Comparative Political Studies
Volume 40 Number 7
July 2007 886-908

10.1177/0010414006296518
http://cps.sagepub.com
hosted at
http://online.sagepub.com

The Necessity of Being Comparative

Theory Confirmation in Quantitative Political Science

Kevin A. Clarke
University of Rochester, New York

The aim of this article is to demonstrate that comparative theory testing is necessary if political scientists wish to make positive statements regarding the confirmation of their theories. Using the tools of formal logic, the author first establishes that theory confirmation is not possible when a theory is tested in isolation, regardless of the statistical approach—falsificationism, confirmationism, or Bayesian confirmationism—employed by the researcher. The author then establishes a necessary and sufficient condition for positive theory confirmation and shows that this condition is met only when two rival theories are tested against one another. Finally, the author discusses two methods of comparative theory testing demonstrating that being comparative, besides being necessary, is also straightforward and practical.

Keywords: *inference; comparative theory testing; falsificationism; Bayesianism*

Comparative politics has, of late, played the host for an extended conversation on inference in political science. King, Keohane, and Verba (1994) began the discussion by asserting the distinctly positivist claim that a single logic of inference underlies all social inquiry. The thread was quickly picked up by prominent comparativists—Laitin (1995), Caporaso (1995), Collier (1995), and Rogowski (1995)—who variously amend or replace that logic of inference with one closer to their own predilections. Brady and Collier (2004, p. 264) revive the discussion and attempt to specify an alternative methodology. Most recently, Johnson (2006, p. 224) addresses

Author's Note: The author thanks Jim Johnson, Gretchen Helmke, Mark Kayser, and Bob Pahre for helpful comments and discussion. Thanks also go to the editor and three anonymous referees for *Comparative Political Studies* for their advice and suggestions. Support from the National Science Foundation (Grant No. SES-0213771) is acknowledged gratefully.

the "baleful consequences" of the positivism that infuses contemporary political science and begins to sketch a pragmatic alternative with a focus on causal explanation.

As Johnson (2006, p. 246) notes, causal explanation is only "one crucial component" of such an alternative. The aim of this article is to make the case for a second crucial component—comparison. Meaningful theory confirmation is not a relationship between a single theory and the data but rather a relationship between multiple theories and the data. Many political scientists are, of course, no strangers to comparative theory testing, and the power of being comparative comes as no surprise to political scientists who find the philosophies of Kuhn (1970) and Lakatos (1970), not to mention Laudan (1977), Miller (1987), and Levey (1996), appealing. The argument I develop makes the claim that when attempting to confirm a theory through its deductive consequences, the standard practice in political science, one must be comparative to make a positive statement regarding theory confirmation.[1]

Why is comparative theory testing so necessary? After all, according to standard positivist accounts, how a theory performs against its rivals is not at issue (Miller, 1987, p. 7). Theory testing in political science, however, is accomplished through deductively derived hypotheses. A researcher deduces a hypothesis from a theory and then tests the hypothesis. If the hypothesis is supported by the data, the theory from which it was derived is confirmed (or "verified" or "supported" or "validated" or "gains confidence"). If the hypothesis is not supported by the data, the theory is not confirmed or falsified. It is precisely this deductive structure, however, that prevents these conclusions from being justified. My goal is to show how we, as political scientists, can successfully use different kinds of statistical tests in concert with this deductive structure to produce meaningful inferences.

Falsificationism, confirmationism, and Bayesian confirmationism are the three main statistical strategies that political scientists employ to test their theories. In making the case for comparison, I demonstrate that theory confirmation is logically impossible when a theory is considered in isolation regardless of which of these strategies is employed. A falsificationist strategy does not allow a researcher to logically falsify a theory—hence, a theory cannot gain confirmation through survival, and a confirmationist strategy, whether Bayesian or not, does not allow a researcher to logically confirm a theory. I establish a necessary and sufficient condition for positive theory confirmation and show that this condition is met only when two rival theories are tested against one another. Finally, I discuss two methods of comparative theory testing and demonstrate that being comparative, besides being necessary, is also simple and practical.

Valid and Invalid Deductive Arguments

A theory is connected to a hypothesis, the actual testable statement that relates two or more variables, through deduction.[2] A hypothesis is deduced, either formally or informally, from a theory and then tested on data. Understanding the connection between this structure and statistical testing requires understanding two kinds of arguments: one valid and one invalid.

The valid deductive argument is known as *modus tollens*, and it forms the basis of both falsificationism and classical hypothesis testing in statistics (see Howson & Urbach, 1993, p. 171). Consider the following valid deduction:

$$\left.\begin{array}{r}\text{If } x \text{ is human, then } x \text{ is mortal,}\\ x \text{ is not mortal}\end{array}\right\} \vdash x \text{ is not human.}$$

Obviously, all humans are mortal. Therefore, if we observe a thing that is not mortal, we can validly conclude that it is not human.

More formally, let T be a theory and H be a hypothesis. *Modus tollens* can then be written as $T \rightarrow H, \neg H \vdash \neg T$, where the conclusion, $\neg T$, can be validly inferred, $\vdash$, from the two premises, "if T, then H" and "not H."

The invalid argument is a logical fallacy known as "affirming the consequent." Consider the following invalid deduction:

$$\left.\begin{array}{r}\text{If } x \text{ is human, then } x \text{ is mortal,}\\ x \text{ is mortal}\end{array}\right\} \nvdash x \text{ is human.}$$

Obviously, not all mortal creatures are humans. Despite the fact that x is mortal, we cannot conclude that x is human. Using the same notation, the fallacy is $T \rightarrow H, H \nvdash T$, where the conclusion, T, cannot be validly inferred, $\vdash$, from the two premises, "if T, then H" and H.

The problem with affirming the consequent is that there are numerous conditions that imply that a thing is mortal, and being human is simply one of them. In the example given above, for instance, x could be a cat, a dog, or even a dandelion, and observing that "x is mortal" does not give us warrant to conclude that any particular one is true.

In what follows, I demonstrate how these two kinds of arguments underlie theory confirmation in political science and the problem that results. We will see that the only way of overcoming this problem is to practice comparative theory confirmation.

Deduction and Theory Confirmation

Although deductive theory confirmation in political science is easiest to see when comparative statics are derived from a formal theory (e.g., Stokes, 2005), it is, in fact, pervasive. Lake and Baum (2001), for instance, are interested in the effect of regime type on the provision of public services. They write of states as “firms that produce public services in exchange for revenue” and that production occurs “within a contestable market” (p. 590). In democracies, where politicians easily enter and exit, the state will “produce relatively larger quantities of goods at relatively lower prices,” and in autocracies, where politicians do not easily enter and exit, the state will “exercise its monopoly power, provide fewer public services, and earn greater rents” (p. 590).

From their theory, Lake and Baum (2001) derive the hypothesis that democracies should produce larger quantities of public services. The authors are explicit about the connection between their theory and the hypothesis. They note that the primary hypothesis follows “by necessity from the logic outlined earlier” in the article (p. 597). Explicitly stated, the logic is as follows:

States produce public services in exchange for revenue.
States produce within a contestable market.
States produce relatively larger quantities of goods at relatively lower prices when barriers and costs are low.

....

Democracies produce larger quantities of public goods.

Whether or not Lake and Buam actually derived their hypothesis from the given assumptions is unimportant. What matters for the argument I am making is the testing relationship between theory and hypothesis. For a researcher to claim that a supported hypothesis, in turn, supports the theory in question, the researcher must claim a relationship between them and that relationship is almost always deductive. Even if Lake and Baum came up with the hypothesis first and then found a theory that entails it, the point is that they claim that the hypothesis, as well as the hypotheses they test, can be derived from the assumptions.

Lake and Baum are not alone in their use of this deductive method; it dominates the literature in quantitative political science. A far from comprehensive survey of recent articles in comparative politics turns up numerous examples, including Conley (2006), Fox (2006), Rohrschneider (2005), and Rudra and Haggard (2005).

To investigate how this deductive structure interacts with statistical tests, we must note the three steps that a researcher goes through in the process of theory testing. Step 1 is to derive a hypothesis from the theory in question. In the Lake and Baum example given above, their theory talks of states as firms that produce public services in exchange for revenue in a contestable market, and from this theory, the authors deduce the hypothesis that democracies produce larger quantities of public goods. I use the symbols $T \wedge K$ to denote a theory and its auxiliary theories and background knowledge.[3] I use the symbols H_0 and H_1 to denote the null hypothesis and the research hypothesis, respectively. Thus, Step 1 for Lake and Baum can be written abstractly as $T \wedge K \rightarrow H_1$, which reads, "If the theory and background knowledge are true, then the hypothesis is true."

In Step 2 of the theory confirmation process, the researcher makes a statement that if the research hypothesis is true, then a certain statistical relation should hold in the data. Confirmationists, for example, would state that if the research hypothesis is true, then the coefficient on the variable of interest should be in a certain direction and of a certain size. Lake and Baum, who turn out to be confirmationists (see the "Conventional Confirmationism" section), argue that if their research hypothesis is true, the coefficient on democracy should be positive. I denote a claim of this sort as $H_1 \rightarrow$ *Coefficient is correct*, which reads, "If the research hypothesis is true, then the coefficient has the correct direction and size."

In Step 3 of the theory confirmation process, the researcher, having decided whether or not the data support the hypothesis, then claims that his or her theory is either falsified or confirmed. In the next few sections, I reconstruct these three steps for the three main quantitative testing strategies in political science: falsificationism, confirmationism, and Bayesian confirmationism. In each case, I show that the deductive structure of theory testing, when matched with these statistical strategies, prevents us from being able to confirm a theory. It is only when theory confirmation is truly comparative that positive statements regarding theory confirmation can be made.

Falsificationism

I begin with falsificationism for two reasons. First, political scientists, although they rarely practice it, often talk about theory testing in terms of falsificationism (see, e.g., King et al., 1994, pp. 100-105). Second, as previously noted, classical testing is based on the idea of falsificationism.[4] In the next section, I turn to a model that is much closer to the actual practices of political scientists, and I demonstrate that the same logical problem exists.

When I reconstruct the three steps discussed in the previous section in the falsificationist setting, all three steps must be written in terms of the null hypothesis, H_0. The reason is that, in practice, classical hypothesis testing concerns the null hypothesis. All political scientists are taught that the two possible outcomes of a hypothesis test are "reject the null" and "fail to reject the null." As no conclusions are to be drawn regarding the research hypothesis, the logic must be in terms of H_0. The practical import of writing the logic in terms of H_0 is that Step 1 assumes the following form, $T \wedge K \rightarrow \neg H_0$. That is, "if the theory is true, then the null hypothesis must be false." Few political scientists write this way, which is why few political scientists are actually falsificationists.[5] Understanding the logic of falsificationism is nonetheless important.

Under falsificationism, a theory gains confirmation by surviving attempts to falsify it (Chalmers, 1982, p. 45). To make my point then, I first show what falsificationists already know—confirmation is logically impossible. Second, I show that confirmation through survival is unjustified because falsification itself is logically impossible. If falsification is not possible, then survival means nothing. The two logics are depicted in Figure 1.

Unjustified Confirmation

The logic on the left-hand side of Figure 1 represents the argument that if the theory is true, then the null hypothesis should be false, and if the null hypothesis is false, then the theory is confirmed. In Step 1, the theory plus auxiliary hypotheses and background knowledge imply that the null hypothesis is false. In Step 2, we see the normal application of probabilistic *modus tollens* used in significance testing.[6] We reject the null hypothesis that the variable has no effect and provisionally accept the alternative hypothesis that the variable does have an effect.[7]

The problem occurs in Step 3. Given that the null hypothesis has been rejected, we cannot logically conclude that the theory should be accepted. If we mistakenly draw that conclusion, we are guilty of affirming the consequent. The problem, as noted in the "Valid and Invalid Deductive Arguments" section, is that the falsity of the null hypothesis is implied by an infinite number of theories. Consider, for example, a null hypothesis that states that the coefficient on a certain variable is zero, $H_0{:}\beta = 0$. Any theory that

Figure 1
The Logic of Falsificationism

Unjustified Confirmation

1. Theory to hypothesis

$$T \wedge K \rightarrow \neg H_0$$

2. Data to hypothesis

$$\left.\begin{array}{r} H_0 \rightarrow P(y \mid H_0) \text{ is high} \\ P(y \mid H_0) \text{ is low} \end{array}\right\} \vdash \neg H_0$$

3. Hypothesis to theory

$$\left.\begin{array}{r} T \wedge K \rightarrow \neg H_0 \\ \neg H_0 \end{array}\right\} \nvdash T \wedge K$$

Summary:

We falsely conclude that the theory is true due to the logical fallacy in (3).

Unjustified Falsification

1. Theory to hypothesis

$$T \wedge K \rightarrow \neg H_0$$

2. Data to hypothesis

$$\left.\begin{array}{r} H_0 \rightarrow P(y \mid H_0) \text{ is high} \\ P(y \mid H_0) \text{ is high} \end{array}\right\} \nvdash H_0$$

3. Hypothesis to theory

$$\left.\begin{array}{r} T \wedge K \rightarrow \neg H_0 \\ H_0 \end{array}\right\} \vdash \neg(T \wedge K)$$

Summary:

We falsely conclude that the theory is false due to the logical fallacy in (2).

includes the variable in question implies that the null hypothesis is false. Even if we were to restrict our attention to plausible alternative theories, there are far too many to allow confirmation of the theory of interest. The actual falsity of the null hypothesis, therefore, cannot serve as evidence for any of these theories.

Falsifying a null hypothesis cannot confirm the theory in question. At best, a researcher can claim that his or her theory has not been falsified. This conclusion has force only if falsifying the theory is possible, and the above logic demonstrates that falsification is not possible.

Unjustified Falsification

The logic on the right-hand side of Figure 1 represents the argument that if the theory is true, then the null hypothesis should be false, and if the null hypothesis is true, then the theory is falsified. In Step 1, the theory plus auxiliary hypotheses and background knowledge imply that the null hypothesis is false. The problem is in Step 2 where, once again, we see the logical fallacy of affirming the consequent. We fail to reject the null hypothesis that the variable has no effect, but we cannot validly conclude that the null hypothesis is true. Assume for the moment that we fail to reject the null hypothesis because the confidence interval includes zero. As with any confidence interval, a number of values exists that may have led to this conclusion. No effect (i.e., the coefficient in question equals zero) is only one of those values. To conclude on the basis that $P(y \mid H_0)$ is high that the variable has no effect is incorrect. The logical mistake in Step 2 denies the falsification that occurs in Step 3. Failing to reject a null hypothesis, therefore, cannot falsify the theory in question.

As we have seen, rejecting a null hypothesis cannot confirm the theory in question and failing to reject a null hypothesis cannot falsify a theory. Confirmation through survival is therefore not a logically justifiable position.[8] Thus, I turn to direct confirmation.

Conventional Confirmationism

Few political scientists are falsificationists, and the arguments reconstructed in Figure 1 no doubt appear odd to many. The disconcerting aspect of that logic is, of course, Step 1. Political scientists are accustomed to deriving a research hypothesis, H_1, from a theory, as opposed to deriving the falsity of the null hypothesis, H_0. The logic of confirmationism starts with the derivation of H_1 from the theory, $T \wedge K \rightarrow H_1$.[9]

Given that Step 1 is written in terms of H_1, the rest of the logic must follow suit. Step 2, therefore, can no longer be written in terms of H_0. The result is that political scientists have largely abandoned strict hypothesis testing in favor of a loose assessment of the direction and size of a coefficient. Signalling this approach is language asserting that an estimated coefficient of a certain size and direction "supports" or is "evidence in favor of" the hypotheses in question. Claims of this sort are antithetical to falsificationism, but they have come to dominate the actual practices of political scientists.

Consider again Lake and Baum (2001), which is a particularly clear example of how confirmationism works in political science. Recall from the section titled "Deduction and Theory Confirmation" that Lake and Baum are looking for evidence that directly supports the deductively derived hypothesis that democracies should produce larger quantities of public services. After running their analyses, the authors state that the results “strongly and consistently support” the hypothesis (Lake & Baum, 2001, p. 609). They note that “as predicted by our primary hypothesis,” the results indicate that “increases in democracy are, under varying conditions, significantly related to increases in the provision of public services” (Lake & Baum, 2001, p. 613). There is no pretense here of falsificationism, and Lake and Baum are admirably clear in their confirmationist testing strategy: A coefficient of the correct size and direction confirms the hypothesis.

I reconstruct the logic of conventional confirmationism in Figure 2. The argument is that if the theory is true, then the coefficient should be correct in direction and size, and if the coefficient is correct in direction and size, then the theory is correct. In Steps 2 and 3, however, we once again see the fallacy of affirming the consequent. The fact that the relevant coefficient is correct in its direction and size does not allow us to conclude H_1. The problem is compounded when the unjustified result from Step 2 is used to affirm the consequent in Step 3.[10] A coefficient that is in the size and direction specified by a hypothesis cannot confirm the theory that implied the hypothesis. As is always the case with logical fallacies in which we affirm the consequent, the same prediction may be a consequence of an infinite number of theories. If an infinite number of theories imply the same prediction, then no theory is confirmed when the prediction is borne out.[11] Confirmation is not possible when a theory is tested in isolation.

Bayesian Confirmation

Some confirmationists, in an attempt to save the programme, have turned to Bayesianism (Earman, 1992). Bayesian confirmation may be used in two different fashions. The first is in conjunction with the deductive method; a hypothesis is derived from theory, and then Bayesian methods are used to test it. Although this formulation is seen as one of the success stories of philosophical Bayesianism (see Earman, 1992, pp. 63-65), I demonstrate that it is of little use to the practical researcher. The problem is once again the deductive structure of theory confirmation in political science.

The other use of Bayesian confirmation is inductive. Given an inclusive set of theories, Bayesian methods could be used to pick the theory most supported by the data. This formulation is also problematic, but the proposed solution will be instructive. I begin with the deductive case.

Figure 2
The Logic of Conventional Confirmationism

1. Theory to hypothesis

$$T \wedge K \rightarrow H_1$$

2. Data to hypothesis

$$\left.\begin{array}{r} H_1 \rightarrow \text{Coefficient is correct} \\ \text{Coefficient is correct} \end{array}\right\} \nvdash H_1$$

3. Hypothesis to theory

$$\left.\begin{array}{r} T \wedge K \rightarrow H_1 \\ H_1 \end{array}\right\} \nvdash T \wedge K$$

Summary:

We falsely conclude that the theory is true, based on incorrect inferences in (2) and (3).

Bayesian Confirmation and the Deductive Method

Bayes's Theorem, to review briefly, is

$$P(H|y) = \frac{P(H)P(y|H)}{P(y)} \propto P(H)P(y|H).$$

That is, the probability of the hypothesis given the data, $P(H|y)$, is proportional to the prior probability of the hypothesis, $P(H)$, multiplied by the likelihood of the data given the hypothesis, $P(y|H)$.[12]

The strength of the Bayesian programme is that it provides the inverse probability statement (the probability of the hypothesis given the data, $P(H|y)$, as opposed to the probability of the data given the hypothesis, $P(y|H)$) that helps us avoid the logical fallacy in the "data to hypothesis" step of Figure 2. Because Bayesian statistics provide the probability of the research hypothesis, H_1, given the data, y, we can validly conclude H_1.[13] The difference between conventional confirmationism and Bayesian confirmation is in Step 2, as follows:

$$\left.\begin{array}{r} H_1 \rightarrow \text{Coefficient is correct,} \\ \text{Coefficient is correct} \end{array}\right\} \nvdash H_1, \quad \left.\begin{array}{r} H_1 \rightarrow P(H_1 \mid y) \text{ is high,} \\ P(H_1 \mid y) \text{ is high} \end{array}\right\} \vdash H_1.$$

We do not commit the fallacy of affirming the consequent here, because unlike conventional statistics, the Bayesian statistics tell us the probability that the research hypothesis is true given the data, $P(H_1 \mid y)$. The Bayesian strategy is outlined in Figure 3.

Figure 3
Bayesian Confirmation

Long Form

1. Theory to hypothesis

$$T \wedge K \rightarrow H_1$$

2. Data to hypothesis

$$\left.\begin{array}{r} H_1 \rightarrow P(H_1 \mid y) \text{ is high} \\ P(H_1 \mid y) \text{ is high} \end{array}\right\} \vdash$$

3. Hypothesis to theory

$$\left.\begin{array}{r} T \wedge K \rightarrow H_1 \\ H_1 \end{array}\right\} \vdash T \wedge K$$

Summary:

We validly conclude that the theory is true based on the inverse probability statement.

Reduced Form

1. Theory to data

$$\left.\begin{array}{r} T \wedge K \rightarrow P(H_1 \mid y) \text{ is high} \\ P(H_1 \mid y) \text{ is high} \end{array}\right\} \vdash T \wedge K$$

Summary:

We validly conclude that the theory is true based on the inverse probability statement

In Step 1, the hypothesis is deduced from the theory, and in Step 2, we conclude that the hypothesis is true based on the inverse probability statement. In Step 3, it appears that we still commit the fallacy of affirming the consequent just as in conventional confirmationism. Note, however, that Step 1 ends with H_1 and Step 2 begins with H_1. We can then substitute for H_1 and write the logic in its reduced form as in the right-hand column of Figure 3. The logic is sound, and confirmation appears possible.

Unfortunately, this rosy scenario fails to hold up when we take a deeper look at Bayesian confirmation. Assume that the probability of the hypothesis is between zero and one, $0 < P(H) < 1$, and the probability of the data is between zero and one, $0 < P(y) < 1$. Furthermore, note that when the hypothesis implies the data, $H \rightarrow y$, the probability of the data given the hypothesis $P(y \mid H)$ is equal to one (Earman, 1992, p. 64).[14] Under these conditions, Bayes' theorem simplifies to

$$P(H|y) = \frac{P(H)P(y|H)}{P(y)} = \frac{P(H)}{P(y)}.$$

Thus, the inverse probability statement on which we have pinned our hopes reduces to the ratio of the prior probability of the hypothesis to the probability of the data. That the probability of the hypothesis given the data is proportional only to the prior probability of the hypothesis is of no concern to the philosophical Bayesian; it is of primary concern, however, to the substantive researcher. A political scientist choosing a noninformative prior would receive no confirmation, and one choosing an informative prior would be fully determining the amount of confirmation. No political scientist could get away with claiming confirmation on this basis. Without being comparative, meaningful theory confirmation is not possible.

Inductive Bayesian Confirmation

Using Bayesian methods, however, does not restrict us to the deductive model. We could lay out all the theories to be considered and let the data, along with our priors, pick the most supported theory. Bayes' theorem, then, would take the following form,

$$P(T_i|y) = \frac{P(T_i)P(y|T_i)}{\sum_{j=1}^{n}[P(T_j)P(y|T_j)]}.$$

The denominator of Bayes' theorem is the theorem of total probability, which states, when there are only two possible theories to consider,

$$P(y) = P(T)\,P(y|T) + P(\neg T)\,P(y|\neg T)$$

In any actual research situation, however, there are always more than two possibilities. If we allow that the theory under consideration, T, is unlikely to be the true theory, then the $\neg T$ term, which is a disjunction of alternative theories, must contain the true theory. To make sense of this, Shimony referred to the "catchall hypothesis" that contains all the unspecified or unknown theories that have some explanatory power regarding the dependent variable (Salmon, 1990, p. 179). $\neg T$, then, is a disjunction of seriously considered alternative theories and the "catchall," $\neg T \equiv T_1 \vee T_2 \vee T_3 \vee T_4 \vee T_{n-1} \vee H_c$.

Earman (1992, p. 168) writes, "the catchall H_c says, in effect, that some as yet uninvented theory is true." Salmon (1990) argues that "among the hypotheses hidden in the catchall are some that, in conjunction with present available background knowledge, entail the present evidence" (p. 191). Both philosophers contend that calculating $P(y|\neg T)$ is not possible.

If $P(y|\neg T)$ cannot be evaluated, then Bayesian induction is not possible. As we cannot know the likelihoods of all the possible theories, particularly those in the catchall hypothesis, we cannot calculate the denominator of Bayes' theorem. Salmon (1990) proposes a solution to this impasse that provides warrant for comparative theory confirmation. Salmon demonstrates that if we restrict the theories under consideration to two, then we can avoid the $P(y|\neg T)$ term. If we write out Bayes' theorem for each of two theories, we see that the denominators are equal, and we can therefore take their ratio,

$$P(T_1|y) = \frac{P(T_1)P(y|T_1)}{P(T_1)P(y|T_1) + P(T_2)P(y|T_2)},$$

$$P(T_2|y) = \frac{P(T_2)P(y|T_2)}{P(T_1)P(y|T_1) + P(T_2)P(y|T_2)},$$

$$\frac{P(T_1|y)}{P(T_2|y)} = \frac{P(T_1)P(y|T_1)}{P(T_2)P(y|T_2)}.$$

The above ratio does not tell us the degree to which the data support Theory 1 or Theory 2. Rather, the ratio tells us the degree to which the data support Theory 1 over Theory 2.[15] Salmon's solution is not a method for confirming a theory in isolation, but rather a method for confirming one theory relative to another. In the next section, I show how we can use Salmon's solution to fix the problems generated by deductive theory testing. Then in the "Quantitative Theory Comparison" section, I show how a substantive researcher can compare theories legitimately.

The Necessity of Being Comparative

We have seen that theory confirmation in the falsificationist and conventional confirmationist cases is defeated by the deductive nature of political science theory testing. A theory might imply a research hypothesis, but support for that hypothesis does not confirm the theory. By the same token, the research hypothesis might imply that a certain pattern should hold in the data, but finding that pattern does not confirm the hypothesis. Bayesian confirmation theoretically overcomes the problem, but it is practical only if knowing or specifying the prior probability of a research hypothesis presents no issues, which is to say never.

Resolving the impasse does not require abandoning the deductive nature of social scientific work. Rather, it requires being comparative. I show this in two steps. First, I show that the problem disappears when a theory is not just a necessary condition for the research hypothesis, $T \rightarrow H_1$, but a necessary and sufficient condition for the research hypothesis. That is, the theory must be a necessary and sufficient condition for the hypothesis to hold, and the hypothesis must be a necessary and sufficient condition for the theory to hold. Second, I show that this condition will hold only when we are being comparative. Thus, by being comparative, we can unproblematically retain the deductive nature of our work.

To achieve Step 1 formally, we make use of the biconditional, $\leftrightarrow$, which is interpreted as "if and only if." Making this change, we can conclude that the theory implies the data,

$$\left.\begin{array}{r} T \wedge K \leftrightarrow H_1, \\ H_1 \rightarrow D \end{array}\right\} \vdash T \wedge K \rightarrow D.$$

Unfortunately, this conclusion is the same one that we came to in the analysis of confirmationism. If the data exist in the correct form, we still cannot conclude that the theory is true,

$$\left.\begin{array}{r} T \wedge K \rightarrow D, \\ D \end{array}\right\} \nvdash T \wedge K.$$

To find a sufficient condition for confirmation, we need to go a step further and argue that the data follow a certain pattern if and only if the hypothesis is true. Modifying the above with the additional biconditional, we get

$$\left.\begin{array}{r} T \wedge K \leftrightarrow H_1, \\ H_1 \leftrightarrow D \end{array}\right\} \vdash T \wedge K \leftrightarrow D.$$

If the theory is a necessary and sufficient condition for the research hypothesis, and the research hypothesis is a necessary and sufficient condition for the data to follow an expected pattern, then the theory is a necessary and sufficient condition for the data to follow an expected pattern.

By combining the two biconditionals, we achieve a necessary and sufficient condition for confirming a theory. A theory may be confirmed if and only if the theory is a necessary and sufficient condition for the data to follow an expected pattern. Thus, we can confirm the theory without affirming the consequent,

$$\left.\begin{array}{r} T \wedge K \leftrightarrow D \\ D \end{array}\right\} \vdash T \wedge K.$$

This logic raises the question of when in political science a statement could be made that the data follow an expected pattern if and only if the theory is true. Even in the physical sciences such statements are rare.[16] Salmon's (1990) result comes into play at this point. We can make such statements only when we are comparing a theory with a direct rival.[17]

Consider two theories, T_1 and T_2, and a null hypothesis, H_0, that states there is no difference between the theories. In Step 1 of the logic, the null hypothesis holds if and only if the two theories are equivalent,

1. Theory to Hypothesis

$$(\mathrm{T}_1 \wedge \mathrm{K}) \equiv (\mathrm{T}_2 \wedge \mathrm{K}) \leftrightarrow \mathrm{H}_0.$$

In Step 2 of the logic, some statistic, τ, takes a value of 0 (or some equivalent value that means the same thing) if and only if the null hypothesis is true. If τ does not equal 0, we reject the null hypothesis,

2. Data to Hypothesis

$$\left.\begin{array}{r} H_0 \leftrightarrow \tau = 0 \\ \tau \neq 0 \end{array}\right\} \vdash \neg H_0.$$

Given this result, we can confirm, in Step 3, that one theory is more supported by the data than the other theory,

3. Hypothesis to Theory

$$\left.\begin{array}{r} (T_1 \wedge K) \equiv (T_2 \wedge K) \leftrightarrow H_0 \\ \neg H_0 \end{array}\right\} \vdash (T_1 \wedge K) > (T_2 \wedge K).$$

We can also write this logic in reduced form as Step 1 ends with H_0 and Step 2 begins with H_0, *4. Theory to Data*

$$\left.\begin{array}{r}(T_1 \wedge K) \equiv (T_2 \wedge K) \leftrightarrow \tau = 0 \\ \tau \neq 0\end{array}\right\} \vdash (T_1 \wedge K) > (T_2 \wedge K).$$

The two theories are equivalent if and only if the statistic, τ, takes a value of 0. When τ is not 0, we can validly conclude that Theory 1 is better supported by the data than Theory 2.

Thus, it is possible to make positive statements about the amount of support a theory receives relative to another theory. Being comparative is necessary when a researcher wishes to make a positive claim for his or her theory. We must remember, however, that passing such a test does not mean that the theory in question is true. It simply means that the theory is better than the available alternatives.

Quantitative Theory Comparison

How do we go about comparing theories in such a way as to be compatible with the logic discussed above? Comparative theory tests come in two flavors: nested and nonnested. One theory is nested within a second theory when the first theory is a special case of the second theory. Two theories are nonnested when neither theory can be expressed as a special case of the other model (Kmenta, 1986). I deal with each case separately and stick to linear regression for ease of presentation.[18]

Comparing Nested Theories

Consider two regression models,[19]

$$\text{Model 1: } \mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon},$$
$$\text{Model 2: } \mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon},$$

where $\mathbf{X}_1$ is $n \times k_1$, $\mathbf{X}_2$ is $n \times k_2$, $\boldsymbol{\beta}_1$, is $k_1 \times 1$, $\boldsymbol{\beta}_2$ is $k_2 \times 1$, and $\mathbf{y}$ and $\boldsymbol{\varepsilon}$ are $n \times 1$. Assume the **X**s are nonstochastic and that all classical assumptions are met.

Model 1 is nested within Model 2 because Model 1 is Model 2 when $\boldsymbol{\beta}_2 = \mathbf{0}$. The easiest way to test Model 1 against Model 2 is to test whether or not $\boldsymbol{\beta}_2$ is equal to zero, $H_0{:}\boldsymbol{\beta}_2 = 0$. The logic of this test meets our criteria,

1. Theory to hypothesis

$$\text{Model 1} \equiv \text{Model 2} \leftrightarrow \text{H}_0$$

2. Data to hypothesis

$$\left.\begin{array}{r} H_0 \leftrightarrow \boldsymbol{\beta}_2 = 0 \\ \boldsymbol{\beta}_2 \neq 0 \end{array}\right\} \vdash \neg H_0$$

3. Hypothesis to theory

$$\left.\begin{array}{r} \text{Model 1} \equiv \text{Model 2} \leftrightarrow H_0 \\ \neg H_0 \end{array}\right\} \vdash \text{Model 1} < \text{Model 2}.$$

In Step 1, the null hypothesis holds if and only if the two models are equivalent. In Step 2, $\boldsymbol{\beta}_2$ is equal to zero if and only if the null hypothesis is true. If we reject the null hypothesis, that result can be used in Step 3 to conclude that the rival models are not equivalent. Model 2 has more support from the data than Model 1 because the vector of coefficients on the additional variables included in Model 2 is not zero.

The actual test can be performed by the well-known F test,

$$F[J, n-K] = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})'\{\mathbf{R}[s^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{R}'\}^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})}{J},$$

where J is the number of restrictions (equal to k_2 in this case), $K = k_{1+} k_2$, s^2 is the estimate of σ^2, $\mathbf{R} = (\mathbf{0}_{k2\times k1}, \mathbf{I}_{k2})$, and $\mathbf{q}$ is a $k_2 \times 1$ vector of zeros. This test is trivial to perform and is familiar to most political scientists.

Comparing Nonnested Theories

Choosing between nonnested theories is generally thought to be harder than choosing between nested theories. The test I discuss here, however, is quite simple.

Consider again two regression models,

$$\text{Model 1: } \mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon},$$
$$\text{Model 2: } \mathbf{y} = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon},$$

where the assumptions are the same as in the previous section. These models are nonnested as neither model is a special case of the other. Artificially nesting the models and then using an F test is not a feasible approach particularly if $\mathbf{X}_1$ and $\mathbf{X}_2$ have variables in common (Greene, 2003, p. 154).

The test I discuss was proposed by Clarke (2003). If we estimate Model 1 and Model 2 with maximum likelihood, the log-likelihoods that are reported with each output are the sum of the log-likelihoods for each individual observation.

If we take the individual log likelihoods from Model 1 and subtract them from the individual log likelihoods from Model 2, half of the resulting differences should be greater than zero and half should be less than zero if the models are equivalent. Thus, the test simply asks whether or not the median of these differences is equal to zero.

The null hypothesis, then, is that the median of the differences in the log likelihoods is equal to zero, $\pi_{0.5} = 0$. The null hypothesis holds if and only if the models are equivalent. The logic of this test also meets our criteria,

1. Theory to hypothesis

$$\text{Model 1} \equiv \text{Model 2} \leftrightarrow \text{H}_0$$

2. Data to hypothesis

$$\left.\begin{array}{r} H_0 \leftrightarrow \pi_{0.5} = 0 \\ \pi_{0.5} \neq 0 \end{array}\right\} \vdash \neg H_0$$

3. Hypothesis to theory

$$\left.\begin{array}{r} \text{Model 1} \equiv \text{Model 2} \leftrightarrow H_0 \\ \neg H_0 \end{array}\right\} \vdash \text{Model 1} < \text{Model 2}.$$

Performing this test is quite simple. Once the differences have been computed, simply count the number of differences greater than zero. If the null hypothesis is true, roughly half should be greater than zero and roughly half less than zero. The number of positive differences is distributed as a Binomial (n, $\theta = 0.5$). (For more on these tests, see Clarke 2001, 2003.)

Thus, comparative theory testing is both eminently possible and consistent with the logic laid out in "The Necessity of Being Comparative" section. Two theories are equivalent if and only if the null hypothesis of no difference is true. The null hypothesis of no difference is true if and only if the data follow a certain pattern: namely, that a particular statistic takes a value of zero (or some equivalent value). If the data do not follow this pattern, then the null hypothesis must be false, which, in turn, means that the theories are not equivalent. One theory, therefore, receives greater support from the data than the other.

Conclusion

Theory confirmation in political science must be a three-way relation between two theories and the data. Being comparative is necessary. Confirming a theory in the absence of its rivals is not possible under falsificationism, conventional confirmationism, or even Bayesian confirmationism. When attempting to falsify or conventionally confirm a theory, we commit the logical fallacy of affirming the consequent. When attempting Bayesian confirmation, we see that the deductive connection between the theory and the hypothesis means that confirmation comes down to the specification of the prior, which is hardly a practical solution. Dealing with the deductive structure of theory confirmation requires the ability to state that a theory is true if and only if a certain pattern holds in the data. Such statements can only be made when one theory is compared with another theory. Meaningful theory confirmation, therefore, can occur only when political science abandons the goal of confirming individual theories and embraces the goal of relative confirmation.

Given the wide acceptance that comparative accounts of scientific progress have received in political science and the fact that some political scientists already embrace comparative theory confirmation, this change should not be disconcerting. Kuhn "often maintained that in actual science the problem is never to evaluate one particular hypothesis or theory in isolation; it is always a matter of choosing from among two or more viable alternatives" (Salmon, 1990, p. 191). Confirming that one theory receives more support from the data than a second theory represents real work and real progress.

Appendix
The Logical Operators

Implication

The symbol for implication is the arrow, $\rightarrow$, which stands for an "if . . . then" statement. $A \rightarrow B$ is read as "If A is true, then B is true." I use the implication symbol to make statements such as, "If some hypothesis is true, then we should observe some statistical relation in the data" or $H \rightarrow D$.

Conjunction and Disjunction

Conjunctions are "and" statements, and they use the symbol $\wedge$. $A \wedge B$ means "both A and B are true." Disjunctions are "or" statements, and they use the symbol, $\vee$. $A \vee B$ means "A is true or B is true or both are true." I use conjunction to symbolize a theory and its auxiliary hypotheses, $T \wedge K$, and I use disjunction to symbolize a group of theories, one of which must be true, $T_1 \vee T_2 \vee T_{n-1}$.

Negation

$\neg$ is the symbol for negation. $\neg A$ stands for "A is false" or "not A" or "it is not the case that A" and $\neg(A \vee B)$ stands for "not A or not B." I use negation to stand for the falsity of a theory or hypothesis, $\neg T$ or $\neg H$.

Logical Validity

The symbol $\vdash$ means "is logically valid" or "therefore" (Lemmon, 1992). If we assume $A \rightarrow B$ and A, then we can logically conclude B. A valid deduction, then, takes the form $A \rightarrow B, A \vdash B$.

If the deduction is not valid, I will use $\nvdash$, $A \rightarrow B, B \nvdash B$. If we assume $A \rightarrow B$ and B, then we cannot logically conclude A.

Biconditional

The symbol for the biconditional is the double arrow, $\leftrightarrow$, which means "If A is true, then B is true" and "If B is true, then A is true." My argument for comparative theory testing relies on two biconditionals: one from the theory to the hypothesis, $T \leftrightarrow H$, and one from the hypothesis to the data, $H \leftrightarrow D$.

Equivalence

$\equiv$ is the symbol for equivalence. $A \equiv B$ means "A and B are either both true or both false." I use the equivalence operator for definitional purposes, $\neg(A \wedge B) \equiv (\neg A \vee \neg B)$; that is, "not A and B is equivalent to not A or B."

Notes

1. To be clear, by *theory*, I mean "a general, systematic account of a subject matter" that constitutes an explanation (Boyd, Gasper, & Trout, 1993, p. 781). Hypotheses are statements

derived from theories that specify a particular relationship between one or more independent variable and the dependent variable (Brady & Collier, 2004, p. 289).

2. This discussion makes use of a number of symbols from formal logic; the Appendix contains a brief primer.

3. Auxiliary theories and background knowledge are any claim or information conjoined with the theory being tested to derive observable predictions from the theory (Boyd et al., 1993).

4. Probabilistic hypotheses are not falsifiable. For example, when flipping a coin n times, how many heads would have to appear before you could conclude whether or not the coin is fair? There is no answer to this question as any number of heads in n tosses is consistent with a fair coin. Fisher's innovation was to "falsify" a statistical hypothesis when the data, given the hypothesis, were unlikely (Howson & Urbach, 1993, p. 174).

5. To be clear, falsificationism is often used in two different ways. Most political scientists are falsificationists in the original sense—falsifiability being the demarcation between scientific theories and nonscientific theories. Few political scientists are falsificationists in the theory testing sense.

6. Classical statistics, as opposed to Bayesian statistics, provide the probability of the data given the hypothesis, $P(y \mid H_0)$. This is the reason for the notation used in Figure 1.

7. Technically, rejecting the null hypothesis has no implications for the alternative or research hypothesis. Substantive researchers routinely ignore this stricture, however.

8. Arguments for falsificationism can be quite nuanced; see Caporaso (1995, p. 458), for example. However complex the relationship between the theory and the hypothesis, though, the logic of Figure 1 still holds when the hypothesis is tested.

9. The term *conventional* is used to distinguish this version of confirmationism from Bayesian confirmationism.

10. Case 3 could be written in a reduced form. As Step 1 ends with H_1 and Step 2 begins with H_1, we could "cut out the middle man" and write,

Theory to data

$$\left.\begin{array}{r} T \wedge K \rightarrow \text{Coefficient is correct} \\ \text{Coefficient is correct} \end{array}\right\} \nvdash T \wedge K.$$

I use the longer form to maintain compatibility with cases 1 and 2.

11. There is a temptation to combine Step 1 from the confirmationist cases, $T \wedge K \rightarrow H_1$, with Step 2 from the falsificationist cases, $H_0 \rightarrow P(y \mid H_0)$ *is high*, but this strategy cannot work. The premises in Step 3 would be $T \wedge K \rightarrow H_1$ and $\neg H_0$, but there is no necessary connection between the falsity of H_0 and H_1.

12. In terms of a hypothesis, background knowledge, and data, Bayes' theorem takes the following form, $P(H \mid y \,\&\, K) = P(H \mid K)P(y \mid H \,\&\, K) \,/\, P(y \mid K)$. I leave out the Ks in the main discussion for notation simplicity.

13. The first premise cannot be written $H_1 \leftrightarrow P(H_1 \mid y)$ because other hypotheses such as $H_1 \,\&\, I$, where I is anything you like, also imply $P(H_1 \mid y)$. The data would confirm these hypotheses as well, but to a lesser extent.

14. A simple example: if being human implies being mortal, $H \rightarrow M$, then the probability of being mortal given being human, $P(M \mid H)$, is equal to 1.

15. $P(y \mid T_i)$ does not go to 1 in this case, as we are not working with the deductive method.

16. Einstein's theory of relativity produced two spectacular examples: the perihelion of Mercury and the bending of light around the sun. Even in Einstein's case, however, things are not so simple. Other theories that make the same prediction could be constructed by using the connector "and" to add irrelevant propositions. (My thanks to Bob Pahre for calling my attention to these examples.)

17. Including hypotheses from different theories in the same equation and testing them individually does not qualify as theory comparison. Statistical equations are not hypothesis-testing machines in which a researcher can include any hypothesis he or she wishes to test. In what follows, I assume that statistical equations reflect a single theory.

18. The point here is not to provide an in-depth discussion of these tests, but rather to demonstrate that comparative theory testing is compatible with the solution in the previous section.

19. In the discussion to follow, I assume that the statistical model reflects the theory accurately.

References

Boyd, R., Gasper, P., & Trout, J. (Eds.) (1993). *The philosophy of science*. Cambridge, MA: MIT Press.

Brady, H. E., & Collier, D. (Eds.) (2004). *Rethinking social inquiry: Diverse tools, shared standards*. Oxford, UK: Rowman & Littlefield.

Caporaso, J. (1995, June). Research design, falsification, and the qualitative-quantitative divide. *American Political Science Review*, *89*(2), 457-460.

Chalmers, A. (1982). *What is this thing called science?* St. Lucia, Australia: University of Queensland Press.

Clarke, K. A. (2001, July). Testing nonnested models of international relations: Reevaluating realism. *American Journal of Political Science*, *45*(3), 724-744.

Clarke, K. A. (2003, February). Nonparametric model discrimination in international relations. *Journal of Conflict Resolution*, *47*(1), 72-93.

Collier, D. (1995, June). Translating quantitative methods of qualitative researchers: The case of selection bias. *American Political Science Review*, *89*(2), 461-466.

Conley, R. S. (2006, June). From Elysian fields to the guillotine? *Comparative Political Studies*, *39*(5), 570-598.

Earman, J. (1992). *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: MIT Press.

Fox, J. (2006, June). World separation of religion and state into the 21st century. *Comparative Political Studies*, *39*(5), 537-569.

Greene, W. H. (2003). *Econometric analysis* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Howson, C., & Urbach, P. (1993). *Scientific reasoning: The Bayesian approach* (2nd ed.). Chicago: Open Court.

Johnson, J. (2006, March). Consequences of positivism: A pragmatist assessment. *Comparative Political Studies*, *39*(2), 224-252.

King, G., Keohane, R.O., & Verba, S. (1994). *Designing social inquiry*. Princeton, NJ: Princeton University Press.

Kmenta, J. (1986). *Elements of econometrics* (2nd ed.). New York: Macmillan.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

Laitin, D. D. (1995, June). Disciplining political science. *American Political Science Review*, *89*(2), 454-456.

Lakatos, I. (1970). Falsification and the methodology of scientific research programs. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge, Vol. 4 of International colloquium in the philosophy of science*. Cambridge, UK: Cambridge University Press.

Lake, D. A., & Baum, M. A. (2001, August). The invisible hand of democracy. *Comparative Political Studies*, *34*(6), 587-621.

Laudan, L. (1977). *Progress and its problems: Towards a theory of scientific growth*. London: Routledge and Kegan Paul.

Lemmon, E. (1992). *Beginning logic*. Indianapolis, IN: Hackett.

Levey, G. B. (1996, March). Theory choice and the comparison of rival theoretical perspectives in political sociology. *Philosophy of the Social Sciences*, *26*(1), 26-60.

Miller, R. W. (1987). *Fact and method: Explanation, confirmation and reality in the natural and the social sciences*. Princeton, NJ: Princeton University Press.

Rogowski, R. (1995, June). The role of theory and anomaly in social-scientific inference. *American Political Science Review*, *89*(2), 467-470.

Rohrschneider, R. (2005, September). Institional quality and perceptions of representation in advanced industrial democracies. *Comparative Political Studies*, *38*(7), 850-874.

Rudra, N., & Haggard, S. (2005, November). Globalization, democracy, and effective welfare spending in the developing world. *Comparative Political Studies*, *38*(9), 1015-1049.

Salmon, W. (1990). Rationality and objectivity in science, or Tom Kuhn meets Tom Bayes. In C. Savage (Ed.), *Scientific theories*, *Vol. 14 of Minnesota studies in the philosophy of science*. Minneapolis: University of Minnesota Press.

Stokes, S. C. (2005, August). Perverse accountability: A formal model of machine politics with evidence from Argentina. *American Political Science Review*, *99*(3), 315-325.

Kevin A. Clarke is an assistant professor in the Political Science Department at the University of Rochester. He is a political methodologist with interests in quantitative theory comparison, philosophy of science, and international relations.