soon find that it has some mystifying features. The material it is made out of is incredibly hard—it cannot be broken, cut, pierced, drilled, or dynamited. All attempts to take images of the inside of Cube using standard technology fail. On one face are several white dots that look as if they are projected onto the face from within Cube. The dots move across the face of Cube, tracing out trajectories over time.

Now suppose we are scientists trying to figure out what is going on inside Cube. We will find, unfortunately, that our options are severely limited, since we have found no way of accessing its interior. What do we do? Perhaps the only thing to do is simply to assume that there are certain lawlike connections between the internal and external states of Cube—that is, the dynamics of the external states depends somehow upon the dynamics of the internal states. We then make hypotheses about (a) the dynamics of the internal states of Cube and (b) the laws that connect the internal states to the external states of Cube. From these hypotheses, we deduce predictions about the dynamics of the external states. If these predictions match our observations of the external states, we say that those hypotheses have been confirmed. This is, of course, essentially the hypothetico-deductive method.

There is something very unsatisfactory about the hypothetico-deductive method when applied to this particular case. Since we have no antecedent knowledge about the internal states of Cube, there is enormous leeway in the hypotheses we can come up with. For any given dynamics of the external states of Cube, there will be many different sets of hypotheses that are consistent with those dynamics. Our theory of the internal states of Cube is massively underdetermined by our observations of the external states. Faced with this predicament, we might give up on the idea that we can gain any knowledge at all about the internal states of Cube and simply change our aim to that of predicting the dynamics of the external states of Cube without making any claim to having any knowledge about the internal states. We will thus become instrumentalists about our theory of the internal states of Cube.

3. Underdetermination and Keplerian Astronomy. The story of Cube is a fiction, but I think there have been cases in the history of science where scientists have found themselves in similar predicaments. Consider the state of planetary astronomy in the sixteenth century. All of our knowledge about the solar system came from the observation of the motions of the planets as they moved across the night sky. More specifically, we can think of

https://doi.org/10.1086/673729 Published online by Cambridge University Press

ourselves as being located inside an immense, hollow, black sphere, on the inner surface of which the constellations are painted. We can then determine the positions of the planets on this sphere, as seen from the earth, and thus record their apparent motions over time. We cannot, however, know how far away a planet is from us merely by looking at it. So we are, in effect, looking at the two-dimensional projection, onto the celestial sphere, of the actual three-dimensional motions of the planets through space. Moreover, although we did not know for sure in the sixteenth century, we are observing these motions from a platform, the earth, that is itself moving.

Drawing out the analogy with the story of Cube, we can think of the apparent motions of the planets as corresponding to the external states of Cube, while the actual three-dimensional motions correspond to the internal states. Like the scientists studying Cube, astronomers in the sixteenth century faced a problem of radical underdetermination. Famously, the apparent motions of the planets across the night sky were compatible with three different theories—the Ptolemaic, the Copernican, and the Tychonic—in which the actual three-dimensional motions of the planets are radically different from each other.

How was this state of underdetermination eventually resolved? Well, suppose the method of astronomy is, as for Cube, hypothetico-deductive. You make hypotheses, deduce the observable consequences of these hypotheses, and then you compare these consequences with actual observations. Since the problem is that there were three theories that could fit the observations to the same degree of precision, we might think that one way of resolving the underdetermination is through increasing the precision in the actual observations. Johannes Kepler showed, however, in the "Introduction" to the *Astronomia Nova*, that with minor modifications, the Ptolemaic, Copernican, and Tychonic systems can be made to give exactly the same predictions for the apparent two-dimensional motions of the planets—they can be made empirically equivalent. Thus, a mere increase in precision of the observations of the apparent two-dimensional motions could not resolve the underdetermination.

What actually happened is that Galileo turned his telescope to the skies and observed that Venus has phases, just like the moon. This situation is inconsistent with the Ptolemaic theory, so it was eliminated from contention. The emergence of a new kind of technology, the telescope, allowed us to bring a new kind of evidence to bear on the question of what the actual motions of the planets are. Let us put history aside for a moment, though, and think philosophically about this situation of underdetermination. From the standpoint of the philosophy of science, the resolution of this particular case of underdetermination is somewhat unsatisfactory, since it came about through an event that was, to a considerable degree, contingent. What if Galileo had

https://doi.org/10.1086/673729 Published online by Cambridge University Press

never decided to turn a telescope to the skies? Could the underdetermination have been resolved in another way?

In fact, prior to Galileo's discovery, Kepler claimed in the *Astronomia Nova* to have an argument that the Ptolemaic theory is not the correct theory of planetary motion. I just got done saying that Kepler showed that all three theories of the planetary motions could be made empirically equivalent to each other and so could not be distinguished on the basis of observations of the apparent two-dimensional motions of the planets. The three theories predict very different motions for the planets through three-dimensional space, however. What could potentially resolve the underdetermination is the measurement of a new variable. Up until the end of the sixteenth century, the observations the astronomers had been using were two-dimensional positional coordinates of the planets against the backdrop of the constellations. What was needed was the measurement of distances in three-dimensional space. But the only data available were the two-dimensional coordinates. So the three-dimensional distances would have to be constructed from the two-dimensional observations.

The methods Kepler used in making these measurements in the *Astronomia Nova* were quite ingenious and very complicated. To be very brief, one of the great contributions of Kepler to astronomy that has been overlooked is his realization that, for any of the three theories of planetary motion, the directly observed two-dimensional motions of a planet such as Mars is the result of two components of three-dimensional motion: the motion of the earth and the motion of Mars for the Copernican theory, the motion of the sun and the motion of Mars for the Tychonic theory, and the motion of the center of the epicycle along the deferent and the motion of Mars along the epicycle for the Ptolemaic theory. Since the apparent two-dimensional motion of Mars is, in a sense, the sum of these two component motions, you can, in effect, subtract out one of these component motions from the apparent two-dimensional motion, in order to determine the other component motion. So a better determination of one of these motions (e.g., the motion of the earth in the Copernican theory) could potentially lead to a better determination of the other motion (e.g., the motion of Mars in the Copernican theory). And these determinations can be played off one another—if I determine the orbit of earth more accurately, it will now allow me to determine the orbit of Mars more accurately, and then I could use my corrected orbit of Mars to determine the orbit of earth even more accurately.[1]

Through a series of determinations of the earth and Mars orbits, Kepler finds in the *Astronomia Nova* that the component of motion corresponding to the earth orbit on the Copernican theory is not circular, and the sun

1. This is an unorthodox view of Kepler's method. See Miyake (2014) for a detailed explanation.

https://doi.org/10.1086/673729 Published online by Cambridge University Press

is not at its center. The same series of determinations, when interpreted in terms of the Ptolemaic theory, would show that the motion of Mars along its epicycle is not circular, and the point of attachment of the epicycle onto the deferent is not at its center. So, as pointed out by Stephenson (1987, 55), this would mean, for the Ptolemaic theorist, that the orbit of Mars has something very much like a miniature version of the earth orbit as its epicycle, with all its complicated motions. Kepler goes on to make the bold pronouncement that if we similarly determined the orbits of Jupiter and Saturn, we would find that their epicycles, too, would look like mirror images of the earth orbit (Kepler 1992, 337). Of course, the Copernican theory furnishes a ready explanation—they all look like images of the earth orbit because they all are images of the earth orbit. There was no reason for all the epicycles to look so much alike according to the Ptolemaic theory, however, so this would make the Ptolemaic theory extremely implausible. So this is a way in which the Ptolemaic theory could potentially have been rejected without the need for the observation of the phases of Venus.

Now, when Kepler thus chooses the Copernican theory over the Ptolemaic theory, this choice would probably best be construed as an inference to the best explanation. But what I think is worth examining is how Kepler ever gets in the position of making this inference. There was a situation of empirical equivalence between three different theories of the planetary motions—empirical equivalence relative to observations of the two-dimensional apparent motions of the planets. Kepler then uses a method whereby he is able to measure the motions of the planets through three-dimensional space, and these new measurements allow him to reject the Ptolemaic theory. Of course, he did not get these new measurements for free. A certain number of assumptions had to be made in order to obtain the measurements of three-dimensional motions from observations of the apparent two-dimensional motion. For example, there is the assumption that the two-dimensional motions of the planets are the result of exactly two component three-dimensional motions, and the assumption that these component motions add together in accordance with the rules of Euclidean geometry.

These assumptions might be questioned by a determined Ptolemaic theorist, but it is also hard to deny that there seemed to be something that was going right about Kepler's theory. Why? Because of the success of his method of "playing off" the determinations of the orbits of the earth and Mars against each other. If the orbit of Mars is not determined to some degree correctly, this would throw off our determination of the orbit of the earth, which would in turn throw off our further determination of the orbit of Mars. This cascade of errors would lead, presumably, to wildly implausible orbits. But this did not happen. So the success Kepler had using this method was seemingly an indication that he was at least latching onto some real, well-behaved feature of the planetary motions. So even if the Ptolemaic theorist rejected

https://doi.org/10.1086/673729 Published online by Cambridge University Press

the original assumptions that were used in the method, the very success of the method would still be a good reason to accept the measurement results themselves.

4. Black Box Measurement and Enabling Assumptions. The previous section is not intended to be a full analysis of the method of Kepler. I simply gave a brief description as an illustration of a way in which we might think about such problems. In any case, I think problems that are similar to that of Cube and planetary astronomy can be found surprisingly often in science. These are problems where we are trying to measure properties of parts of systems that are inaccessible to us, based upon observations of accessible parts of those systems. For example, seismologists try to determine the properties of the deep regions of the earth, from observations of seismic waves at the surface of the earth (note that there is an extensive nonphilosophical literature on inverse problems, which is similar to the kind of problem I am describing here, although this literature generally focuses on mathematical aspects). This kind of problem can be found in various types of neuroimaging as well.

Here is a more general characterization. Suppose there is a complicated, partially inaccessible system that I want to acquire knowledge about. A complicated system is one that consists of many parts, those parts having various properties and relations with each other. A system is *partially inaccessible* if we can only confidently measure a proper part of the properties of, and relations between, the parts of that system. I call the properties that we can confidently measure the *accessible properties*. I also sometimes speak of *accessible parts*, by which I simply mean the parts of the system that have properties that we can confidently measure. In order to determine the properties and relations of the inaccessible parts, we must make inferences based upon what we know about the accessible parts. For example, in the case of planetary astronomy, we want to determine three-dimensional distances between planets. In the case of seismology, we want to determine properties like density at various points deep within the earth. The kinds of facts about the inaccessible parts that are being inferred are the specific values of variables that represent properties of the inaccessible parts. I am, in other words, talking about measurement. I thus refer to this kind of problem as *black box measurement*.

Let us consider again the example of Keplerian astronomy. There were three different theories of the motions of the planets that were consistent with the observed two-dimensional motions of the planets at that time. In order to overcome this underdetermination problem, the measurement of a new variable was required. I believe that this process of the construction of new measurements of properties of the inaccessible parts of a partially in-

https://doi.org/10.1086/673729 Published online by Cambridge University Press

accessible system, namely, black box measurement, is not well understood. In particular, in many such cases, new measurements are enabled by making certain assumptions about the systems being investigated. A central philosophical question with regard to black box measurement is how these enabling assumptions can be justified. I claimed that, in the case of Kepler, some justification for the measurements, and the assumptions that went into them, came from the very success of the method that he used. But exactly what kinds of such enabling assumptions can there be, and how does the justification of such assumptions work in general? Below, I lay out a basic framework for thinking about these questions.

First, since we are trying to measure properties of inaccessible parts of a complicated system through observation of its accessible parts, we need to presuppose some kind of relation between the accessible and the inaccessible parts. Note that we can often be very confident about the results of certain measurements in cases where we need to presuppose such a relation in order to carry out the measurement. For example, I do not have direct access to the amount of electric current flowing through a wire, but I can have great confidence in the value I measure using a galvanometer. At least part of this confidence has to do with what I call *antecedent familiarity*. If an object is of a type that is familiar to me, I can safely assume certain facts about that object. For example, I know that the laws of mechanics will apply to cannonballs, so I can use them to determine how long it will take a cannonball to drop from a given height. Thus, if I could measure very precisely the amount of time it takes for a cannonball to fall, I could use that measurement, very confidently, to determine the height from which the cannonball is dropped. There have been cases in the history of science, however, where we have wanted to know facts about an object that is utterly unlike anything else we knew of at the time. The solar system is a good example of such an object. For all astronomers knew in the sixteenth century, the solar system could have been radically different from anything else we knew of, so it was hard to know what a reasonable assumption to make about the solar system was.

There are two difficulties in carrying out a black box measurement that would make us lack confidence in such a measurement, particularly if the system we are making the measurement on is antecedently unfamiliar. The first difficulty is theory mediation. You have to make measurements of the properties of the inaccessible parts of the system, based upon observations of the accessible parts. In order to make such measurements, you need to presuppose that a particular nomic relation holds between the accessible properties and the inaccessible properties. If the relation you use to make the measurement is not known antecedently, then the question naturally arises as to how you can know that the measurement is correct.

https://doi.org/10.1086/673729 Published online by Cambridge University Press

This difficulty has been described and dubbed the *problem of nomic measurement* by Chang (2004, 59–60).

This problem is compounded, however, by a second difficulty in the case of black box measurement, which I call the *combining of effects*. Again, the root of this difficulty is that you have to make measurements of inaccessible properties based upon observation of accessible properties. Suppose the system you are making a measurement upon is complicated. If so, there could be more than one part of the system that has an effect on the accessible parts. If you want to measure a property of one of those parts, you might have to separate out, or decompose, the effects of the various parts on the accessible part. If you do not antecedently know the composition of the system, however, you might not know exactly how to carry out the decomposition. If so, you might not be confident that the measurement you make using it is correct.

So, in order to carry out a black box measurement, you must, in some cases at least, make assumptions about (1) the nomic relations between the accessible and the inaccessible properties and (2) the way in which the effects of the system combine together. An important question here is, how can these assumptions be justified? Now, in cases where we are dealing with systems that are antecedently familiar, we can find some justification for these assumptions. For example, in seismology, we are almost certain that the laws of physics do apply to the material deep inside the earth, so there is no question of the justification of the applicability of the laws of physics when we want to set up a black box measurement of the properties of this material. There is, however, some question about the physical properties of this material, including its elastic properties. But assumptions must be made about these elastic properties in order to set up the proper nomic relations (using wave equations) in order to allow us to make measurements of the properties of the deep interior of the earth. Further, we do not have much knowledge about what the structure of the deep interior is like, so there are assumptions that must be made about the way in which the effects of the system combine together.

Seismologists are in a much better position with regard to measuring properties of the deep interior of the earth, however, than planetary astronomers in the sixteenth century were with regard to the planets. Whereas seismologists are very confident that the laws of physics apply to the interior of the earth, there was no reason to think, in the sixteenth century, that the laws of physics apply to the heavens. To go a bit further with this point, one way in which we can view Newton's accomplishment in the *Principia* is that he showed that you can set up black box measurements in order to measure mechanical, not just kinematic, properties of the planets. And a central question with regard to Newton's method is the justification of the assumptions he needs in order to make these measurements. How was it at all justified to

https://doi.org/10.1086/673729 Published online by Cambridge University Press

apply the laws of mechanics to the motions of the planets, when we had very little evidence at the time that they apply to the planets at all?[2]

5. Strategies for Justification. I think a detailed study of actual historical cases is needed to answer the question of how such assumptions get justified. But I have a couple of proposals about strategies that might be used. In what follows, I will refer to figures 1 and 2 below. To explain them briefly, the circles represent properties of a system, and the arrows represent nomic relations between these properties.

The first strategy is converging measurement. Suppose there is some system that we can represent by figure 1. There is a node *X* with two arrows out from it, arrow *a* to node *Y*, and arrow *b* to node *Z*. Suppose both *Y* and *Z* are accessible properties, that is, we have a way of measuring their values confidently, but we do not have too much confidence in the nomic relations *a* and *b*. In this situation, there are two different ways of measuring *X*, through observation of *Y* using relation *a*, and through observation of *Z* using relation *b*. If we carry out both measurements, and we get approximately the same result, that is, they converge, this is good reason to think that the measurements are good, and that the measurement of *X* is correct. Note that we can get converging results even if the relations *a* and *b* are not strictly true of the system—it could be the case that, say, relation *a* simply holds to a good approximation under the circumstances of the measurement. It turns out that we have more reason to believe that the measurement result itself is correct than the assumptions we made in order to make the measurement.

There are other strategies besides converging measurement in which we can get the enabled measurements themselves to be evidence that the assumptions were correct. The following strategy is what I call *decomposition*, which I illustrate in figure 2. Here, the accessible property *Z* is affected by both the inaccessible property *X*, via the relation *a*, and the inaccessible property *Y*, via the relation *b*. Suppose we do not have too much confidence in the relations *a* or *b*, and we want to measure *X*. We might first try guessing the effect of *Y* on *Z*, subtracting that effect out, and then measuring *X* using the relation *a*. We now have a way of modeling the effect of *X* on *Z* using the relation *a*. Now subtracting that effect out, we measure the value of *Y* using the relation *b*. Using this new value of *Y*, we model the effect of *Y* on *Z*. We subtract out that effect and measure a more refined value for *X*. Using this new, refined value for *X*, we model the effect on *Z*, and we now come up with a new, refined measurement for *Y*. One example of this method is the way in which Kepler carried out his measurement of the Mars orbit. If my measurements of *X* and *Y* seem to be converging on certain val-

2. I have discussed these issues briefly in a review (Miyake 2013) of William Harper's recent book on Newton.

https://doi.org/10.1086/673729 Published online by Cambridge University Press

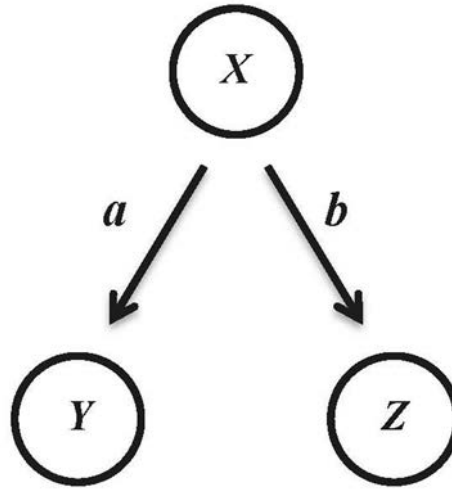


Figure 1.

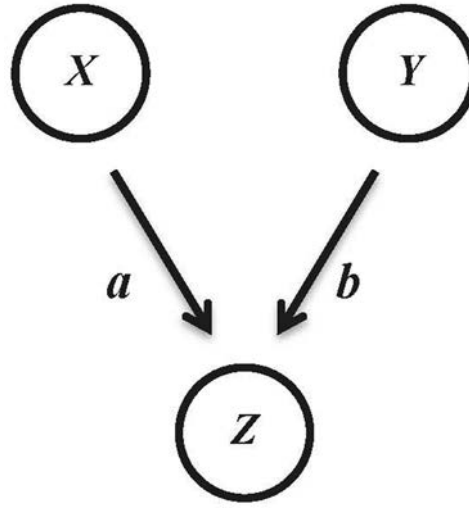


Figure 2.

ues, then this is good evidence that this relational structure is approximately correct and the relations *a* and *b* are also at least approximately applicable. Why? Suppose the relation *a* is not approximately applicable. Then when we model the effect of *X* on *Z* and subtract out this modeled effect in order to measure *Y*, we do not expect to get a good value when we measure *Y*.

https://doi.org/10.1086/673729 Published online by Cambridge University Press

Then, when we model the effect of Y and subtract it out to measure X, we should expect this measurement not to give a good value for X, and thus it should not agree with the previous value for X. Thus, if the sequence of values for X is converging, this is evidence that the values for X and Y are correct.

There are a couple of things I should note here. First, I have been treating X and Y as single-valued variables, but we can generalize this picture so that X, Y, and Z are vectors, or continuous functions. In fact, for all of the cases I have actually examined in detail, including seismology and planetary astronomy, the accessible properties and inaccessible properties are represented as continuous functions. Second, we could also have much more complicated structures, involving many more properties and nomic relations. I do not claim that these are the only ways in which enabling assumptions might be justified—as I said, these are mere proposals, and I think we need to examine more historical examples of black box measurement. I should note that there is already some philosophical literature on converging measurement (e.g., Myrvold and Harper 2002; Forster 2007), and the strategy of decomposition can, I think, be construed as a type of bootstrapping (Glymour 1980) or demonstrative induction (Norton 2003). What I want to emphasize here is that philosophers have not given enough thought to the process of choosing enabling hypotheses. Suppose you want to make a black box measurement. In order to do so, you need to choose enabling hypotheses that make this measurement possible. But you also need to choose them in such a way that minimizes the risks if the hypotheses turn out to be wrong.[3] Part of this involves choosing hypotheses that will eventually allow for the strategies of converging measurement or decomposition to be carried out. Other factors must be considered, however. For example, a major risk of decomposition is the possibility that there could be unknown parts of the system that are having an effect on the accessible properties. So the strategy of decomposition must include something that I call *exploration*—searching for these unknown parts of the system. The framework presented in this article is useful for thinking about such issues.

6. Conclusion. In this article, I have introduced the idea of a black box problem and have made the case that these are interesting problems that deserve a detailed philosophical analysis. I think there are several important cases in the history of science in which such black box measurements have been done. Notably, this picture can show us how some cases of underdetermination could potentially get resolved. I have taken some steps towards such an analysis in this article, but more work must be done to flesh out this picture.

3. George Smith has written on this issue in Newton (see Smith 2002).

https://doi.org/10.1086/673729 Published online by Cambridge University Press

REFERENCES

Chang, Hasok. 2004. *Inventing Temperature: Measurement and Scientific Progress*. Oxford: Oxford University Press.
Forster, Malcolm. 2007. "A Philosopher's Guide to Empirical Success." *Philosophy of Science* 74:588–600.
Glymour, Clark. 1980. *Theory and Evidence*. Princeton, NJ: Princeton University Press.
Kepler, Johannes. 1992. *New Astronomy.* Trans. William H. Donahue. Cambridge: Cambridge University Press.
Miyake, Teru. 2013. "Essay Review: Isaac Newton's Scientific Method." *Philosophy of Science* 80 (2): 310–16.
———. 2014. "Underdetermination and Decomposition in Kepler's *Astronomia Nova*." *Studies in History and Philosophy of Science*, forthcoming.
Myrvold, Wayne, and William Harper. 2002. "Model Selection, Simplicity, and Scientific Inference." *Philosophy of Science* 69:135–49.
Norton, John. 2003. "A Material Theory of Induction." *Philosophy of Science* 70:647–70.
Smith, George. 2002. "The Methodology of the *Principia*." In *The Cambridge Companion to Newton*, ed. I. B. Cohen and George Smith. Cambridge: Cambridge University Press.
Stephenson, Bruce. 1987. *Kepler's Physical Astronomy.* New York: Springer.

https://doi.org/10.1086/673729 Published online by Cambridge University Press