crítica

Universidad Nacional Autónoma de México

Instrumentalism Revisited
Author(s): Elliott Sober
Source: *Crítica: Revista Hispanoamericana de Filosofía*, Vol. 31, No. 91 (Apr., 1999), pp. 3-39
Published by: Instituto de Investigaciones Filosóficas, Universidad Nacional Autónoma de México
Stable URL: http://www.jstor.org/stable/40104484
Accessed: 09/10/2013 12:22

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Instituto de Investigaciones Filosóficas and *Universidad Nacional Autónoma de México* are collaborating with JSTOR to digitize, preserve and extend access to *Crítica: Revista Hispanoamericana de Filosofía.*

http://www.jstor.org

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

ARTÍCULOS

CRÍTICA, *Revista Hispanoamericana de Filosofía*
Vol. XXXI, No. 91 (abril 1999): 3–39

INSTRUMENTALISM REVISITED*

ELLIOTT SOBER
Department of Philosophy
University of Wisconsin

Like many of the good ideas that the logical empiricists had, instrumentalism did not receive the attention it deserves. In part this was because of what I'll call *the lightning rod effect*. Here's how you can witness this phenomenon in the comfort of your own home (children should not do this without adult supervision):

> Put a good idea next to a bad one. Someone will then refute the bad idea. Then people will think that the good idea as well as the bad one have both been demolished.

You don't always get the lightning rod effect when you follow these instructions, but it occurs often enough that it deserves a name.

* I am grateful to Martin Barrett, Ellery Eells, Branden Fitelson, Malcom Forster, Ilkka Kieseppä, Theo Kuipers, Greg Mougin, Diana Raffman, Larry Shapiro, Chris Stephens, Neil Tennant, and Mark Wilson for useful discussion and to Branden Fitelson and Tina Eliassa-Rad for doing the computer simulations that allowed them to prepare the figure. A short version of this paper will appear in the *Proceedings of the 1998 World Congress of Philosophy*.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

The logical empiricists said some good things about epistemology and scientific method. However, they associated those epistemological ideas with some rather less good ideas about philosophy of language. There *is* something epistemologically suspect about statements that cannot be tested. But to say that those statements are *meaningless* is to go too far. And there *is* something impossible about trying to figure out which of two empirically equivalent theories is true. But to say that those theories are *synonymous* is also to go too far.

My goal in this paper is not to resuscitate all these positivist ideas, but to revisit just one of them. Instrumentalism is the idea that theories are instruments for making predictions. Of course, no one would disagree that this is *one* of the things we use theories to do. In just the same way, no one could disagree with the emotivist claim that *one* of the things we do with ethical terms like "good" and "right" is to express our feelings of approval and disapproval. Instrumentalism and emotivism become contentious, and therefore interesting, when these claims are supplemented.

The most familiar sort of supplementation that they received is semantic. Emotivism claimed that ethical statements are neither true nor false; instrumentalism said the same thing about scientific theories. This negative semantic claim is what is usually intended when it is said that ethical statements "merely" express feelings of approval and disapproval and that scientific theories are "merely" instruments for making predictions. I don't propose to go down that well-trodden path again. Setting ethics to one side, I am happy to assume that scientific theories have truth values. Arguments to the contrary rest on bad theories of meaning. Theoretical terms have meanings that transcend what can be stated in some proprietary observation language (even assuming, controversially, that there is such a thing as an observation language); the fact that the meanings of terms

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

like "mass" and "charge" are not exhausted by statements about meter readings doesn't show that statements about mass and charge are neither true nor false.

If instrumentalism as I construe it does not deny that theories have truth values, what is there left to discuss? The claim that theories are instruments for making predictions now seems to be the boring point that this is one of the many things that scientists do with theories; scientists also use theories to explain regularities, to impress their colleagues, and for other purposes as well. Can instrumentalism do anything better than teeter between falsehood and triviality? I think it can. The important point of instrumentalism is methodological, not semantic. It is the idea that *theories are to be judged by their ability to predict*. Of course, if this just means that predictive accuracy is *one* of the criteria that scientists use and ought to use for evaluating theories, then we again have a truism. But instrumentalism goes further. The claim is that predictive accuracy is the *only* consideration that matters in the end —it is the unique ultimate goal that scientists bring to bear in evaluating theories. Instrumentalism does not deny that theories are and ought to be judged by their simplicity, their ability to unify disparate phenomena, and so on. However, instrumentalism regards these considerations as relevant only in so far as they reflect on a theory's predictive accuracy. If two theories are predictively equivalent, then a difference in simplicity or unification makes no difference, as far as instrumentalism is concerned. Simplicity, unification, even fitting the data at hand are simply means to the end of securing predictive accuracy.

I won't try to show in the present essay how simplicity, unification, and other *desiderata* are related to predictive accuracy, although that's an important task for a proper defense of instrumentalism to undertake (see Forster and Sober 1994 for discussion). What I do want to consider

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

is how instrumentalism, thus construed, is related to a quite different proposal concerning how theories should be evaluated. Scientific realism has many formulations; the one of interest here says that the goal of science is to find theories that are true (Van Fraassen 1980). Understood in this way, realism does not say that our present theories are true, nor does it offer a substantive account of what truth is. It specifies an end; scientific theories are to be judged by how well they manage to attain that end.

How are the goals of truth and predictive accuracy related? Is it possible to choose between instrumentalism and realism as accounts of scientific practice? In *The Structure of Science*, Ernest Nagel (1979, p. 139) says that the difference between instrumentalism and realism is merely "verbal". In any concrete situation in which scientists have to make a decision, the two philosophies make the same recommendations. We'd expect scientists to be doing the same things, regardless of whether their ultimate goal is to find true theories or to find theories that are predictively accurate. Truth and predictive accuracy seem to go hand-in-hand. Nagel's contention is an important one, if we are to determine whether anything interesting can be salvaged from instrumentalism.

Just as instrumentalism as a methodological claim needs to be separated from instrumentalism as a semantic thesis, we also need to separate the methodological idea from what I'll term the realm of the *personal*. Scientists are people and different people have different goals and the same person may have many goals. To say that successful prediction is *the* goal of science thus seems to deny the multiplicity of goals that exists within and among scientists (Putnam 1975, pp. 233–234). My reply is that instrumentalism is not a claim about the goals of individual *scientists* or of scientific *institutions* so much as it is a claim about the goals of *scientific inference*. This requires clarification. Inference

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

rules are abstract entities, like numbers and functions. Apparently, inference rules have "goals" no more than the number 17 has a goal.

What I mean is that when scientists distinguish a "good" inference from a "bad" one, they do so, whether explicitly or implicitly, by assuming a set of desiderata that the inference should satisfy. Their evaluative comment is a claim about whether the inference at hand fills the bill. Rules of inference are tools whose justification depends on one's aims (Hempel 1979, Forster 2000). *Modus ponens* is a good rule to follow if you want your inference to be truth-preserving. But if your goal is to draw conclusions that go beyond what the premisses assert, then *Modus ponens* is not what you should use —you should use an ampliative mode of inference instead.

This means that the debate in philosophy of science about realism and instrumentalism could be conducted in two ways. One could argue about whether scientists *should* seek true theories or theories that are predictively accurate. Or, one could argue about what scientists in fact are getting at when they distinguish good inference from bad. It is the latter question that interests me; I am not going to discuss what the categorical imperatives are to which scientists should swear allegiance. Rather, the task at hand is to examine the inferential practices at work in science and try to infer from them what their goal is. This is something like what evolutionary biologists do when they observe a phenotypic trait and try to guess what function the trait subserves. In both cases, one is engaged in a project of "reverse engineering". As the analogy with biology suggests, our question is not automatically settled by asking scientists what they think they are trying to achieve. Scientists may or may not be able to describe the rules they follow; and even if they can, they may or may not have anything useful to say about what the desiderata are that

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

"good" inferences typically satisfy and that "bad" inferences typically do not.

The gloss I have given of the dispute between realism and instrumentalism differs somewhat from an interpretation one frequently hears. The more usual idea is that realism and instrumentalism are claims about normative methodology, not about descriptive psychology or sociology. The question, so it is said, is not about what scientists actually think, but what they ought to think. I do not reject the idea that philosophers can make normative recommendations about science; after all, *anyone*, *even a philosopher*, can make a normative suggestion, which should be judged on its merits, not on the basis of the union card that the recommender happens to carry. Rather, the point is that the normative suggestions made in philosophy of science are different in character from those made in ethics. Ethics is in the business of discussing categorical imperatives. I don't see much of a role for this in philosophy of science. Philosophers of science, like statisticians, advance hypothetical imperatives —"if your goals are such-and-such, then X is a practice you should embrace while Y is a practice you should eschew". Formulating hypothetical imperatives is part of the project of reverse engineering; this project is descriptive and normative at the same time.

Let us return, then, to Nagel's very good question about the seeming equivalence of instrumentalism and scientific realism. The problem he posed is illustrated in the following two-by-two table:

		Believed to be	
		More predictively accurate	Less predictively accurate
Believed to be	True	H1	
	False		H2

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

If scientists have to choose between hypotheses H1 and H2, what will they do? If they think that H1 is true and that H2 is false, and that H1 will be more predictively accurate than H2, they presumably will choose H1 over H2. However, this doesn't tell us whether they are really after theories that are predictively successful, or theories that are true, or both. When the competing hypotheses fall on the main diagonal of this table, the two factors "truth" and "predictive accuracy" are hopelessly confounded.[1]

To test realism against instrumentalism as claims about the goals of science, we should try to find a situation in which the *desiderata* of truth and predictive accuracy come into conflict. What would be nice is a pair of hypotheses H3 and H4 that are taken to have these properties:

		Believed to be	
		More predictively accurate	Less predictively accurate
Believed to be	True		H3
	False	H4	

If scientists prefer H3 over H4, this result favors scientific realism —they are opting for the theory they think is true, even though they think it will be less predictively accurate than its competitor. On the other hand, if they prefer H4 over H3, this result favors instrumentalism, since it shows that scientists are prepared to sacrifice truth to improve predictive success.

Could there be an anti-diagonal case of this sort in which truth and predictive accuracy come into conflict?[2] Let's

[1] The use of two-by-two tables to represent the interaction of two desires is borrowed from the treatment of psychological egoism and altruism in Sober and Wilson (1998), Chapter 7.

[2] It's easy to invent situations in which truth and predictive accu-

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

begin by examining more carefully an argument that seems to support Nagel's contention that this is impossible. The argument in question is *the sure thing argument*:

> Suppose that T_1 is true and that T_2 is false. Then you can deduce only true predictions from T_1; however, if you use T_2 to make predictions, it is an open question whether the predictions you deduce from it will be true. Hence, if your goal is true prediction, you should use T_1 rather than T_2; T_1 is a "sure thing".

This argument has two limitations. It construes prediction as deduction; yet, many predictions involve nondeductive inferences. In addition, the argument describes successful prediction in terms of the dichotomous concept of *truth*, rather than in terms of the matter-of-degree concept of *predictive accuracy*.

It is possible to supplement the sure thing argument with an argument that has neither limitation. The new argument still isn't perfectly general, but it is worth seeing that the

racy "conflict", but not in a sense that is relevant to the disagreement between realism and instrumentalism concerning what acceptance involves. Suppose H3 is a true hypothesis about geology and H4 is a false (though predictively accurate) hypothesis about economics. If one wants to make predictions about economics, it is no surprise that one should use H4 rather than H3. This, however, is no argument in favor of instrumentalism, since H3 and H4 are not *competing* hypotheses. They don't make contrary predictions, and so we are not forced to decide which predictions to trust. A similar point pertains to an idea that Mark Wilson (forthcoming) has emphasized —that it often happens that theories in mathematical physics that are regarded as true are mathematically intractable, and so are not very useful as devices for making predictions. To make predictions, applied mathematicians opportunistically employ various tricks, shortcuts, and idealizations. Here, the true theory "makes predictions" in the sense of entailing them; the trouble is that human beings can't figure out what those predictions are. So once again, there is no practical problem of being faced with contrary predictions.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

thought behind the sure thing argument isn't completely tied to the two limitations just noted. Suppose your goal is to estimate the mean value (μ) that a continuous trait has in a population, assuming that the trait is normally distributed with variance σ^2. For example, you might want to say what the average height is in a population of corn plants in which height has the familiar bell-shaped distribution. You construct an estimate (θ) of the population mean in some way, perhaps by examining some of the individuals in the population. How accurately will θ predict new data that you draw from the same population? The predictive inaccuracy of θ is defined as its average distance from the mean values that will be found in new samples. Suppose you sample ten corn plants, determine that their average height is s_1, return the sample to the population, then sample ten individuals again, noting that their average height is s_2, and so on. The sample averages s_1, s_2, . . . will differ. Although an accurate estimate of the population mean may occasionally be quite distant from a given sample mean, on average it will be quite close.[3]

It is no surprise that the most predictively accurate estimate is the population mean itself. If you use μ to predict the sample averages, you won't achieve perfect predictive accuracy, but you'll do better than if you use any other value. This intuitive idea is vouchsafed by the following fact. Suppose your value for θ is obtained by sampling a number of individuals in an initial sample, whose mean is s_0; if you use s_0 as your value of θ and you then use this value of θ to predict the means that will be found in new samples, then the predictive accuracy of θ is defined as follows:

[3] The more general definition of the predictive accuracy of an estimate θ, $A(\theta)$, is its expected log-likelihood; $A(\theta) = E(\text{log-likelihood of } \theta)$. When error is symmetrically distributed around the true value, inaccuracy may be defined as the average (expected) distance.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

$$A(\theta) = A(\mu) - [(\mu - \theta)^2/2\sigma^2].$$

The right side of this equation has its maximal value when $\mu = \theta$. In other words, nothing can be more predictively accurate than the truth; the more distant the estimate θ is from the true population mean μ, the worse, on average, θ will do in predicting new data. Nagel's intuition that truth and predictive accuracy go hand-in-hand is vindicated in this type of inference problem.

However, just like the sure thing argument, the present argument also has its limitations. First, even if it is true in this type of inference problem that nothing is more predictively accurate than the truth, it doesn't follow that this holds for *all* inferences. If we can find even one situation in which truth and predictive accuracy come into conflict, we'll have an interesting test case for instrumentalism. The other limitation is both more serious and more subtle. If the true hypothesis is the one that is most predictively accurate, does it follow that the best way to make accurate predictions is to try to find the truth? This may seem to follow, but in fact it does not. The following principle is false:

(*) If you want to maximize A and T maximizes A, then the best way to maximize A is to try to maximize T.

To see why (*) is a bad principle, suppose that if you aim at T, you will have a very small probability of finding T (though if you do, you'll maximize A), but that if you aim at something else, you'll have a better chance of scoring a high value with respect to A.

The defect in the (*) principle is not limited to what it says about the goals of truth and predictive accuracy. Suppose you go to the bus terminal and want to get on a bus that will take you as close as possible to Fred's house.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

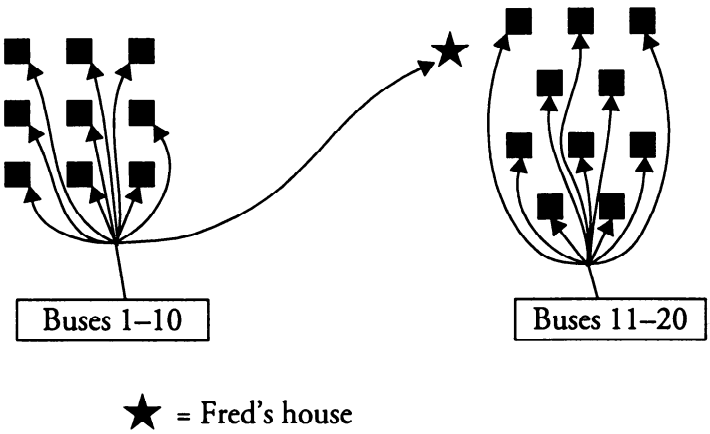
The buses in the city are numbered. You know that one of the buses that is numbered 1–10 goes right to Fred's door, but you don't know which one it is. You also know that no bus that is numbered 11–20 goes there. Should you take a bus numbered 1–10? Well, you've got a 10% chance of getting right to Fred's house if you take a low-numbered bus, and a 0% chance of being dropped at Fred's house if you take a high-numbered bus. So, if your goal were simply to be delivered to Fred's door, you should take a low-numbered bus.

But now let's add a little more information. Suppose that one of the buses numbered 1–10 takes you right to Fred's door, while the other nine take you very far away; on the other hand, all of the buses numbered 11–20 go very near Fred's house, though none of them goes right to his door. The bus routes are as depicted in Figure 1. If your goal is to get as close as possible to Fred's house, you should take a bus numbered 11–20. The point is this: even if a bus with a low number is the one that goes closest to Fred's house, it doesn't follow that the best way to get close to Fred's house is to take a low-numbered bus. Similarly, even if the true hypothesis is the one that is most predictively accurate, it doesn't follow that the best way to maximize predictive accuracy is to try to find the truth; the (*) principle is false. This suggests that there may be inference problems in which trying to find the truth and trying to maximize predictive accuracy lead to different decisions. The bus example suggests that this may be possible even if no hypothesis is more predictively accurate than the truth.[4]

[4] In Sober (1998), I argue that the (*) principle underlies a fallacious argument in the surprise examination problem, in which one concludes that the teacher should assign a probability of zero to giving an exam on the last day of the semester if her goal is to surprise the students.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

FIGURE 1



The example I want to describe in which the quest for truth and the quest for predictive accuracy lead to different decisions is an extremely mundane problem in statistical inference. Suppose you examine two populations of corn plants and want to determine whether the mean height (μ_1) in the first population is the same as the mean height (μ_2) in the second. You sample a number of plants from each population and have to figure out from these samples which hypothesis to accept. In formulating the question as one about "acceptance", I leave open whether "acceptance" means *believing that the hypothesis is true* or *believing that it will be predictively accurate.*[5] The question is whether the practices of scientists in this

[5] Although I'll formulate the problem in terms of the concept of "acceptance", this is a matter of convenience; the dichotomous concept of acceptance could be replaced with the concept of degree of belief. Formulated in the latter way, the question would be whether the goal of science is to say how probable it is that various hypotheses are true, or to say how predictively accurate one should expect those hypotheses to be.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

routine testing problem provide evidence as to what the underlying goal is of scientific inference.

The two hypotheses, then, are as follows:

(Null) $\mu_1 = \mu_2$

(Diff) $\mu_1 \neq \mu_2$.

The (Null) hypothesis has this name because it says that the difference between the population means is zero. The conventional practice in science is to compare the sample means (θ_1 and θ_2) drawn from the two populations and determine whether they are significantly different. If they are, you reject (Null) and accept (Diff). If the two samples do not differ significantly, you decline to reject (Null) and you also do not accept (Diff).

What does it mean for θ_1 and θ_2 to differ "significantly?" The idea is that θ_1 and θ_2 are sufficiently different that there would be only a small probability (5% is the conventional choice) that their values could be that different (or more) if in fact the null hypothesis were true. For example, if the two corn populations have the same mean height, then it is exceedingly improbable in sampling 10 plants from each population that one would end up with $\theta_1 = 50$ inches and $\theta_2 = 68$ inches, and with the 10 plants in the first sample tightly clustered around 50 inches and those in the second tightly clustered around 68 inches. On the other hand, with only 10 plants drawn from each population, it would not be especially improbable, if the (Null) hypothesis were true, to obtain samples in which $\theta_1 = 60$ inches and $\theta_2 = 62$ inches and with the two samples each showing a variance of one inch around the observed mean values. Whether an observed difference is statistically significant depends on the sample size; if 1000 plants were drawn from each population, a difference of 2 inches between the sample means, given a sample variance

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

of one inch, would be significant (see Sokal and Rohlf 1969, pp. 220–223 for details).

This is just standard statistical practice. I suggest that this practice is completely irrational if the goal in science is to discover whether (Null) or (Diff) is true. This is because we know with virtual certainty, before the samples are drawn and afterwards as well, that the (Null) hypothesis is false. This hypothesis says that the two populations have exactly the same mean height. Who could believe this —that their average heights aren't just close, but are exactly the same, down to a thousand decimal places and beyond? If "not rejecting" the (Null) hypothesis means believing it, then scientists are crazy. And if "not rejecting" means remaining agnostic, then scientists are also crazy, in that they are refusing to assent to a proposition (namely, Diff) that is obviously true.

On the other hand, if the goal is to choose hypotheses that will be predictively accurate, the routine scientific practice I have described makes sense. To flesh out this suggestion, I need to explain how (Null) and (Diff) are used to make predictions. The idea is that a model makes predictions about new data by using the old data to estimate the values of the adjustable parameters it contains. Suppose your sample means are $\theta_1 = 60$ inches and $\theta_2 = 62$ inches. If you use these observations to identify the likeliest estimates of the parameters found in the two hypotheses, you'll obtain:

L(Null) $\mu_1 = \mu_2 = 61$ inches.

L(Diff) $\mu_1 = 60$ inches and $\mu_2 = 62$ inches.

L(Null) is the likeliest member of (Null) in the sense that it confers a probability on the observations that is greater than the probability entailed by any other assignment:

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

$p[\theta_1 = 60$ inches and $\theta_2 = 62$ inches $\mid \mu_1 = \mu_2 = 61$ inches] >
$p[\theta_1 = 60$ inches and $\theta_2 = 62$ inches $\mid \mu_1 = \mu_2 = x$ inches], for any $x \neq 61$.

L(Diff) is the likeliest member of (Diff) for the same reason:

$p[\theta_1 = 60$ inches and $\theta_2 = 62$ inches $\mid \mu_1 = 60$ inches and $\mu_2 = 62$ inches] >
$p[\theta_1 = 60$ inches and $\theta_2 = 62$ inches $\mid \mu_1 = x$ inches and $\mu_2 = y$ inches], for any $x \neq 60$ and for any $y \neq 62$.

If L(Null) is the likeliest member of (Null) and L(Diff) is the likeliest member of (Diff), how do L(Null) and L(Diff) compare to each other? The answer is that L(Diff) has the higher likelihood:

$p[\theta_1 = 60$ inches and $\theta_2 = 62$ inches $\mid \mu_1 = 60$ inches and $\mu_2 = 62$ inches] >
$p[\theta_1 = 60$ inches and $\theta_2 = 62$ inches $\mid \mu_1 = \mu_2 = 61$ inches].

Even though L(Diff) fits the *old* data better than L(Null) does, most scientists will expect L(Null) to do a better job of predicting *new* data in the circumstance described. They will suspect that L(Diff) has *overfit* the old data; that is, they'll suspect that the small difference between the two sample means is just sampling error —it is noise, not signal. This doesn't mean that they in their hearts believe that the two populations have exactly the same mean height when they see two sample means that are only slightly different. I've already claimed that no one could or should believe that. Scientists regard the small

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

difference in sample means as misleading as far as the task of predicting *new* data is concerned.

The procedure whereby scientists use hypotheses that contain adjustable parameters to predict new data can be diagramed as follows:

$$\begin{array}{ccccc} & \text{Null} & & & \\ & \left.\begin{array}{c}\\ \\ \\ \end{array}\right\} & \rightarrow & \text{L(Null)} & \searrow \\ \text{Old Data} & & & & \text{New Data} \\ & \left.\begin{array}{c}\\ \\ \\ \end{array}\right\} & \rightarrow & \text{L(Diff)} & \nearrow \\ & \text{Diff} & & & \end{array}$$

Given the old data, one *deduces* that L(Null) is the likeliest member of (Null) and that L(Diff) is the likeliest member of (Diff); these likeliest members, in turn, make *probabilistic* predictions about new data.

How are the predictive accuracies of (Null) and (Diff) to be understood? This is a slightly different question than the one we asked earlier about the predictive accuracy of an estimated value of θ, since (Null) and (Diff) contain adjustable parameters. However, the idea of average performance in a series of prediction problems provides the common thread. As shown in the above flow chart, we imagine a two-part process in which parameters are estimated from an old data set and a new data set is predicted on that basis. A model is predictively accurate to the degree that this process, on average, generates predictions that are close to the means observed in new data. An accurate model may on occasion find itself faced with an old data set that leads it to do a poor job of predicting new data; but, on average, an accurate model will come close to the means found in new data.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

Notice that (Diff) is a *family* of hypotheses —it is comprised of the infinitely many specific hypotheses obtained by assigning different values to μ_1 and μ_2. We are imagining that one of these specific hypotheses is true; if only you could find that member of (Diff) and use it to predict new data, you'd be doing as well as it is possible to do, since nothing works better than the true specific hypothesis in the inference problem we are considering. The problem is that you don't know which member of (Diff) is true, even though you know in advance that one of them is. You also know that no member of (Null) is true, but, curiously, if the sample means differ only modestly, (Null) is apt to yield more accurate predictions of new data than (Diff).

I hope the analogy with the bus problem is becoming clear. You know that one of the low-numbered buses goes right to Fred's house, and that none of the high-numbered buses does so. This is analogous to your knowing that (Diff) is true and (Null) is false. If your goal were literally and only to reach Fred's door, then you should take a low-numbered bus. Likewise, if your goal were literally and only to find the true hypothesis, then you should choose (Diff) and reject (Null). However, real bus-riders aren't like this and real scientists aren't either. Real bus-riders want to get as close as possible to their destinations; they think that the saying "a miss is as good as a mile" is absurd. And real scientists want accurate predictions and the more accurate the better. What this means is that it can make sense for bus-riders to take a bus that they know in advance cannot take them precisely to their destination; rather, they choose a high-numbered bus because that bus can be expected to come closer to Fred's house than a low-numbered bus. And scientists will use (Null) to predict new data rather than (Diff) when the sample means differ only a little because they expect the false hypothesis (Null)

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

to deliver more accurate predictions in this instance than the true hypothesis (Diff).[6]

There is an important asymmetry between (Null) and (Diff). If (Diff) is true, it can sometimes be better to use (Null) to predict new data. However, if (Null) is true, it can never be better to use (Diff) to predict new data. This asymmetry is described in the following table:

	(Null) is more predictively accurate	(Diff) is more predictively accurate
(Null) is true	possible	impossible
(Diff) is true	POSSIBLE	possible

I've written one of the entries in this table in capital letters because it is the one that matters to the dispute between realism and instrumentalism.

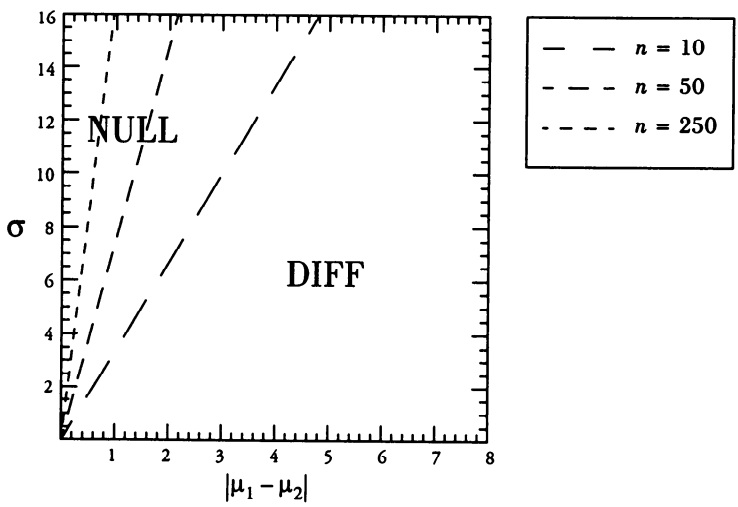
The three possibilities represented in this table are depicted in more detail in the accompanying figure, which summarizes the results of a large number of computer simulations. The values for the two population means are assumed to fall between 0 and 100; the x-axis represents cases in which the difference between the two population means $| \mu_1 - \mu_2 |$ is less than 8. The y-axis represents some possible values for the within-population variance, σ^2. In the square, (Diff) is true practically everywhere; (Null), on the other hand, is simply the line that comprises the y-axis —it is true when $| \mu_1 - \mu_2 | = 0$. When n=10 individuals are drawn from each population, (Diff) will probably

[6] As noted in footnote 5, describing acceptance and rejection as a dichotomous choice is incidental to the argument of this paper. If scientists assigned probabilities to hypotheses, what probability should they assign to (Null) and (Diff)? I am suggesting that they should assign (Null) a very low probability of being *true*, both before and after they look at data; however, the data may indicate that (Null) has the higher expected degree of *predictive accuracy*.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

be more predictively accurate in the region shown; (Null) is more predictively accurate when it is true, but also in a region in which it is false. As the sample size is increased, the region in which (Diff) will be more predictively accurate increases in size; this consequence of increasing sample size is depicted for the cases of n=50 and n=250.[7]

FIGURE 2



This figure depicts the "metaphysics" of the relation between truth and predictive accuracy, so to speak. It shows the region of parameter space in which a false hypothesis has a higher degree of predictive accuracy than a true one; this says nothing about whether or how scientists are able to determine which hypothesis are true and which will make more accurate predictions. So far, I've addressed this two-part epistemological question simply by appealing to

[7] These simulations closely agree with the analytic solution that Branden Fitelson obtained, according to which (Null) will be more predictively accurate (in expectation) than (Diff) precisely when $|\mu_1 - \mu_2| < 1.34898\ \sigma/\sqrt{n}$.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

experience. First, I've claimed that our general knowledge of the world tells us, with as much certainty as we're ever liable to have, that (Diff) is true and (Null) is false. Second, I've said that it is part of the day-to-day experience that scientists have when they use models to make predictions that (Diff) can be expected to make less accurate predictions than (Null) when the sample means differ only a little. These facts about the practical knowledge that scientists have would be enough to show that the behavior of scientists favors instrumentalism over realism, at least in the context of the problem of choosing between (Null) and (Diff). However, there is more to be said about the epistemology of this problem. It isn't just the lived experience of scientists that is relevant here; there is a mathematical theory that undergirds the expectations that scientists have about the inference problem at hand.

That undergirding was provided by the Japanese statistician H. Akaike (1973, 1977; see the excellent review in Burnham and Anderson 1998). Akaike proved a theorem that shows how one can obtain an unbiased estimate of the predictive accuracy of a family of hypotheses. A family has some number of "adjustable parameters". To explain what this means, and to make more precise the way in which (Null) and (Diff) can be used in prediction, I need to describe more carefully what the two hypotheses say:

(Null) $| \mu_1 - \mu_2 | = 0 + N(0, \sigma^2)$
(Diff) $| \mu_1 - \mu_2 | = \beta + N(0, \sigma^2)$, where $\beta \neq 0$.

The (Null) hypothesis says that there is no difference between the two population means, but that sampling from the two populations and examining the difference between the two sample means (θ_1 and θ_2) is subject to error; variation in the sampled difference is described by a normal distribution with mean 0 and variance σ^2. The (Diff) hy-

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

pothesis says that the two population means differ by a value $\beta \neq 0$, and that this difference is also subject to the sampling distribution given. Notice that (Null) has one adjustable parameter (σ^2) and (Diff) has two (β, σ^2). L(Null) and L(Diff), on the other hand, contain no adjust*able* parameters; these specific hypotheses are obtained from (Null) and (Diff) by using the data to substitute constants for parameters; their "parameters" have all been adjust*ed*. Akaike's theorem says that the predictive accuracy of a family can be estimated by attending to two considerations —how well the likeliest member of the family fits the evidence at hand and how many adjustable parameters (k) the family contains:

> An unbiased estimate of the predictive accuracy of the family F =
> Log-likelihood[L(F)] – k.

In our example in which the sample means were θ_1 = 60 inches and θ_2 = 62 inches, L(Diff) has a higher log-likelihood than L(Null); that is, the data are more probable according to the hypothesis L(Diff) than they are according to L(Null). However, the estimated predictive accuracy of a family doesn't depend just on how well its likeliest member accommodates old data. Akaike's theorem says that the estimate also should take account of the complexity of the family. (Diff) receives the higher penalty for complexity than (Null), since (Diff) contains more adjustable parameters. Thus, if (Diff) and (Null) fit the data about equally well, one should expect the simpler hypothesis, (Null), to be more predictively accurate.[8]

[8] If σ is known in advance and thus need not be estimated from the data, then (Null) has zero adjustable parameters and (Diff) has one, and the penalty term in Akaike's theorem has the form σ^2k.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

Although Akaike's theorem makes no overt mention of sample size, this consideration influences estimates of predictive accuracy, just as it influences the decision in conventional statistics as to whether the (Null) hypothesis should be rejected. When $\theta_1 = 60$ inches and $\theta_2 = 62$ inches and 10 individuals are sampled from each population, the likelihoods of L(Null) and L(Diff) will be fairly close together. However, L(Diff) would be much more likely than L(Null) if the same sample means were obtained by examining 1000 individuals from each population. For a fixed quadruple of sample means and sample variances, simplicity matters more in the estimate of predictive accuracy for smaller data sets than for larger ones.

There is still some controversy in statistics about Akaike's results. For example, there are model selection criteria on the market that impose different penalties for complexity (see Burnham and Anderson 1998, pp. 70–73, and Forster 1999 for discussion). However, this does not affect the philosophical question I'm addressing. Different criteria sometimes provide different advice about which models one should use to predict new data; however, there is no disagreement about the fact that a false model can sometimes be more predictively accurate than a true one.

It also is no objection to the argument I've presented to point out that scientists usually use standard Neyman-Pearson statistics to evaluate (Null) and (Diff), and these procedures are conceptually quite different from the ones that the Akaike framework recommends. Even if Neyman-Pearson statistics provided a satisfactory account of why one should sometimes prefer (Null) over (Diff), the argument would still go through. However, I believe that Neyman-Pearson statistics doesn't do a very good job of justifying the methods that the framework advocates. In the example I've been discussing, one is told to reject the (Null) hypothesis if and only if the difference between θ_1

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

and θ_2 is statistically significant. It is a matter of definition that a significant difference θ_1 and θ_2 is very improbable if (Null) is true. Presumably, this is a reason to "reject" the (Null) hypothesis, or at least to reduce one's degree of belief in it, only if (Diff) does a better job of accommodating the observations. But what probability does (Diff) confer on a significant difference between θ_1 and θ_2? The trouble is that (Diff) doesn't make any predictions. This is why Neyman-Pearson statistics focuses its attention on the predictions made by the (Null) hypothesis; as far as Neyman-Pearson statistics is concerned, (Diff) is a mystery that is passed over in silence.

The Akaike framework provides a very different analysis. (Diff) makes predictions in the same way that (Null) does; one uses the old data to estimate the values of adjustable parameters and then uses L(Diff) and L(Null) to predict new data. When θ_1 and θ_2 differ only modestly, one prefers (Null) over (Diff) because the former has the higher estimated predictive accuracy. When the observed difference is larger, one makes the opposite decision. The beautiful thing about the Akaike approach is that it throws light on the properties of *both* competing hypotheses. This is why I think that the Akaike framework provides a better explanation than the Neyman-Pearson framework of why the behavior of scientists makes sense, even though scientists usually appeal to Neyman-Pearson and often have never even heard of Akaike. However, I want to emphasize that the argument for instrumentalism doesn't depend on the Akaike framework's being correct. I mention it to point out that it isn't just the *de facto* practices of scientists that underwrite my claim that a false theory sometimes can be expected to be more predictively accurate than a true one; there is, in addition, a mathematical explanation of why this is so.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

At the start of this paper, I discussed the Nagelian motto "nothing is more predictively accurate than the truth". What is the status of this principle in the context of the present problem? As the previous table and figure show, (Null) can be more predictively accurate than (Diff) even when (Null) is false and (Diff) is true. However, (Null) and (Diff) are each infinite disjunctions. If (Diff) is true, let's denote the true member of (Diff) as T(Diff). It certainly is correct that T(Diff) has a higher degree of predictive accuracy than any other member of (Diff) or (Null). The Nagelian formula fails when it is applied to *families* of hypotheses; however it is correct when it is applied to the specific members of those families that contain no adjustable parameters.

Although instrumentalism is often thought of as a version of empiricism, the construal of instrumentalism presented here differs from the constructive empiricism of Van Fraassen (1980). According to Van Fraassen, the goal of science is to find theories that are empirically adequate, which means, roughly, that they are true in what they say about observables. However, in the example under discussion, (Diff) is empirically adequate while (Null) is not. If so, the goal of predictive accuracy and the goal of empirical adequacy are different. If my argument shows that instrumentalism is a better account of the goals of science than realism, it also shows that instrumentalism is superior to constructive empiricism.

Where does this discussion leave the general issue of realism versus instrumentalism? I've described a standard problem of statistical inference and have argued that the goal that scientists pursue in their efforts to solve this problem is one of predictive accuracy, not truth. I do not conclude from this example that scientists *always* aim at predictive accuracy rather than truth. First, there are many inference problems in which the two goals coincide; instru-

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

mentalists cannot cite such inference problems as evidence for their position, but neither can realists. In this circumstance, there is no issue worth discussing between realism and instrumentalism, just as Nagel claimed. Furthermore, the analysis I've suggested for the case of (Null) and (Diff) does not rule out the possibility that there may be *other* inference problems in which truth and predictive accuracy conflict, and where scientists prefer the hypothesis they think is true and predictively inaccurate over one that they think is false and predictively accurate. In terms of the second two-by-two table, perhaps there are situations in which the choice is between H3 and H4 and scientists prefer H3. Realists need to produce such examples. We can't rule out the possibility, in advance, that *global* instrumentalism and *global* realism are both false. Perhaps scientific practice is sufficiently diverse that *local* instrumentalism and *local* realism are both correct.

Can the inference problem of choosing between (Null) and (Diff) be reinterpreted so as to accord with the dictates of scientific realism? One suggestion in this vein is that the (Null) hypothesis should not be interpreted literally. It says that the two population means are exactly the same, but perhaps this gets glossed by scientists as the idea that the mean values are *approximately* the same. If this is how scientists interpret (Null), then perhaps they are not being irrational if they sometimes believe that the hypothesis is true. It is ironic that a realist should make this suggestion, since realism as a semantic thesis often involves an insistence on literal readings of scientific theories. But irony aside, the question is whether this suggestion accords with scientific practice. *Prima facie*, it does not; when scientists judge whether θ_1 and θ_2 differ significantly, they are talking about the probability of the sample means' being that different or more, conditional on the (Null) hypothesis'

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

being true. The numbers they use are obtained by taking the (Null) hypothesis to mean what it says.

If scientists interpret the (Null) hypothesis as saying that the two population means are within ϵ of each other, what value do they assign to ϵ? No matter what (nonzero) value the realist suggests for ϵ, the practice of science turns out to be irrational. If ϵ is tiny (say, 10^{-10} inches), then "$| \mu_1 - \mu_2 | < \epsilon$" remains a hypothesis that we know to be false, and so scientists are behaving irrationally when they fail to reject it. On the other hand, if scientists assign ϵ a larger value (say, 2 inches) in their construal of (Null), then their statistical practice is irrational for another reason. If scientists interpret the (Null) hypothesis as saying that the means are no more than 2 inches apart, then they should *not* reject the (Null) hypothesis when they find that θ_1 and θ_2 differ by 1 inch in a large sample. However, this is precisely what they do. This argument generalizes to any setting of ϵ, large or small. The behavior of scientists shows that they interpret (Null) literally.

Another way to try to bring this inference within the orbit of realism is to suggest that scientists are being good realists because they are trying to choose hypotheses that are *close* to the truth. When scientists judge that (Null) will be more predictively accurate than (Diff), based on the data at hand, they are inferring that the values for μ_1 and μ_2 specified in L(Null) are probably closer to the true population means than are the values given in L(Diff). My reply is that this is correct, but the fact remains that when scientists fail to reject (Null) they are failing to reject a hypothesis that they know is false, and when they fail to accept (Diff) they are failing to accept a hypothesis that they know is true. If the goal of scientific inference were merely to assign truth values to (Null) and (Diff), the behavior of scientists would be very different.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

The argument for instrumentalism that I have constructed is predicated on the assumption that it is rational for scientists to refuse to reject the (Null) hypothesis when they obtain sample means θ_1 and θ_2 that differ only a little. I used this rationality assumption to argue that scientists are seeking predictive accuracy, not truth, in this instance. However, why accept the assumption that scientists are rational when they act like this? Perhaps scientists really *are* trying to discover the truth but are just doing a poor job of selecting methods to achieve that end. How are we to tell whether scientists are rational instrumentalists or irrational realists? One way to address this question is to present the argument of this paper to scientists themselves. Once they understand the argument, will they change their behavior and reject the (Null) hypothesis no matter what the data are? I suspect that they will not. Of course, this in itself, does not conclusively prove that they are rational; after all, they may be irrational about their choice of inference methods and impatient when they listen to arguments made by annoying philosophers. However, I still think that the test just sketched would provide evidence.

My colleague Haskell Fain once described a similar problem about basketball. How do we know that the players are trying to score baskets, rather than trying to miss and doing a bad job of it? One test would be to say the following to basketball players —"look, if you're trying to miss baskets, here are some things you can do that will allow you to increase your effectiveness in attaining that goal". If players take the proffered advice and start to score fewer baskets, we have evidence that their goal is to miss; however, if they reject the advice and keep doing the same old things, we have evidence that their goal is to score.

Although the inference problem of testing (Null) against (Diff) has considerable generality, I think the case for instrumentalism goes further. I've focused on an inference

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

problem in which scientists know that truth and predictive accuracy conflict. But there is another type of inference problem that also bears on the issue of instrumentalism versus realism. It is depicted in the following table:

		Believed to be	
		More predictively accurate	Less predictively accurate
Believed to be	True		
	False	H4	H2

This is a good way to describe the problem that scientists face when they try to choose between different *idealizations* (Forster and Fitelson, unpublished). If the *only* goal of science were to find true theories, then H4 and H2 should be equally unsatisfactory. However, the instrumentalist has an obvious explanation of why scientists prefer H4 over H2.

In fairness to scientific realism, I have to admit that there is yet another type of inference problem that needs to be considered:

		Believed to be	
		More predictively accurate	Less predictively accurate
Believed to be	True	H1	
	False	H4	

If scientific practice provided methods for choosing between theories that are *predictively equivalent*, that would favor realism over instrumentalism. Many philosophers of science seem to think that this is precisely what considerations of parsimony, unification, etc. permit scientists to do. I do not, but that is a subject for another occasion (see Sober 1996 for discussion).

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

In the present paper, I have described four types of inference problem. The first is one in which truth and predictive accuracy are thought to *coincide* (H1 versus H2); here the behavior of scientists provides no information about whether realism or instrumentalism is right. The second type of problem is one in which truth and predictive accuracy *conflict* (H3 versus H4); I described an inference problem of this type and argued that the behavior of scientists suggests that instrumentalism is more plausible than realism. But even if I'm right about this example, that doesn't justify instrumentalism *tout court*. Instrumentalism and realism, as I've construed them, are monistic doctrines; each says that scientific inference is aimed at a single ultimate goal. If scientists sometimes prefer a false theory over a true one because they expect the former to be more predictively successful, this shows, at minimum, that the pursuit of truth is not their *only* ultimate aim. Instrumentalism is consistent with this result, but so is the pluralistic position that says that truth and predictive accuracy are both ultimate goals. The same ambiguity must be recognized in the third and fourth types of inference problem that I described. In the third, when scientists choose one idealization over another (H4 versus H2) because they expect the former to be more predictively successful, this doesn't establish that successful prediction is their *only* goal. And in the fourth type of problem, if scientific inference sometimes allows one to discriminate between predictively equivalent hypotheses (H1 versus H4), that doesn't show that truth is the *only* ultimate goal in scientific inference. Both instrumentalists and realists may need to expand their horizons; these monistic theories have the virtue of simplicity, but scientific inference may be variegated enough that a more complex model is warranted.

Indeed, a *prima facie* argument that favors pluralism over instrumentalism is already at hand. In the discussion

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

of (Null) and (Diff) I claimed that scientists are reasonable in believing that (Diff) is true *and* that (Null) will be more predictively accurate when the sample means θ_1 and θ_2 differ only a little. This judgment is a conjunction and I take it that the argument for each conjunct is made by using principles of inference that are part of "the scientific method". Were scientists to disown the argument that assigns truth values to (Null) and (Diff),[9] that might suggest that their only interest, qua scientists, in those hypotheses concerns the issue of predictive accuracy. But this, I take it, is not what they do or should do.

If this is right, then there is no univocal notion of "acceptance" that is potentially influenced both by judgments about truth and judgments about predictive accuracy. This is how I formulated the problem of choosing between (Null) and (Diff), but that formulation made sense only because the problem was to test the monistic theories of realism and instrumentalism against each other. However, if pluralism is the right way to view the goals of science, we should think of scientists as accepting that (Diff) *is true* and as (sometimes) accepting that (Null) *is more predictively accurate*. There needn't be, in addition, some all-things-considered notion of acceptance that takes these two judgments into account and delivers a decision about which hypothesis to "really" accept.

Even if scientists have these two irreducibly distinct goals, the question remains, within a pluralistic framework, of which goal is more important in shaping the trajectory of

[9] The argument for thinking that (Null) has got to be false is Bayesian, but not in a way that should be controversial. A population can be thought of as having different probabilities of exhibiting different average heights, depending on the mix of genotypes and microenvironments that pertain. The chance that two populations should end up with exactly the same mean height is zero. In just the same way, the chance that an infinitely sharp pin dropped on a line a mile long will land exactly at point p, rather than anywhere else, is zero.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

scientific change. We need to look at episodes of change in fundamental theory and at more mundane cases in which one lower-level hypothesis replaces another and see whether scientists are looking for truth, or for predictive accuracy, or for both. Both the macro- and the micro-evolution of scientific ideas matter to this question. If the example of (Null) and (Diff) is any guide, we may speculate that scientific change that is data-driven is driven by the goal of predictive accuracy; one needn't look at data from the two populations of corn plants to see that (Null) is false and that (Diff) is true.

What, then, was the point of revisiting instrumentalism? Once the lightning rod effect is seen for what it is, we can see that methodological instrumentalism is not refuted by the fact that semantic instrumentalism is mistaken. And when truth and predictive accuracy coincide, we can see that methodological instrumentalism is not refuted by the fact that scientists prefer true theories over false ones. The more relevant situation is the one in which the potential goals of truth and predictive accuracy conflict. Here we found a case in which it is standard scientific practice to not reject a hypothesis that is palpably false and to not accept a hypothesis that is obviously true; this practice makes sense if scientists are looking for theories that are predictively accurate. The example of (Null) and (Diff) shows that instrumentalism was right to claim that realism had missed something. However, the fact that instrumentalism contains a kernel of truth does not show that it is true. In looking for a case in which instrumentalism could be tested against realism, we also seem to have found a case in which pluralism can be tested against instrumentalism. Instrumentalism was worth revisiting even if we decide in the end that it isn't true.

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

Appendix on Verisimilitude[10]

Can scientific practice with respect to (Null) and (Diff) be explained by the hypothesis that scientists want their theories to be "close to the truth"? Popper's endorsement of this conception of the goal of science gave rise to the verisimilitude program in philosophy of science (Niiniluoto 1987, 1998). Although the idea of closeness to the truth has been explicated in a number of ways, the verisimilitude of a statement is always taken to reflect its truth value *and* its information content. Because of this second component, (Null) may be closer to the truth than (Diff), even though (Null) is false and (Diff) is true. After all, (Diff) is almost a tautology whereas (Null) has a lot of content.

Since (Diff) is a disjunction, how is its distance from the truth related to the distance from the truth of its disjuncts? If each population's possible mean height is scaled to fall between 0 and 100 inches, then the two hypotheses can be represented in terms of a line:

```
(−100) ---- t ----- 0 ---------- (+100)
              |←--d-->|
```

This line represents possible values of the mean of the first population minus the mean of the second. (Null) says that the true value t = 0; (Diff) says that the true value is somewhere else. Suppose that the true value (t) is d units from 0. Then the distance from the truth of (Null) is d. If (Diff)'s distance from the truth is defined as the average distance of the different disjuncts from the truth

[10] I am grateful to Theo Kuipers for calling my attention to the relevance of the verisimilitude idea to the argument of this paper and for his help in understanding how Niiniluoto's quantitative notion of verisimilitude applies to the example of (Null) and (Diff).

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

(with each disjunct having equal weight), then the average distance from the truth of (Diff) is:

$$[d/200]\{d/2\} + [(100-d)/200]\{(100-d)/2\} + [100/200]\{50 + d\}.$$

Here the numbers in square brackets [. . .] are probabilities of falling in different intervals, and the numbers in curly brackets { . . . } represent distances from the truth to the centers of those intervals. (Null) is closer to the truth than (Diff) if and only if

$$d < d^2/400 + (100-d)^2/400 + \tfrac{1}{2}(50+d),$$

which simplifies to $0 < (d\text{-}100)^2$. This is true for all values of $d \neq 100$; when $d = 100$, the two sides are equal. So (Diff) is never strictly closer to the truth than (Null), and one can know this *a priori*. Yet, we have seen that (Diff) is sometimes more predictively accurate, and whether this is so is something one can't know *a priori*. It follows that distance to the truth cannot be defined in the way described, if one wishes to say that scientists consult the data to determine whether (Null) or (Diff) is apt to be "closer to the truth".

Niiniluoto (1987, p. 216) gives a different formula to define the distance D(I,t) from an interval hypothesis I to the truth t, which has a point value. He uses two parameters p and q (each having values between 0 and 1):

$$D(I,t) = p[Dmin(t,I)] + q[Dsum(t,I)].$$

Here Dmin(t,I) is the minimum (normalized) distance between t and I, and Dsum(t,I) is the normalized integral of the distances between t and the members of I. From this it follows that

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

$$D(\text{Null},t) = p[d/200] + q[0] = pd/200$$

$$D(\text{Diff},t) = p[0] + q[1] = q.$$

Which hypothesis is closer to the truth depends on the truth (d), but also on the values of p and q, which represent how much one values truth and how much one values content. Distance from the truth therefore reflects the values of the inquirer as well as features of the world. Furthermore, if there is no uniquely correct mix of values that an inquirer ought to have, there will be no uniquely correct answer to the question of whether (Null) is closer to the truth than (Diff).

Although the distance from the truth of (Null) and (Diff) in the sense defined depends on d, p, and q, it is worth recalling from the previous figure that the predictive accuracy of the two models depends on d, σ^2, and n. Here I am talking about the *definition* of predictive accuracy, not about Akaike's method for constructing an unbiased *estimate* of that quantity. Predictive accuracy depends on perfectly objective features of the populations under study (including the manner in which samples are drawn from those populations), which are independent of the values of the inquirer.

How might Niiniluoto's idea of distance from the truth be harnessed epistemologically? Does it make sense to say that scientists decide between (Null) and (Diff) by deciding which is "closer to the truth" in the sense just defined? One might try to flesh out this idea by saying that scientists use the data to obtain a maximum likelihood estimate of d, and assign values to p and q as a matter of scientific convention. Depending on how these conventions are described, the verisimilitude idea might be able to mimic the policies recommended by the Akaike framework; if so, it probably can be made to reproduce the advice given by other rules

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

as well. However, it is hard to see how the verisimilitude framework can be used to decide which policy is best to pursue. Alternatively, it may be that the verisimilitude approach can use the goal of predictive accuracy as a way of spelling out its concept of closeness to the truth.

If the verisimilitude approach can be brought into agreement with the Akaike approach, in the sense that estimates of predictive accuracy induce the same ordering of hypotheses as estimates of their closeness to the truth, what becomes of the question of how each is related to the debate between realism and instrumentalism with which we began? Perhaps at this point we should take up Nagel's suggestion that the difference is merely verbal —we have moved to an H1 versus H4 problem in which closeness to the truth and predictive accuracy coincide. But if we do, we should not forget that the Nagelian formula —that nothing is more predictively accurate than the truth— is true in some cases, but false in others.

REFERENCES

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle", in B.N. Petrov and F. Csaki (eds.), *Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, pp. 267–281.

—— (1977), "On the Entropy Maximization Principle", in P. Krishniah (ed.), *Applications of Statistics*, Amsterdam, North Holland, pp. 27–41.

Burnham, K. and D. Anderson (1998), *Model Selection and Inference*, Springer, New York.

Forster, M. (1999), "The New Science of Simplicity", in H. Keuzenkamp, M. McAleer, and A. Zellner (eds.), *Proceedings of the Tilburg Conference on Simplicity*, Cambridge University Press, Cambridge.

—— (2000), "The Hard Problems in the Philosophy of Science —From Kuhn, to Lakatos, and Beyond", in R. Nola

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

and H. Sankey (eds.), *After Popper, Kuhn and Feyerabend —Issues in Theories of Scientific Method*, Kluwer, North Holland.
—— and B. Fitelson (unpublished), "Do Scientists Value Theories for their Truth or their Predictive Accuracy? A Bayesian Analysis".
—— and E. Sober (1994), "How to Tell When Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions", *British Journal for the Philosophy of Science*, 45, pp. 1–35.
Hempel, C. (1979), "Scientific Rationality, Analytic versus Pragmatic Perspectives", in T. Geraets (ed.), *Rationality Today*, University of Ottawa Press, Ottawa.
Nagel, E. (1979), *The Structure of Science*, Hackett Publishing, Indianapolis.
Niiniluoto, I. (1987), *Truthlikeness*, D. Reidel, Dordrecht.
—— (1998), "Verisimilitude —the Third Period", *British Journal for the Philosophy of Science*, 49, pp. 1–29.
Putnam, H. (1975), "Craig's Theorem", in *Mathematics, Matter, and Method —Philosophical Papers*, vol. 1, Cambridge University Press, Cambridge.
Quine, W. (1969), "Epistemology Naturalized", in *Ontological Relativity and Other Essays*, Columbia University Press, New York.
Sober, E. (1996), "Parsimony and Predictive Equivalence", *Erkenntnis*, 44, pp. 167–197.
—— (1998), "To Give a Surprise Exam, Use Game Theory", *Synthèse*, 115, pp. 355–373.
—— and D. Wilson (1998), *Unto Others —the Evolution and Psychology of Unselfish Behavior*, Harvard University Press, Cambridge.
Sokal, R. and J. Rohlf (1969), *Biometry*, W.H. Freeman, San Francisco.
Van Fraassen, B. (1980), *The Scientific Image*, Clarendon Press, Oxford.
Wilson, M. (forthcoming), "The Unreasonable Uncooperativeness of Mathematics in the Natural Sciences", *The Monist*.

Recibido: 11 de enero de 1999

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions

RESUMEN

En filosofía de la ciencia, el instrumentalismo a menudo se interpreta como la tesis semántica de que las teorías científicas no son ni falsas ni verdaderas, sino meros instrumentos útiles para predecir. El instrumentalismo que se examina en este trabajo es epistemológico, no semántico. Doy por hecho que las teorías son falsas o verdaderas; la cuestión es si el objetivo de la inferencia científica es explicar cuáles teorías son verdaderas, o explicar cuáles teorías son predictivamente exactas. Argumento que el propósito de encontrar teorías que son verdaderas y el de encontrar teorías que son predictivamente exactas pueden entrar en conflicto. Enseguida analizo un problema de inferencia estadística por medio del cual argumento que las prácticas de los científicos tienen sentido si consideramos que se proponen descubrir cuáles teorías son predictivamente exactas, pero que no tienen sentido si consideramos que se proponen descubrir cuáles teorías son verdaderas.

[*Traducción: Claudia Chávez A.*]

This content downloaded from 128.104.46.196 on Wed, 9 Oct 2013 12:22:17 PM
All use subject to JSTOR Terms and Conditions