

Perspectives on Psychological Science

<http://pps.sagepub.com/>

What If Social Scientists Had Reviewed Great Scientific Works of the Past?

David Trafimow and Stephen Rice

Perspectives on Psychological Science 2009 4: 65

DOI: 10.1111/j.1745-6924.2009.01107.x

The online version of this article can be found at:

<http://pps.sagepub.com/content/4/1/65>

Published by:



<http://www.sagepublications.com>

On behalf of:



[Association For Psychological Science](http://www.sagepub.com/content/4/1/65)

Additional services and information for *Perspectives on Psychological Science* can be found at:

Email Alerts: <http://pps.sagepub.com/cgi/alerts>

Subscriptions: <http://pps.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Issues in Publishing, Editing, and Reviewing

What If Social Scientists Had Reviewed Great Scientific Works of the Past?

David Trafimow and Stephen Rice

New Mexico State University

ABSTRACT—*One might question whether the great works in the history of science would get good reviews if subjected to the type of reviewing process to which psychologists are forced to submit their manuscripts. In some ways, behavioral scientists are too critical, and in other ways they are insufficiently so. To explore these issues, we imagine that great works from the history of nonsocial sciences were submitted for review in behavioral science journals and present simulated editor letters summarizing the comments of behavioral science reviewers. The philosophical underpinnings and justifications of the arguments are discussed, and recommendations for improved reviewing are offered.*

Few people would be willing to assert that progress in the behavioral sciences has been as impressive as the progress made in other sciences. Usually, when people discuss these matters, they present reasons to justify the differences: the other sciences have existed for longer; the behavioral sciences are more difficult because the mind is less tangible than the body, the world, or the universe; there is more funding for other sciences than for the behavioral sciences; and so on. Although some of these justifications may have some merit, there is another possibility that behavioral scientists rarely consider: perhaps they are not as effective in their scientific reasoning and in the way they evaluate scientific research.

It is not the most enjoyable thing in the world to have one's area criticized. In addition, abstract arguments about the philosophy of science tend to have little effect on behavioral scientists. To sidestep both of these problems, we present great works from other sciences and imagine that they have been

submitted to psychology journals to be evaluated by editors and reviewers. For each great work, we present a simulated editor letter that summarizes the comments of simulated reviewers. Subsequent to the presentations and editor letters, we provide an analysis of the letters. Our hope is that this will render our points more enjoyable, more concrete, and easier to understand than if they were presented in the usual abstract format. Also, doing it this way provides an element of drama that otherwise would be difficult to achieve. And if the reader gains an increased appreciation of the controversies and accomplishments in other areas of science, we would consider that to be a plus, particularly if the lessons are eventually applied successfully to psychology.

We tried hard to make the reviews to be presented resemble reviews that we have actually experienced either in response to manuscripts that we have submitted to journals or in response to manuscripts of others on which we were reviewers (a situation in which, consequently, we were able to read the comments of the other reviewers). Some of the reviews also resemble opinions contemporary with the great works; it is not only behavioral scientists who have committed solecisms of reasoning, as other scientists have done so as well. A major goal is to dramatize that a reviewer who wishes to find fault is always able to do so. Therefore, the mere fact that a manuscript can be criticized provides insufficient reason to evaluate it negatively. Rather, one must consider the possible gains against the possible losses involved, and if the former outweigh the latter, the work should be evaluated positively in spite of the potential criticisms.

NONSOCIAL SCIENCE SUBMISSIONS TO PSYCHOLOGY JOURNALS

The 25,000 Mile Spherical Earth

Brief Description

The ancient Greeks noted that there were problems with the Persian notion of a box Earth (in which the sky was assumed to

Address correspondence to David Trafimow, Department of Psychology, MSC 3452, New Mexico State University, P.O. Box 30001, Las Cruces, NM 88003-8001; e-mail: dtrafimo@nmsu.edu.

extend down to the sides of a flat Earth to keep the seas from pouring out). For example, what kept the box from falling? One could, of course, argue that the box was held up by pillars, but in that case what are the pillars standing on? Clearly, one can go on in this direction ad infinitum. Worse yet, why do objects fall in the direction they do and not in other directions? A way to circumvent the problems of infinite regression and falling objects was to conclude that the Earth was spherical and define “down” as in the direction of the center of the sphere (e.g., Aristotle). Given this realization, it became clear how to determine the circumference of the Earth. Eratosthenes and his associate placed sticks in the ground at Syene and at Alexandria (500 miles north of Syene) during the summer solstice when the sun was nearly directly overhead at Syene, but not at Alexandria. The sun obviously did not cast much of a shadow in Syene, but it did in Alexandria. By measuring the angle of the shadow in Alexandria and using geometry, Eratosthenes was able to calculate that the Earth was approximately 25,000 miles in circumference (a recent determination, around the equator, suggests it is a little over 24,902 miles). Imagine that Eratosthenes had submitted a manuscript to a psychology journal with the argument that the Earth’s circumference is 25,000 miles.

Editor Letter

Dear Eratosthenes,

I am sorry to be the bearer of bad news, but the reviewers were unanimous in their negative opinion and, based on my independent reading, I was forced to agree. I sent your manuscript to four reviewers, all recognized experts in the area, and they all identified different areas of weakness. For example, Reviewer A, who identifies himself/herself as Hecataeus of Miletus, agrees with you to the extent of saying that the Earth is not a rectangle.¹ But this reviewer believes that the Earth is a flat disc and states that your notion of a sphere simply does not accord with sensory observations. Reviewer B felt that you cherry picked among the empirical findings. This reviewer takes your point about the curved shadow cast by the Earth upon the moon suggesting a spherically shaped Earth, but notes that there are alternative explanations. For example, perhaps the Earth is a flat disc (as Reviewer A suggested) but that it is oriented in space in the precise way necessary to cast a curved shadow. Or perhaps the shadow is not caused by the Earth at all, but by something else. Given the plethora of empirical observations in favor of the Earth being flat, you simply need more evidence to make your argument plausible and to eliminate potential alternative explanations.

Reviewer C focused more on your reasoning. This reviewer admits your point that positing a flat Earth, of whatever shape, causes problems with infinity. But this reviewer also states that it

is better to have a problem with infinity than to make a proposal that is obviously wrong. This reviewer also feels that your reasoning is too complex and will not be understood by our readers. Reviewer D agreed with Reviewer C but also noted that even if you were correct about the Earth being spherical, your comments about the circumference of the Earth should not be taken seriously. First, the stick that you put in the ground in Alexandria might not have pointed straight up, thereby biasing your measurement of the angle of the shadow. Second, there was no independent verification of your measurements, thereby reinforcing the possibility that your measures were biased. For example, how can you be sure that the distance between Syene and Alexandria is 500 miles—do you have any validity data for the measures you used? Third, your dependence on geometry is itself a possible problem. Perhaps there is a problem with one of Euclid’s assumptions. Finally, if the circumference of the Earth were really 25,000 miles, it would mean that the area of the Earth is extremely large, which would not be in accord with the estimates of recognized luminaries in the field, such as Hecataeus of Miletus. You should consult his works before making any more outrageous statements that fail to accord with the literature.

DECISION: Reject

The Heliocentric System and Galilean Relativity

In the year of his death (1543), Nicolas Copernicus’s *De revolutionibus orbium coelestium* (*On the Revolutions of the Heavenly Spheres*)—in which he determined that the planets circle the sun—was published. Although the heliocentric system was simpler, more elegant, and allowed for easier calculations than did the geocentric system, it was extremely slow to be accepted. Because if the Earth moves around the sun, why don’t we feel it? Also, if the Earth moved around the sun, then the positions of the planets against the stars should seem to shift, according to the phenomenon known as parallax. Unfortunately, the planets were too far away for this phenomenon to be observed with the measures that were available at the time. Both of these arguments were used as points against the heliocentric system. Galileo Galilei argued for a heliocentric system in a way that dealt with both of these problems. Galileo argued that we do not feel the movement of the Earth because we are moving with it, and so, relative to our movement, the Earth does not move and there is nothing to feel. But if one takes the stars as the frame of reference, than the movement of the Earth would be obvious. And as far as the lack of the parallax phenomenon is concerned, Galileo argued that the stars are too far away for the parallax phenomenon to be detected by the methods available at the time.² Of course, this meant that the universe had to be quite large, which was a controversial claim at the time.

¹Actually, Hecataeus of Miletus predated Eratosthenes by more than 2 centuries, but we hope that the reader will forgive the anachronism.

²Perhaps ironically, improvements on Galileo’s invention (or reinvention) of the telescope eventually made it possible to detect parallax.

Editor Letter

Dear Galileo,

I have received reviews from four recognized experts in the field and none of them argue in favor of publication. Reviewers A and B both point out that the Ptolemaic system, particularly as it was augmented recently by Brahe, results in more precise calculations than does your theory. Reviewer B admits that this precision comes at the cost of a more complicated theory but nevertheless feels that because the geocentric system is adequate, your theory is simply unnecessary. Reviewer C focused on your relativity argument (that we and the Earth move together and so although the Earth does not move relative to us, it does move relative to the stars), which he/she feels is necessary to make your system work. Unfortunately, although this reviewer finds your relativity argument to be extremely clever, there is very little in the way of empirical evidence. Rather than presenting a rigorous set of empirical findings to support that argument, you seem to depend more on anecdotal evidence and clever rhetoric. How do we know for sure that an object dropped from the top of the mast of a sailing ship would describe a parabola from the reference point of someone on shore but would describe a straight line from the reference point of the person who dropped it? If you are going to make extraordinary claims, then you have to back them up with extraordinary data. Finally, Reviewer D points out that everything you say is so unlikely that only irrational people would believe it, despite the fact that it could be correct.

In addition to the comments made by the reviewers, I would add that your justification of the lack of the parallax phenomenon seems unlikely at best. You claim that the stars are so far away as to prevent us from detecting the phenomenon with the measures that are currently available, but there is no reason to believe that the universe is sufficiently large for this to be so, and there are no respectable scientists who believe that. In summary, you ask the reader to believe too many unlikely things: your relativity argument, with its unsupported assumptions, and your distance argument, which seems more a fantasy than anything that a reasonable person would take seriously.

DECISION: Reject

Newton's Laws of Motion

Newton's laws of motion can be stated quite simply. The first law states that "every object in a state of uniform motion tends to remain in that state of motion unless an external force is applied to it." The second law describes how force changes motion: Force = Mass \times Acceleration. According to Nobel Laureate Leon Lederman (1993), this is the most important single equation in the history of physics, despite the fact that the variables in the equation do not have independent definitions. The third law states that action equals reaction or, more precisely, if Object A exerts a force on Object B, then Object B exerts an equal

and opposite force on Object A. For example, if the Earth exerts a force on an apple, the apple exerts an equal and opposite force on the Earth. Furthermore, this is a requirement for all forces: gravity, magnetism, electricity, and so forth. The implications these simple laws suggest for understanding and predicting states of the universe are legion.

Editor Letter

Dear Isaac,

I had the good luck to obtain reviews from four experts and all within the relatively quick period of 3 months. Although there were some positive comments, the general consensus was negative. Reviewer A felt that your theory was too abstract and mathematical to be of interest to scientists who are interested in the real universe. In comparison with other, similar theories, even as far back as those from the ancient Greeks, your theory does not have the descriptive richness that is required. Reviewer B also complained about this, but is more willing to tolerate it if you could back it up with data. But, in fact, you do not do so. Therefore, Reviewer B considers your theory to be nothing but an ingenious story.

Reviewer C made the most serious criticism. This reviewer points out that you fail to provide independent definitions of your terms. What is mass? What is force? What is acceleration? Reviewer C acknowledges that you can define each of these in terms of the others, but that merely makes your theory circular. To avoid the circularity, at least one of these terms has to be defined independently of the others, and you fail to do so. Consequently, there is no way to know what your theory actually means. Perhaps if your theory included more descriptive richness, as Reviewer A recommends, it would help with Reviewer C's problem.

Reviewer D also pointed out a serious shortcoming. According to this reviewer, your laws merely summarize what others, such as Galileo and Kepler have found, without actually adding anything new to the literature. In fact, your first law is a restatement of Galileo, and your other laws do not really tell us anything new. For example, Kepler's laws explain the orbits of the planets, and Reviewer D states quite frankly that there is simply no need for your manuscript. After my own independent reading of your manuscript, I agree with the reviewers' criticisms. Therefore, I am declining to publish your manuscript. I am sorry to have had to be the bearer of bad news, but I trust that you will find the reviewers' comments useful if you decide to submit your manuscript to another journal. Thank you for considering our journal as an outlet for your research.

DECISION: Reject

William Harvey and Systemic Circulation*Brief Description*

Harvey's teacher, a scientist named Hieronymus Fabricius, laid forth the claim that there were valves in veins; however, he was

unable to explain their function to Harvey's satisfaction. In his quest to provide a more thorough explanation for the purpose of valves, Harvey began analyzing the larger question regarding how blood moves through the circulatory system. Eventually, he discovered that blood was pumped by the heart, passed through the body, and then returned to the heart, where it was recirculated. This closed system opposed the previously accepted hypothesis (Galen) that there were two types of blood: venous blood, which originated in the liver, and arterial blood, which originated in the lungs. This model predicted that blood moved through the body and was eventually consumed.

In a series of animal experiments, Harvey measured how much blood passed daily through the heart by estimating the heart's capacity (1.5 oz.), the frequency of the heart's pumping mechanism, and the amount of blood expelled with each pump (one sixth of an ounce). By his calculations, the liver would have to produce 540 lbs of blood per day if Galen was right! Harvey's findings indicated that blood flowed through the body and the heart in two loops—pulmonary circulation connected the heart to the lungs, whereas systemic circulation connected the heart to other vital organs.

Harvey also reported that veins would only allow blood to flow in one direction. He noticed that when he put a ligature on the upper arm of a patient to cut off blood flow, the area below the ligature became pale and cool, whereas the area above the ligature became swollen and warm. Once the ligature was slightly loosened, blood from the arteries flowed into the arm, causing the lower portion of the arm to begin to swell and become warmer. Veins became more visible due to the swelling, revealing the tiny valves that his teacher had discovered. Harvey was unable to push the blood back down the arm via the veins, but was easily able to push blood up the veins. Harvey theorized that veins only allowed one-way flow to the heart; these same veins used valves to prevent blood from returning to the arteries.

Possibly, Harvey's most controversial argument pertained to the transfer of blood from arteries to veins. There was no obvious way for such transfer to take place, and so Harvey theorized about the existence of tiny blood vessels, too small to be seen by the naked eye, where the transfer could occur. These hypothetical blood vessels were termed capillaries, and their existence was later demonstrated by Marcello Malpighi.

Editor Letter

Dear William,

I was fortunate enough to get three reviewers who are experts in your area. In addition, although one of the reviewers considers himself/herself to be inexperienced at mathematics while, at the same time, being an expert on blood production (Reviewer A), the others have sufficient mathematical expertise to evaluate your calculations (Reviewers B and C). Although the mathematical experts support the validity of your calculations, none of the reviewers supports publication of your manuscript. Reviewer A points out that your whole system depends on an

extremely unlikely assumption—you have to assume that there is a way for blood to be transported directly from arteries to veins, which is, at best, an unlikely assumption. To further compound the problem, nobody has been able to detect these “capillaries” to which you so glibly allude. Even you are forced to admit that you cannot directly measure them and your argument that the capillaries are too small to be detected strikes this reviewer as “fanciful.”

Reviewers B made an argument that I see as an extension of the argument made by Reviewer A; if you are allowed to postulate entities that cannot be detected, anything can be explained, which means that your theory is not falsifiable. No matter what blood is found to do, you can provide a post hoc explanation by posing the existence of entities that cannot be detected. Furthermore, as Reviewer C points out, to make your argument logical, you have to assume that entities that are too small to see can handle a heck of a lot of blood! This forces yet another ridiculous assumption that there are a zillion capillaries, and it makes the theory even more capable of post hoc explanations and further decreases its susceptibility to falsification. My own independent reading of your article leads me to agree with the reviewers and so I am unable to accept your manuscript for publication. Although your argument is extremely clever, your dependence on entities that cannot be directly measured and the general inability of your theory to make predictions that might be falsified preclude publication in a journal as competitive as ours happens to be. You might consider sending your work to a philosophical journal, where speculative arguments such as yours might be evaluated more favorably.

DECISION: Reject

The Triumph of Measurement and the Death of Phlogiston Theory

Brief Description

Why are some objects more combustible than others? Based partly on the work of Johann Joachim Becher, Georg Ernest Stahl proposed an explanation based on a substance termed *phlogiston*. When an object burns, phlogiston gets used up, and what is left over is simply the dephlogisticated substance or *calx*, which cannot burn. Support for this idea was obtained in numerous experiments in which objects were weighed prior to burning and found to be heavier than the calx that remained subsequent to burning. According to phlogiston theory, it was the removal of the phlogiston upon burning that caused the decrease in weight.

There were many problems with phlogiston theory that were either ignored or justified with auxiliary assumptions. For example, according to phlogiston theory, the rusting of metals is also due to the loss of phlogiston, but at a slower rate than the burning of wood. But metal weighs more after rusting than before it, which seems to be quite inconvenient for the theory.

As an example of how phlogiston theory was justified by additional assumptions, Rutherford found that after burning a candle in enclosed ordinary air, it became impossible to burn anything further in that air. Instead of concluding that there was a problem with the theory, he concluded that burning the candle had caused the air to become saturated with phlogiston, and so the air would no longer accept any more phlogiston, which is why nothing else would burn in it; he termed it *phlogisticated air* which we now call *nitrogen*. Similarly, Priestly found that objects burned particularly well in a gas that he created from a mercury calx. Why should this happen if burning is a function of phlogiston residing in the object itself? Like Rutherford, he refused to question the theory. Rather, he called the gas *dephlogisticated air*, and assumed that it accepted phlogiston with unusual eagerness, thereby facilitating combustion. This *dephlogisticated air* was later termed *oxygen* by Antoine-Laurent de Lavoisier, whose precise measurements and enlightened reasoning were largely responsible for the death of phlogiston theory.

Editor Letter

Dear Antoine,

I have now obtained letters from four world-class experts in phlogiston theory, which you claim to have disconfirmed. None of them finds your alternative explanations of their findings to be convincing, nor do they believe that your experiment disconfirms phlogiston theory. To clarify our criticisms, let us consider your experiments. You heated metals in closed containers of ordinary air and found that a calx formed on the surface of the metals, but only up to a point where further heating had no effect. As Reviewer A pointed out, this merely serves to further support phlogiston theory. Clearly, the air had absorbed all of the phlogiston it was capable of absorbing and was not able to absorb any more, thereby preventing the formation of any more calx.

Reviewer B took issue with the continuation of your experiment and, in particular, with your reasoning. To continue with your experiment, you weighed the calx and found that it weighed more than the metal itself had weighed, yet the weight of the whole container was unchanged from before heating. You then reasoned that if the metal had lost phlogiston, it would have decreased, rather than increased, in weight. Further, you found that when you opened the container, air rushed inside the container, thereby demonstrating that some of the air had gotten used up by becoming part of the calx and formed a partial vacuum. All of this implies, according to your manuscript, that combustion and rusting are caused by the addition of some portion of the air, rather than the loss of phlogiston, which you claim does not exist anyway. But as Reviewer B takes pains to point out, your whole superstructure depends on your ability to make precise measurements, and Reviewer B finds it unlikely that your measures really were that precise.

Reviewer A also questioned this. Both reviewers felt that your dependence on the exactness of your measures, and what they

found to be byzantine reasoning, is simply unconvincing. Finally, according to both reviewers, your conclusion that phlogiston does not exist is unacceptable, both because there is a wealth of literature indicating that phlogiston does exist, and because you cannot prove a negative, which you were foolish enough to attempt.

Reviewer C further takes issue with your explanation of combustion. Your theory cannot account for the fact that many objects lose mass after burning, which clearly supports phlogiston theory and does not support your theory unless additional assumptions are added.

Finally, Reviewer D mentioned that others also have investigated the heaviness of calx, without rejecting phlogiston theory. Therefore, the fact that you also obtained this finding contributes nothing new to the literature, nor does it force the rejection of phlogiston theory.

My own independent reading of your manuscript resulted in a reaction similar to those of the reviewers. In addition to their points, I would also like to suggest that your data provide an insufficient inductive basis for your general law of the conservation of mass. Consequently, I am forced to reject your manuscript. I know this is not the news that you wanted to hear but our journal is very competitive, and we are forced to reject the majority of manuscripts that we receive. Nevertheless, I thank you for considering our journal as an outlet for your work, and I hope that you will continue to consider us in the future.

DECISION: Reject

The Michelson–Morley Experiment

Brief Description

By the late 1800s, there was much accumulated evidence disconfirming Newton's particle theory of light and thereby supporting the notion that light is a wave. But this created a problem. Waves must have a medium through which they can travel, and it is not clear what this medium is. How can light from the stars reach us if space is a vacuum? The popular answer was that the universe is filled with such a medium, termed *luminiferous ether*. Albert Abraham Michelson and Edward Williams Morley combined their talents in an attempt to detect and measure the luminiferous ether.

They invented a device—an *interferometer*—that was able to split a beam of light into two beams that travelled at right angles to each other and were then brought back together again. The motion of the Earth through the ether should have caused an "ether wind," which would have differentially influenced the speed of the two light beams, thereby causing them to be out of phase with each other when brought back together again. However, the experiment did not work; Michelson and Morley continually failed to obtain the predicted effects. Eventually, Ernst Mach concluded that there is no ether, which later supported Planck and Einstein's theories that light behaves like

particles in some ways and thus no medium is necessary for its propagation through space. The worth of Michelson's failed experiment, along with his other work in optics, was eventually recognized, and he received a Nobel Prize in physics in 1907, the first American to achieve this.

Editor Letter

Dear Albert,

Frankly, I am not sure how to react to your work and neither is the panel of expert reviewers. As you will see, the reviewers brought up many different points that do not necessarily agree with each other. Reviewer A noted your inability to come to a solid conclusion and feels uneasy about accepting a paper like that because papers in a journal as competitive as ours are supposed to result in contributions. Given that you cannot come to a definitive conclusion, where is your contribution? Reviewer A comments further that perhaps the problem is simply that your device does not work!

Reviewer B reinforces the argument made by Reviewer A. This reviewer points out that, if your experiment is taken at face value, it would mean that there is no luminiferous ether, which means that there is no way that light could travel through space. But light does travel through space, thereby demonstrating that your experiment can not be taken at face value. The most likely possibility is that your device is simply not sensitive enough, which might explain why your effect was so small. Reviewer B suggests an additional possibility that perhaps the Earth drags the ether with it as it rotates on its axis and orbits the sun and that this is why little or no ether wind was detected in your experiment.

Reviewer C complains that it is wrong to draw conclusions from null effects. You did not perform any inferential statistical analyses, and even if you had, the null hypothesis significance testing procedure only allows you to conclude that there is a difference, not that there is not one. Reviewer C feels that there are a host of possibilities such as that your device did not work, that your measure of phase shifts was insensitive, or that you messed up the study in some other way.

Interestingly, Reviewer D disagrees with Reviewer C. Reviewer D analyzed the data you presented, using the traditional null hypothesis significance testing procedure and actually obtained a significant result. Thus, Reviewer D's analysis suggests that your experiment was a resounding success that provides a strong case for the existence of the luminiferous ether. Therefore, my suggestion is that you perform a major revision of this work, highlighting the statistical test performed by Reviewer D, and argue that your data provide strong support for the existence of the luminiferous ether. If you choose to do this, I will send it out again to all four reviewers for their comments. You have 60 days to perform the revision, and you should include a cover letter detailing all of the changes you made, as well as the changes suggested by the reviewers that you decided not to make with your reasons for not making them. Please inform me as soon

as possible about your intentions. Thank you for considering our journal as an outlet for your research.

DECISION: Major revision requested

Einstein's Special Theory of Relativity

Brief Description

The Michelson–Morley experiment left physics in a state of crisis. It was clear from Galilean relativity that the velocity of objects depended on one's frame of reference, but it also seemed clear from Maxwell's work that the speed of waves (including light) was constant relative to the medium through which the waves traveled. As light was clearly a wave, it should have a constant speed relative to the medium through which it traveled—the luminiferous ether—except that the Michelson–Morley experiment demonstrated that there was no ether! To address the mystery, Einstein proposed that light propagates in packets of energy called quanta, with the idea that light exhibits both particlelike properties and wavelike properties. Quanta of light were sufficiently particlelike to traverse the reaches of space without the help of a medium (ether).

The large principle that made it all work, however, and that unified Galilean relativity with Maxwell's work, was the way in which Einstein generalized Galilean relativity to apply to Maxwell's electro-magnetic radiation as well as to objects. Specifically, the speed of light is constant, regardless of the relative motion of the bodies involved and regardless of one's frame of reference. This idea seems to contradict common sense. For example, if a person runs at 6 miles per hour and throws a spear at 30 miles an hour in the same direction, the speed of the spear relative to an observer sitting in the bleachers would be approximately $6 + 30 = 36$ miles per hour. But if the same runner shined a flashlight, which emitted light at speed c , then the speed of the beam of light with respect to the observer would still be c , and would not be $c + 6$ miles per hour. To reiterate, the speed of light is always c , no matter what the observer's frame of reference!

The special theory of relativity had startling consequences. For example, it predicted that objects traveling near the speed of light (e.g., distant galaxies) should seem to contract. Another consequence is time dilation; time should pass more slowly for a person on a fast-moving space ship than it would for a person standing on the Earth's surface. To understand why, suppose Observer A shines a beam of light at a mirror at the other side of the vehicle in which he or she is enclosed, at distance d , and measures the time it takes for the light to traverse the distance to the mirror and make the return trip—the time the light takes to go a distance of $d + d$ (see Figure 1). In contrast, assume that Observer B is in a similar vehicle that is moving with respect to Observer A and perpendicular to a beam of light that Observer B directs at a mirror. From Observer B's perspective, the light travels the same distance as seen from Observer A's perspective,

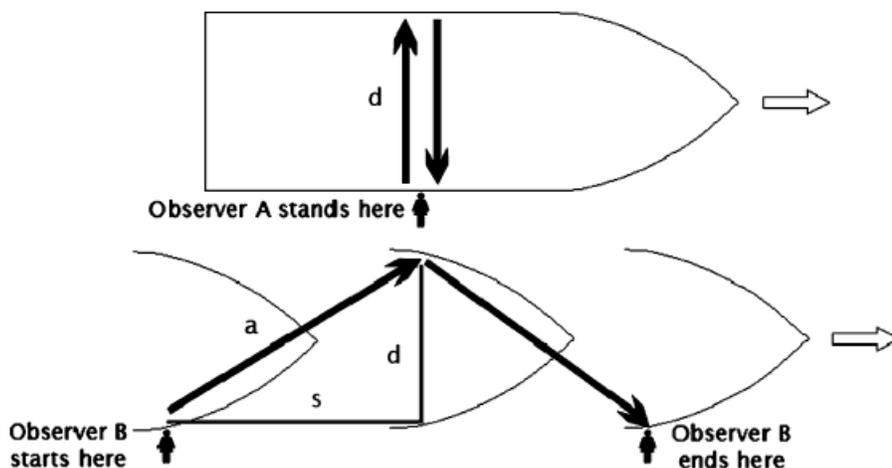


Fig. 1. An illustration of Einstein's special theory of relativity, where d is the distance across the ship, s is the distance moved, and a is the diagonal path of the light beam as it appears to Observer A. Note that according to the Pythagorean Theorem, $a = \sqrt{d^2 + s^2}$.

which, again, is $d + d$. But from Observer A's perspective, the light has to traverse two diagonal paths rather than two straight ones, and so the distance each way is longer than d , thereby rendering the total distance as greater than $d + d$. To accommodate the greater distance while still keeping the speed of light constant at c , it is necessary for time to dilate, which is precisely what Einstein predicted (and this has been empirically supported).

Mathematically inclined readers might be interested to know that the solution to the problem is a consequence of using the Pythagorean theorem to find the distance of the diagonal path. The physicist Richard Wolfson (2003) provides an easy-to-follow proof. The resulting formula for time dilation is presented below as Equation 1, where t' is the time between two events measured in a reference frame where they occur in the same place, t is the time between two events measured in a reference frame where they occur at different places, and v is the relative speed of the two reference frames, as a fraction of c (Wolfson, 2003, p. 106). Note the square root sign in Equation 1, which comes from the Pythagorean Theorem.

$$t' = t\sqrt{1 - v^2} \quad [\text{Eq.1}]$$

Editor Letter

Dear Albert,

Your manuscript provoked different reactions from the expert reviewers. Reviewer A felt that you failed to make any new predictions. This reviewer focused on your prediction pertaining to the contractions of objects moving at high speeds and noted that two other researchers, Lorentz and Fitzgerald, had both made similar independent proposals. Reviewer A concluded that if you want to make a theoretical contribution, you need to make a prediction that is different from those that others have already proposed.

Reviewer B also felt that you failed to make a new contribution. But whereas Reviewer A focused on your failure to make a new prediction, Reviewer B focused on the theory itself and noted that you simply generalized what Galileo had said about relativity without actually proposing a new principle. At best, your theory provides an incremental contribution, which is not sufficiently strong to justify publication in a journal as competitive as ours.

Reviewer C feels that your arguments are outrageous and that they fly in the face of a great deal of accumulated work—in fact, the reviewer found your arguments about time dilation to be simply insane. In less loaded words, Reviewer C thinks you are wrong and recommends rejection on that basis.

Reviewer D actually liked your work, and used words like “brilliant,” “innovative,” and even went so far as to predict a Nobel Prize for you based on this work. Unfortunately for you, because three reviewers recommended rejection and only one reviewer recommended acceptance, it is clear that the numbers are not in your favor, and I am forced to reject the manuscript. I realize that you may disagree with my decision, but a mathematician as sophisticated as you are should be able to understand that with three votes against and only one vote in favor, you lose by a net amount of two votes!

DECISION: Reject

Norman Ernest Borlaug's Green Revolution

Brief Description

Among his many contributions to Agronomy, Borlaug invented the concept of a *double wheat season*. He was originally working in Chapingo, a village east of Mexico City, where farmers were having problems with poor soil and crop rust. He then proposed adding a second growing season by transporting harvest seeds

up to the Yaqui Valley, just east of the Gulf of California. By taking advantage of the differences in altitude and temperature, he was able to grow crops twice a year, instead of just once, as usual.

Borlaug's hypothesis turned out to be true, with an additional unexpected benefit. Photoperiodism (flowering in response to changes in length of days and nights) had previously prevented wheat varieties from adapting well to alternative environments due to the changes in sunlight. However, by shuttle breeding across two different environments with a different amount of sunlight and rain, the wheat seeds became resistant to photoperiodism and were able to adapt to a variety of conditions. This finding went against all prior knowledge of wheat harvests, and it allowed new projects to be started around the world without having to rely on individual breeding programs that were tailored to specific geographic regions. Borlaug's work resulted in a huge number of people being fed who otherwise would have starved.

Editor Letter

Dear Norman,

I was fortunate to obtain reviews from four experts in your area of agriculture. Although there were some positive comments, the majority of them was negative. Starting first with the praise, Reviewer A was very much in favor of publication, mostly because he/she thought that your proposal would be of great benefit to society if it worked; also, this reviewer felt that there was little enough to lose in attempting it.

Reviewer B sang a different tune, the high note of his/her criticism being that you failed to present any basic theory and that your proposal was not based on any basic theory in the literature. Reviewer B recognizes that this is an applied journal but nevertheless maintains that even applied research should be based on solid theoretical grounds, as is the tradition with our journal. Reviewer B then goes on to point out that your proposal contradicts the well-established principle that seeds need time to revitalize themselves for germination.

Reviewer C complained that, even if you are correct in every way, your idea is simply impractical in that it requires much more work than the current agricultural method. How are you going to induce people to do all of that extra work?

Reviewer D suggested several alternative explanations for altitude and temperature effects, and I urge you to pay close attention to them (see attached reviews). Unfortunately, some of these explanations suggest that your scheme is unlikely to pan out. This reviewer believes that you need to collect more data to eliminate the alternative explanations before your proposal can be accepted. My own independent reading is consistent with Reviewers B, C, and D and also with the conservative tradition in science, which states that claims have to be well established before they can be accepted. Perhaps, after you collect the necessary data, you will consider us again as an outlet for your work.

DECISION: Reject

ANALYSES OF THE EDITOR LETTERS

There are some common threads running through the reviewers' comments that were summarized by the editor letters and will be addressed in the following subsections. There were also some less common points raised and these will also be addressed.

Plausibility

In some cases, reviewers preferred the "received" thinking, including the thinking supported by authorities, to radical new ideas. Reviewers A and D simply disagreed with Eratosthenes and found reasons to state that the notion of a spherical Earth, to which ordinary geometry could be applied, was implausible. This is also true of Reviewer D and the editor in the case of Galileo's work—they found the argument about relativity and the implication about the size of the universe to be implausible (and they also did not want to disagree with the Pope). Similarly, Reviewer A found Harvey's assertion about the existence of capillaries to be implausible, despite the fact that the opposing theory was clearly much less plausible than that, given Harvey's calculations. Similar comments apply to reviews pertaining to Lavoisier, Einstein, and Borlaug.³ In the case of Lavoisier, it is interesting that the reviewers and many scientists at the time continued to believe in phlogiston theory despite the fact that Lavoisier's experiments showed it to be implausible and contradicted by data. Perhaps scientists are too willing to dismiss new ideas as implausible and too willing to accept old ideas as plausible.

Is There Anything New?

Several of the works were criticized because they allegedly contributed nothing new, and this is possibly the most common reason for rejecting manuscripts. But the criticism can take on different forms. Reviewers A and B pointed out that the successes of the Ptolemaic view made Galileo's theory unnecessary, especially as Galileo was unable to make more precise predictions about the movements of the planets. Reviewer D suggested that all Newton did was summarize what Galileo and Kepler had done without adding anything new (and we know of 21st-century psychologists who have said this too). In a more experimental vein, Reviewer A accused Michelson and Morley of being unable to come to a solid conclusion about the existence of the luminiferous ether, thereby meaning that nothing new had been contributed. And Einstein was also accused of not contributing anything new, but for different reasons by different reviewers.

³Many people believe that Einstein's theory was instantly accepted. But as Hawking (2001) stated, it was not, and it was the target of a good deal of opposition, even from Michelson. In fact, when Einstein was awarded the Nobel Prize in 1921, the citation made no mention of relativity, which was still controversial!

According to Reviewer A, Einstein was not able to make a prediction that differed from what Lorentz and Fitzgerald had made about the contractions of objects moving at large velocities. In contrast, Reviewer B focused on the theory of relativity rather than the predictions that came out of it and concluded that the theory was merely a generalization of Galilean relativity and therefore only an incremental contribution at best. How seriously should we take these accusations?

To start with, consider that the different criticisms are at different levels. Was there a new experimental paradigm, a new finding, a new hypothesis, a new theory, or a new unifying principle? Few works make new contributions at all of these levels simultaneously, and so it is very easy to focus on one level, note that a new contribution has not been made, and not consider the other levels, thereby leading to the conclusion that nothing new was contributed when a great deal may have been contributed at other levels. All of the reviewers who made criticisms that nothing new was contributed were guilty of this; they failed to focus on the levels where the contributions were greatest. In addition, some of the reviewers were wrong even if the discussion is confined to the level at which the reviews were focused. The most obvious example is the criticism that Einstein's contraction hypothesis was nothing new given the contraction hypotheses of Lorentz and Fitzgerald. But as Einstein (1961) himself pointed out, the Lorentz and Fitzgerald contractions were ad hoc hypotheses to Maxwell's theory, whereas Einstein's contraction fell out naturally from relativity theory, rather than being an ad hoc hypothesis. And concerning the other criticism that Einstein's theory was merely a generalization of Galilean relativity, this allegedly incremental contribution simultaneously unified classical physics with Maxwell's work, resulting in predictions that were unprecedented in the history of physics. This is hardly an incremental contribution!

It is interesting to consider the issue of whether anything new was contributed in conjunction with the issue of plausibility. For example, the reviewers believed that Einstein's theory failed to contribute anything new, but Reviewer C stated that it was implausible. If Einstein was really saying nothing new, then why was Reviewer C so outraged? Clearly Einstein was contributing something new, and the editor should have seen that.

Alternative Explanations

Another common reason for rejecting manuscripts is that reviewers often suggest alternative explanations to account for the findings. But although alternative explanations may seem plausible if one only considers the point at hand, they are often implausible when more factors are considered. For example, Reviewer B suggested an alternative explanation to Eratosthenes as to why the Earth casts a curved shadow on the moon. Instead of assuming a spherical Earth, this reviewer suggested that the Earth might be a perfectly placed disk. But is this really reasonable when more factors are considered? How would Re-

viewer B account for the appearance of the shadow at different times? Or consider the alternative explanation for the "failure" of the Michelson–Morley experiment put forth by Reviewer B, who stated that the Earth drags the ether along with it. Although this seems plausible at first blush, it implies friction between astronomical bodies and the ether, in which case Newton's laws should not apply as well as they actually do. In addition, such friction further implies that astronomical bodies should slow down over time. Our point here is not that alternative explanations should be discounted but that they must be considered with care. Even if they are plausible within the context of the issue at hand, they might not be plausible when additional factors are considered.

Complexity

For both Eratosthenes and Lavoisier, at least one reviewer complained that the work was too complex. In consideration of this issue, it is useful to distinguish between at least two types of complexity. First, a theory can be complex because it has many assumptions; this kind of complexity is clearly a disadvantage. Second, however, a theory might make use of only a few assumptions with complex reasoning from those assumptions to the conclusions or predictions. This type of complexity can have merit and be unifying, as we saw in Lavoisier's work, providing that the reasoning is logically valid. If and when reviewers complain about complexity, editors should not automatically assume that this is a vicious thing. The complexity might have merit and be a reason for accepting rather than rejecting the manuscript.

Methodology and Statistics

When data are obtained that challenge the beliefs of reviewers, one of the most convincing ways of justifying rejection is to complain about the methodology. If the measures can be argued to be invalid, then the findings do not need to be taken seriously. The 25,000 mile circumference of our planet (Eratosthenes), the death of phlogiston theory (Lavoisier), and the failure to detect ether (Michelson and Morley) were all argued by various reviewers to be due to invalid measures. Reviewers suggested that perhaps the distance between the sticks was not measured correctly, that Lavoisier's measurements were not really as precise as they needed to be, or that Michelson and Morley's interferometer did not work correctly. In one sense, the reviewers are clearly correct; there is no way to be absolutely sure that the measures were as valid as they needed to be. But it is always possible to argue this, even if validation studies have been performed, and so the argument itself is not sufficient reason to justify the rejection of manuscripts. Do the reviewers have good reasons to believe that Eratosthenes got the distance wrong, that Lavoisier's measurements were imprecise, or that Michelson and Morley's interferometer did not work? If not, the

mere assertion that something might have been amiss is surely insufficient.

The measurement issue is often integrated with a statistical issue, and this is best seen by considering the example of Michelson and Morley's research more deeply. Specifically, if an obtained result is not statistically significant, then reviewers often take this as evidence that the measure was not sufficiently sensitive (or that the manipulation was too weak). After all, had the measure been sufficiently sensitive, a statistically significant result might have been obtained. But if a statistically significant effect is obtained, then the measure was clearly sufficiently sensitive to have obtained it.

With this in mind, consider the comments made by Reviewers C and D. Reviewer C complained that Michelson and Morley failed to conduct a significance test, whereas Reviewer D actually performed the test and obtained statistical significance. Consequently, Reviewer C concluded that there was no effect, and Reviewer D concluded that there was one. It is interesting to note that Carver (1993) went back to the Michelson–Morley experiment, actually performed this statistical test, and obtained a statistically significant finding! Had physicists bought into behavioral science methodology, they would have come to the same conclusion that Reviewer D came to, with deleterious consequences for the development of physics in the following years. The Carver reanalysis of the Michelson–Morley data suggests that there is something horribly wrong with the null hypothesis significance testing procedure that dominates the behavioral sciences (but see Sawilowsky, 2003, for a qualification). This has been discussed in detail by several researchers, but these writings have been given insufficient attention (e.g., Bakan, 1966; Cohen, 1994; Rozeboom, 1960; Schmidt, 1966; Schmidt & Hunter, 1997; Trafimow, 2003, 2005; see Trafimow, 2006b, for a review).

Falsification

Psychological theories are often accused of being incapable of falsification (see Trafimow, *in press*, for a recent review), and this criticism was also applied to Harvey's work. In particular, all of the reviewers and the editor felt that the assumption of unseen blood vessels (capillaries) allowed Harvey's theory to account for any conceivable set of findings. However, a close look at the notion of falsification from a philosophy of science perspective undercuts this criticism.

Philosophers have long been aware that it is impossible to prove theories to be true on the basis of predictions confirmed by experiment because the predictions might have worked out for another reason; such reasoning is an example of the logical fallacy known as affirming the consequent. In contrast, if a prediction fails, then one can, by the valid logic of *Modus Tollens*, conclude that the theory is wrong. Therefore, philosophers, most notably Popper (1934/1959), have recommended that scientists should attempt to disconfirm theories rather than

confirm them, which implies that the theories must be capable of falsification, or else the whole enterprise is doomed from the start.

But matters are not this simple, as Lakatos (1978) pointed out. This is because predictions are not made solely on the basis of a theory, but on the basis of a theory combined with auxiliary assumptions. If a prediction fails, it is not necessarily the theory that is at fault; an auxiliary assumption might be to blame. Thus, it becomes clear immediately that absolute falsification is as impossible as absolute verification. Therefore, if no theories are falsifiable, then this is hardly a reasonable criterion for evaluating a theory.

A possible way out for the reviewers might be to not insist on absolute falsification as an evaluative criterion. Rather, one could argue that theories only need to be capable of falsification within the limits of the auxiliary assumptions used to derive the predictions. In this case, many theories are falsifiable, and it is no longer unreasonable to hold Harvey to that standard. But this introduces another problem; the falsification criterion now fails to exclude any theories. To see why this is so, consider again that it is the addition of auxiliary assumptions to the theory at hand that allows researchers to derive predictions from theories. So how can one know whether a theory is, in principle, falsifiable? The answer would be to combine it with auxiliary assumptions and derive predictions. If at least one set of auxiliary assumptions results in a testable prediction, then the theory is falsifiable. But what if the reviewers cannot think of any sets of auxiliary assumptions that result in at least one testable prediction? Does that mean the theory is not falsifiable? Clearly, this is not so as some other person might be able to do it. Unless one has tried to combine every one of an infinite set of auxiliary assumptions with the theory at hand and has always failed to derive a testable prediction, that person cannot justifiably denounce the theory as being incapable of falsification. As demonstrations of the invalidity of subjective judgments of theories not being falsifiable, consider that the use of creative auxiliary assumptions has resulted in the falsification of aspects of psychoanalytic theory (Freud & Breuer, 1895) and the theory of reasoned action (Fishbein & Ajzen, 1975), which have been deemed by numerous researchers to be incapable of falsification (Trafimow, Brown, Grace, Thompson, & Sheeran, 2002; Trafimow & Sheeran, 1998; see Trafimow, *in press*, for a review).

There is another way in which auxiliary assumptions are overwhelmingly important. In the behavioral sciences, highly mathematical theories are sometimes criticized as being too abstract and having insufficient descriptive richness. For example, Reviewer A criticized Newton's theory on this basis—a criticism that a careful consideration of the role of auxiliary assumptions negates. Although theories explain how constructs relate to each other, it is auxiliary assumptions that make it possible to instantiate specific information into the slots provided by the theory. In Newton's theory, for example, information about the planets is not mentioned, yet Newton made very clear

predictions about planetary motions. This was done by adding auxiliary assumptions about the present positions and velocities of the planets rather than deriving them from Newton's laws themselves. If it is up to auxiliary assumptions and not theories to provide descriptive richness, then theories should not be rejected because of arguments about descriptive richness.

Circular Reasoning

Reviewer C noted that Newton failed to provide independent definitions of force, mass, and acceleration, thus inducing circularity into Newton's theory and justifying the recommendation to reject. But is circular reasoning really so vicious? To see why it might not be, suppose that an independent definition of mass is desired and that Newton could supply it. That independent definition would have to include at least one theoretical term, call it T_1 , but we then would require a definition of that term. Newton could then define T_1 in terms of T_2 , but then a definition of T_2 would be required, and so on, ad infinitum. Thus, Reviewer C's criticism, which seems reasonable on the surface, actually requires Newton to perform the impossible as a requirement for publication. Put more generally, any set of theoretical terms forces one to have primitive terms that are not defined (e.g., mass), circularity, or infinite regression. Therefore, the mere fact of these provides an insufficient justification for rejection unless there is a further issue.

Applied Work

In addition to the foregoing criticisms, applied work is often subject to additional criticisms such as a lack of theory and a reviewer's accusation that "it will never work in the real world." We saw these criticisms in the letter evaluating work by Borlaug. It seems almost like common sense that if a discovery can help a lot of people, then it is worth making even if there is no brilliant accompanying theory, but common sense sometimes goes out the window when applied research is evaluated by reviewers. We are more sympathetic to the criticism that the research would not work in the real world, but even with this criticism we would like to underscore two potential problems. One problem is that, at least in the case of Reviewer C's criticism of Borlaug, the reviewer failed to explain why Borlaug's idea would not work. Nor did the reviewer provide any evidence that it would not work. Moreover, even if Reviewer C had been correct in arguing that there were technical problems that would prevent Borlaug's methods from working, someone else—perhaps someone not even in the agricultural area—might have been able to find a way to circumvent the technical difficulties. For example, if Reviewer C felt that people would not be willing to put forth the work to use Borlaug's method, perhaps a social psychologist could have found a way to motivate them (fortunately, this turned out not to be necessary).

Of course, even in applied work, one can apply the usual criticisms. For example, Reviewer D suggested an alternative

explanation for altitude and temperature effects that the editor did not actually describe, but to which Borlaug was referred. This saved the editor from having to deal with whether the alternatives really were sufficiently compelling so as to provide a strong reason for rejecting the manuscript. It also allowed the editor to make the familiar decision based on the votes of the reviewers, coupled with a weak justification based on the conservative nature of science and the usual call for more data. To us, it seems questionable whether the conservative nature of science is always a good thing. When a proposal is made that has the potential to save many lives, it is perhaps justifiable to be less conservative, particularly if the downside risks are low.

What Should Be Valued?

Several of the reviewers' comments are based on stated or unstated value judgments. For example, Reviewer A points out the inability of Michelson and Morley to come to a definitive conclusion, a standard basis in the behavioral sciences for rejecting articles. But sometimes it is more important to illuminate an important puzzle than to solve a minor one.

Or consider that Reviewer C criticized the lack of data in support of Galileo's relativity principle. This argument is quite consistent with the emphasis that behavioral scientists place on data. But should Galileo really be required to have data to back up his theory? Is it not sufficient to propose a brilliant theory with the possibility that others can find the relevant data? Arguably, there is an overemphasis on data in the behavioral sciences. Even in journals that specialize in theoretical articles, much space is devoted to data rather than to theory. And even in such journals, reviewers often argue that theories are not ready for publication—not because there is a problem with the theories, but rather because a literature has not yet been built up to properly support them. One might reasonably argue that if a strong empirical record is necessary to justify the publication of a theoretical article, then this greatly discourages the publication of new ideas for which an empirical base has yet to be established.

An additional value judgment is made salient when one considers that the rejection of Galileo's work was strongly flavored by political correctness. The politically correct belief, at the time, was that the Earth was at the center of the universe. Obviously, Galileo went against political correctness and was punished for it. How many behavioral science theories that deal with currently hot issues such as prejudice, political action, and others are accepted because they are consistent with what academics want to believe rather than because they are actually good theories?

DISCUSSION

As the foregoing cases illustrate, there are many reasons for rejecting articles (see Table 1 for a summary). Some of these

TABLE 1
Reasons for Rejection

Reasons for rejection	Eratosthenes	Ptolemaeus	Galileo	Newton	Harvey	Lavoisier	Michelson	Einstein	Borlaug
Disagreement with authority	X		X			X			
Reviewer disagrees with theory	X		X			X	X	X	X
Cherry picking the data	X								
Alternative explanations	X					X	X		X
Reasoning too complex	X					X			
Bad methodology	X					X	X		
Data can be explained by previous theory			X			X			
Not enough data			X	X	X				
Too abstract/Needs more description				X					
Need definitions				X					
Nothing new				X		X		X	
Lack of direct evidence					X				
Not falsifiable					X				
Attempts to prove a negative						X			
No solid conclusion							X		
Too applied – Not theoretical enough									X
It will not work									X
Outvoted	X		X		X	X		X	X

reasons are not justified, in general, whereas others may be valid at times, if used judiciously. Rather than discuss them in more detail, we proceed directly to a discussion of the process of evaluation.

Clearly, reviewers use combinations of mental processes, and it is unlikely that we will be able to capture all of them here, particularly without the help of any formal research on it. Nevertheless, there are some processes that are familiar to us all that merit some brief discussion. For example, in the “Gatekeeper” process, reviewers perceive themselves as guards who prevent anything flawed from getting in and poisoning the field. Clearly, to some extent, this is appropriate, but as we have seen, it can be taken too far. An example of this is when reviewers note every case where an author breaks some sort of “rule,” duly notes it in his or her review, and then uses the violations as the reason for recommending rejection. The foregoing case studies provide numerous examples of this, but there are many more that did not come up. For example, many papers get rejected because the author failed to provide a “manipulation check.” A cursory examination of this argument makes it seem like a good one. Imagine that a researcher manipulates attitude and obtains an effect on behavior. A reviewer notes that the researcher failed to include a manipulation check and asks how the researcher can be sure that the manipulation really influenced attitude. Perhaps the manipulation worked for a different reason that could be eliminated if a manipulation check were included.

But let us consider this issue more deeply. Suppose the researcher had included a manipulation check and found, in fact, that the manipulation influenced responses on the attitude measure that was now included. Would this really demonstrate that the reason the manipulation influenced behavior is because

of attitude? Clearly, the answer is negative. It is entirely possible that the manipulation influenced attitude and another variable, and it is the other variable, rather than attitude, that caused the effect on behavior. In addition, even if this point were disregarded, how can the reviewer be sure that the measure used for the manipulation check was valid? Does this measure need to be compared with another measure, which itself might not be valid, thereby necessitating comparison to yet another measure? Should the researcher be required to build a nomological network around the manipulation before being allowed to publish? Furthermore, including a manipulation check can disturb the experiment. If the manipulation check takes place before the main dependent measure, then it might interfere with the effect of the manipulation. Or, even if it does not, one then could argue that the manipulation check is a necessary condition for the effect to be observed. And if the manipulation check takes place after the main dependent measure, then it is possible that the effect of the manipulation dissipated by the time the manipulation check took place. Our specific point here is not that manipulation checks should never be performed, only that they often should not be performed and that the reviewer who recommends one should be very careful to make sure that the benefits really would outweigh the costs (see Kidd, 1976; Perdue & Summers, 1986; Sawyer, Lynch, & Brinberg, 1995, for fuller discussions). For example, a manipulation check might be valuable to explore a possible reason for failure if the experiment does not work out. Our more general point is that it is often a good idea to break the rules, and reviewers should not penalize researchers for doing so without a good reason.

Some reviewers also consider how strongly the data support the proposed theory. Clearly, to the extent that it is easy to think

of alternative explanations for the findings, they provide a less convincing case for the theory. Although few would argue that alternative explanations are unimportant, we hope that the foregoing case studies demonstrate that their mere presence does not justify rejection. One should evaluate the quality of the alternative explanations. Are they really as plausible, or more plausible, than the proposed explanation? Are they as parsimonious as the proposed explanation? How far did the reviewers have to stretch for the alternative explanations? Are there data in the literature or in the research at hand that contradict the alternative explanations?

Sometimes researchers come up with reasons other than alternative explanations to argue that the data do not provide strong support for the theory, and this is often the case where intervening variables are concerned. Suppose an author argues that manipulation X causes intervening variable I , which in turn causes dependent variable Y . As support, the researcher manipulates X and obtains an effect on Y but neglects to measure I . Does this failure to measure I justify rejection? We believe it depends on the precise nature of the research and whether the reviewer is able to generate a plausible alternative explanation. Suppose that the researcher had measured I and found that both I and Y were influenced by X , how much would this increase support for the theory? Possibly, the increased support would be trivial; for example, it could be that X influenced I and Y for different reasons. A reviewer could argue for some kind of path analysis to test whether X really influenced Y through I , but correlation does not provide a strong case for causation and neither do the multiple correlations upon which path analyses are based. In fact, Trafimow (2006a) recently demonstrated that although path analyses can address random invalidity (unreliability), they cannot satisfactorily handle nonrandom invalidity, which virtually guarantees that, even when there is no mediation, it will nevertheless be found, provided that the sample size is sufficiently large. Our point is not that reviewers should never require measures of hypothesized intervening variables. Rather our point is that if reviewers are going to do this, they should be required to explain precisely what alternative explanation the measure is going to eliminate. And if they recommend a path analysis, they should again be required to explain precisely what alternative explanation the path analysis is going to eliminate. Reviewers should not recommend extra measures or extra analyses just because they are fashionable in the field, they are routinely performed, or because the reviewers themselves do it in their own research.

In addition to gatekeeping and evaluating the relations between theory and data, some reviewers try to evaluate whether the idea is important. Obviously, this is highly necessary though subjective; what is important to one person might not be important to another. We can only urge reviewers to be careful and to try to consciously evaluate all parts of the contribution, rather than the most salient part or the part that is most relevant to the reviewers' interests. We hope that reviewers will not simply decide that the author is not making a contribution on the basis

of disagreements between the reviewers and authors. Dissenting views should be encouraged rather than discouraged, and perhaps reviewers should ask themselves not whether the author is right or not (because the answer will be "no" as long as the reviewer has a different opinion!), but whether the idea deserves to be aired. If the author's idea would be important—if it were right—then it might provide a strong reason for airing it even if the reviewer thinks it is unlikely to be true. Editors could aid in this process by soliciting reviews from researchers who do not have a stake in the issue or by being more willing to overrule reviewers who do have a stake in the issue. We recognize that there is likely to be a correlation between having a stake in the issue and being expert in the issue, but it might be worth sacrificing some expertise to increase open-mindedness and tolerance on the part of the reviewers.

Lastly, reviewers often evaluate the submitted research in terms of its relation to previous literature. They consider submitted research to be "good" when it has a strong connection to previous literature or "bad" when it is not well connected. It is not surprising that reviewers feel good when submitted research has clear connections to previous literature. This increases the likelihood that the reviewers will be cited! In addition, even if the reviewer does not have a personal stake in the submitted research, he or she gains a nice feeling of linear progress if the submitted research is strongly connected to previous literature. Furthermore, demonstrated connectedness with previous literature assures reviewers that the author is well read. Unfortunately, insisting that submitted research be strongly connected with previous literature can cause reviewers to reject new ways of thinking. Sometimes, when an idea is really novel, part of what makes it novel is precisely the fact that it does not depend to any great extent on the ideas of previous researchers. Incremental progress from article to article can be considered a desirable thing to have, but truly revolutionary ideas are more exciting and psychology might benefit if behavioral scientists were to become less insistent that everything be connected to previous literature.

Given that there are many ways of failing to appreciate the worth of great works, some of which stem from particular evaluative processes employed by behavioral scientists, how can reviewers make sure to recognize great works when they are sufficiently fortunate to encounter them? We have no definitive answer, but we can suggest some potential warning signs, some of which were mentioned earlier in passing. Most obviously, if the theory seems unlikely to be true, it might be a crackpot idea but it also might be brilliant, as brilliant theories often seem unlikely to be true. Relatedly, if the theory makes predictions that seem ridiculous, the ridiculous predictions may be evidence in favor of the creativity of the theory rather than evidence that the theory should never see the light of day. Also, if the reviewer has a difficult time getting his or her mind around the concept, then perhaps the author has a truly innovative or revolutionary perspective. An additional warning sign is if the proposed theory causes the established literature to be seen in a

different light. Rather than assuming that the author has merely put “old wine in new bottles,” the reviewer should consider that there really is something basic that is new, or else it would not imply a reinterpretation of the known literature. A further warning sign is if the theory seems to have consequences that are in opposition to the literature. It could be that the author is not sufficiently acquainted with the literature, but it also could mean that the author is on an inspired new track and that the contradicted literature is either wrong or wrongly interpreted. Finally, if there is a lack of an empirical track record in support of a theory, a reviewer should consider that the theory truly is novel and deserves to be aired rather than interpreting this as a reason to insist on more data. Perhaps a general question reviewers should ask themselves is whether the theory, if it were true, would be important. If so, then perhaps the reviewer’s opinion that the theory is unlikely to be true should be deemphasized.

In conclusion, several great works of the past were presented and criticized. In turn, the criticisms themselves were analyzed and found to be wanting. Several more ways of criticizing submitted manuscripts were also reviewed in the Discussion section, and these were also found to be problematic. It might be tempting to interpret the present arguments as reasons to be less critical or to accept more papers. On the contrary, our position is that researchers should be more critical and make special efforts to pick out the papers that have the possibility of profoundly changing the field (and perhaps accept slightly fewer “business as usual” papers to make room). Hopefully, the present examples from other sciences have dramatized that critical thinking includes criticizing the potential criticisms, determining if they really are justified, and considering that they might not be. Another aspect of critical thinking is to evaluate the positive characteristics of the submission and consider that these might outweigh the negative ones, even if there are clear negatives to complain about. Yet a further aspect of critical thinking is to determine whether the authors could do something to improve the paper, even if the current evaluation is negative. Above all, do not be the next person to squelch a potentially great work because of ill-considered criticisms, even if the criticisms are standard in the field.

REFERENCES

- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, *66*, 423–437.
- Carver, R.P. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education*, *61*, 287–292.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997–1003.
- Einstein, A. (1961). *Relativity: The special and the general theory* (Robert W. Lawson, Trans.). New York: Crown Publishers.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Freud, S., & Breuer, J. (1895). *Studies on hysteria (Studien über Hysterie)* (N. Luckhurst, Trans.). New York: Penguin Group.
- Hawking, S. (2001). *The universe in a nutshell*. New York: Bantam Books.
- Kidd, R.F. (1976). Manipulation checks: Advantage or disadvantage. *Representative Research in Social Psychology*, *7*, 160–165.
- Lakatos, I. (1978). *The methodology of scientific research programmes: Philosophical papers, Volume 1* (J. Worrall & G. Currie, Eds.). Cambridge, United Kingdom, UK: Cambridge University Press.
- Perdue, B.C., & Summers, J.O. (1986). Checking the success of manipulations in marketing experiments. *Journal of Marketing Research*, *23*, 317–326.
- Popper, K. (1959). *Logik der Forschung* [The logic of scientific discovery]. New York: Basic Books. (Original work published 1934)
- Rozeboom, W.W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin*, *57*, 416–428.
- Sawilowsky, S. (2003). Deconstructing arguments from the case against hypothesis testing. *Journal of Modern Applied Statistical Methods*, *2*, 467–474.
- Sawyer, A.G., Lynch, J.G., & Brinberg, D.L. (1995). A Bayesian analysis of the information value of manipulation and confounding checks in theory tests. *Journal of Consumer Research*, *21*, 581–595.
- Schmidt, F.L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods*, *1*, 115–129.
- Schmidt, F.L., & Hunter, J.E. (1997). Eight objections to the discontinuation of significance testing in the analysis of research data. In L. Harlow, S.A. Mulaik, & J.H. Steiger (Eds.), *What if there were no significance tests?* (pp. 37–64). Mahwah, NJ: Erlbaum.
- Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries: Surprising insights from Bayes’s theorem. *Psychological Review*, *110*, 526–535.
- Trafimow, D. (2005). The ubiquitous Laplacian assumption: Reply to Lee and Wagenmakers. *Psychological Review*, *112*, 669–674.
- Trafimow, D. (2006a). Multiplicative invalidity and its application to complex correlational models. *Genetic, Social, and General Psychology Monographs*, *132*, 215–239.
- Trafimow, D. (2006b). Using epistemic ratios to evaluate hypotheses: An imprecision penalty for imprecise hypotheses. *Genetic, Social, and General Psychology Monographs*, *132*, 431–462.
- Trafimow, D. (in press). The theory of reasoned action: A case study of falsification in psychology. *Theory & Psychology*.
- Trafimow, D., Brown, J., Grace, K., Thompson, L., & Sheeran, P. (2002). The relative influence of attitudes and subjective norms from childhood to adolescence: Between-participants and within-participants analyses. *American Journal of Psychology*, *115*, 395–414.
- Trafimow, D., & Sheeran, P. (1998). Some tests of the distinction between cognitive and affective beliefs. *Journal of Experimental Social Psychology*, *34*, 378–397.
- Wolfson, R. (2003). *Simply Einstein: Relativity demystified*. New York: W.W. Norton.