The Rule of Reproducibility and Its Applications in Experiment Appraisal
Author(s): Xiang Chen
Source: *Synthese*, Vol. 99, No. 1, Experiments and Scientific Change (Apr., 1994), pp. 87-109
Published by: Springer
Stable URL: http://www.jstor.org/stable/20117888
Accessed: 02/12/2009 15:49

XIANG CHEN

THE RULE OF REPRODUCIBILITY AND ITS APPLICATIONS IN EXPERIMENT APPRAISAL

1. INTRODUCTION

Many students of science agree that reproducibility is the essential basis that guarantees both cognitively and socially the validity of experimental results. According to Popper, for example, reproducibility is a demarcation criterion for objective observational and experimental knowledge. He argues that "[w]e do not take even our own observations quite seriously, or accept them as scientific observations, until we have repeated and tested them. Only by such repetitions can we convince ourselves that we are not dealing with a mere isolated 'coincidence', but with events which, on account of their regularity and reproducibility, are in principle inter-subjectively testable" (1968, p. 45). Sociologists have also regarded reproducibility as the most important institutionalized norm for experiment appraisal. According to Zuckerman, "the institutionalized requirement that new contributions be reproducible is the cornerstone of the system of social control" (1977, p. 92). Some philosophers further suggest that the requirement of reproducibility should be understood as a mandatory rule for experimental practices. In particular, the rule of reproducibility can be expressed in the form of a hypothetical imperative: if one wants to produce experimental knowledge, then one ought to conduct experiments whose results are reproducible (Hones, 1990, p. 586). This suggests that repetition is a necessary procedure for examining the validity of experimental results, and that scientists will without exception apply the rule of reproducibility whenever they need to evaluate experimental findings.

However, the results of some recent studies cast doubt on the proposed normative or mandatory status of the reproducibility rule. Based on detailed analyses of experimental discoveries in contemporary physics, Franklin summarizes a set of strategies that scientists used in their practices to achieve validity in their experimental findings (1986, pp. 166–84). Among this set of strategies, repetition is only one of the possible means that, according to Franklin, are neither necessary nor

Synthese **99**: 87–109, 1994.

sufficient for the validation of experimental results (Franklin and Howson, 1988, p. 426). Also, based on interviews with a group of biochemists, Mulkay and Gilbert report that scientists only occasionally make efforts to repeat what somebody else did, and always have several conceptions of what a valuable or proper repetition should be (1986, p. 22). These empirical studies suggest that the role of the reproducibility rule in experiment appraisal and the process of experiment repetition are more complicated than some philosophers and sociologists expect.

The purpose of this paper is to examine the complexities involved in experiment repetition, and to explore the epistemological and social foundation of the reproducibility rule. In the following sections, I first illustrate the complexities in the application of the reproducibility rule by analyzing a historical case. This is the debate on the analysis of sunlight in the 1840s. The main themes of this debate were whether a particular experimental finding should be counted as experimental knowledge, and how the reproducibility rule should be properly applied to evaluate this experimental result. This historical episode vividly shows that the reproducibility rule was not mandatorily applied in experiment appraisal. The rule was only applicable under certain conditions, which were not logically defined by the rule itself but determined contextually by scientists.

I further examine how traditional philosophy of science and recent sociology of science interpret these complexities in experiment repetition. I argue that neither the traditional philosophical account which entirely relies on logical reasoning nor the recent sociological account which appeals to social conventions can provide an appropriate explanation of experiment appraisal. Finally, I explore the epistemological and social foundation for the application of the reproducibility rule from an alternative perspective. Inspired both by Wittgenstein's account of rule-following and by recent studies of categorization by cognitive psychologists, I provide a different interpretation for the practice of experiment appraisal in the historical case. I conclude that it is scientists' practices that provide a bedrock for the application of the reproducibility rule, and that the result of experiment appraisal relies on both cognitive and social factors.

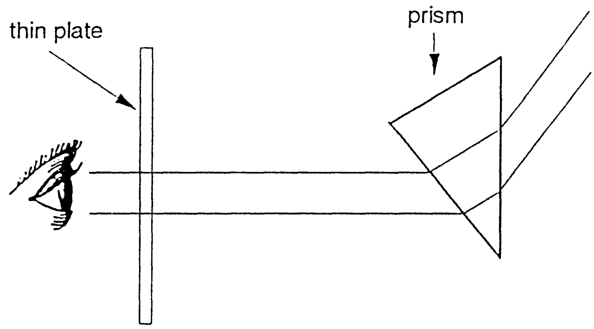

Fig. 1. Brewster's experiment on solar spectrum.

2. THE DEBATE ON THE ANALYSIS OF SUNLIGHT

By examining the solar spectrum produced by a prism, Newton, in the seventeenth century, concluded that there were seven primary colors in sunlight (1952, p. 126). However, Newton's conclusion on the colors of the solar spectrum was challenged by David Brewster in the 1830s. Brewster was one of the most prestigious British scientists in optics during the early nineteenth century, with an especially strong reputation in optical experiment. His contemporaries regarded him as "the father of experimental optics" (Whewell, 1967, Vol. 2, p. 373). In a paper presented to the Royal Society of Edinburgh in 1831 and published in its *Transactions* in 1834, Brewster presented a series of new experimental findings inconsistent with Newton's observations. In his experiments, Brewster examined the impact of absorbing materials on different rays of sunlight. His experimental apparatus included a prism and a plate of colored glass. A narrow beam of sunlight entered a dark room and passed through the prism; the spectrum emerged from the prism, passed through the plate of colored glass and directly into the eye of the observer (Fig. 1). When Brewster interposed a plate of purplish-blue glass, about one-twentieth of an inch thick, he found that all the orange and a large part of the green light in the spectrum disappeared. By varying the absorbing material, Brewster could eliminate the indigo and the violet light from the spectrum. According to Brewster, all of these results indicated that the orange, green, indigo, and violet colors in the spectrum were not primary. He concluded that the red, yellow, and blue rays were the only primary colors, and that all the others were

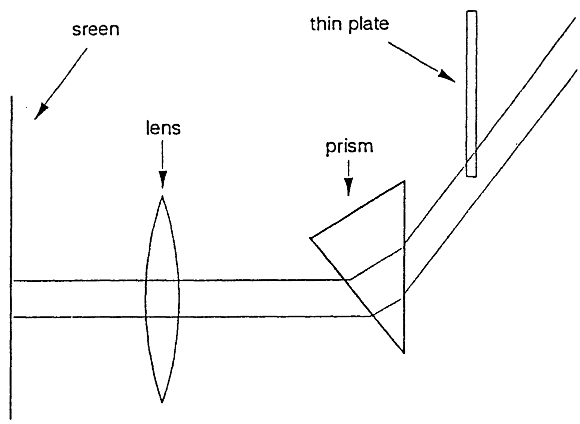

Fig. 2. Airy's experiment replication.

compounds, each of them consisting of red, yellow, and blue light in different proportions (1834, pp. 124–35).

Brewster's new analysis of sunlight relied entirely on his experimental findings: interpositions of absorbing materials caused changes of color in the spectrum. In 1833, George Airy, also one of the most reputable British scientists in optics during the early nineteenth century, repeated Brewster's experiment, attempting to examine these experimental findings directly. The experimental setting of Airy's repetition was, in principle, identical with that of Brewster's original experiment. However, Airy added a lens in front of the prism to obtain a better image of the spectrum. Instead of observing it directly with the eye, he used a piece of paper as a screen to receive the spectrum. Another significant difference was the position of the absorbing material. Airy placed the plate of colored glass between the light source and the prism, rather than between the prism and the observer, as in Brewster's experiment (Fig. 2). The results of this repetition, however, were negative. Under most circumstances, Airy said that he did not observe any change of color in the spectrum caused by absorbing materials. Although in two or three instances he found that the blue color was extended, these alterations disappeared when he prevented the screen from being illuminated by white light coming directly from the light source. Airy finally concluded that "no change was produced in the qualities of the colours" (1847, p. 75).

Although Airy obtained negative results in his experiment, he was reluctant to publish his findings. He only made an oral report of his experiments to the Cambridge Philosophical Society, in which he merely described his experimental findings, but did not connect them to Brewster's original experiments. Later he recalled:

> I never drew up in writing any orderly statement either of the experiments or of the reasoning connected with them. I had the pencil notes, written each at the instant of making an experiment; and from these notes, or rather from the strong recollection of every experiment detailed in the notes, I made my oral statement (as far as the mere experimental facts were concerned) to the Cambridge Philosophical Society. (Ibid., p. 73)

After the oral presentation, Airy did not preserve the penciled notes. "These notes I have mislaid", he said, and later could not find them (Ibid.).

Airy's reluctance to publish his results indicates, perhaps, that he had doubts about his own findings. Although he was well known as a theoretical analyst and mathematical calculator, Airy was not an authority in experimental optics. Also, this replication was his first experimental work studying the impact of absorbing materials on light. By contrast, Brewster had conducted experiments in this field since the early 1820s. Thus, the members of the scientific community would be likely to interpret the negative results of Airy's repetition not as a challenge to Brewster's experiments, but as evidence of Airy's own failure. The fate of Airy's notes indicates that Airy himself may have interpreted the results of his repetition in this way. Airy's doubts about his own experiment suggest, perhaps, that he was concerned about the qualifications of experimenters, both the original and the repeater, in his practice of experiment appraisal.

The first open criticism of Brewster's experimental findings came from William Whewell, in his *History of the Inductive Sciences* published in 1837. In a brief footnote, Whewell mentioned that Brewster's experimental result, that absorbing materials could change some colors in the spectrum, "has, however, been denied by other experimenters" (1967, Vol. 2, p. 287). Whewell here referred to Airy's experiment, which was unknown to Brewster. But Whewell did not reveal Airy's name, nor any details of the experiment. Whewell's remark made Brewster uncomfortable. In a review of Whewell's History, Brewster asked Whewell to give the details of the alleged experimental denial (Brewster, 1837, p. 72). The complaint from Brewster forced Whewell to release more information about Airy's work. In the second edition of his *History* (published in 1847), Whewell wrote that "Mr. Airy repeated [Brewster'] experiments with about thirty different absorbing substances, and could not satisfy himself that in any case they changed the colour of a ray of given refractive power" (1967, Vol 2, p. 288). At the same time, Whewell asked Airy to publicize his results. In response to Whewell's request, Airy published a paper in the *Philosophical Magazine* (in 1847), entitled 'On Sir David Brewster's New Analysis of Solar Light', providing details of his repetition of Brewster's experiment.

In his 1847 paper, Airy's confidence in his experiment conducted more than a decade ago, increased dramatically. He claimed that "I have no hesitation in saying that no form of experiment anterior to my own has been such as to place its conclusions beyond doubt" (1847, p. 76). When Airy had just completed his repetition, he clearly had doubts about the results of the experiment. Apart from the verbal report to the Cambridge Philosophical Society, he decided to keep his results private, and gave up his attempt to challenge Brewster by employing the reproducibility rule. But fourteen years later, he firmly believed that the results of his repetition were not open to doubt and that they could be used to challenge Brewster's experimental findings, despite the fact that he had not kept any written record of his experiment.

That Airy in 1847 became absolutely confident in his own experiment reflected a dramatic change in the field of optics. By the mid 1840s, the undulatory theory of light, the theory to which Airy was firmly committed, had convincingly demonstrated its explanatory superiority over its rival – the emission theory of light. Accompanying this change, Brewster's prestige decreased, because of his persistent support of the emission theory. On the other hand, Airy's influence and prestige rose considerably. For example, because his *Tract* on optics (1831) had become the textbook for the Mathematical Tripos at Cambridge, more and more Cambridge-trained physicists saw nothing but his version of the undulatory theory. Moreover, Airy in 1835 became the Astronomer Royal, one of the most prestigious scientific positions in nineteenth-century Britain. In this new situation, the obstacle that prevented Airy from employing the reproducibility rule to challenge Brewster in the early 1830s disappeared. In terms of experimenters' qualifications, Airy could now say that his competence was the same as, or even higher than, that of Brewster.

Although Airy's repetition now satisfied the condition concerning experimenters' qualifications, at the same time it encountered new troubles. After he made the oral presentation to the Cambridge Philosophical Society in 1833, Airy gave up his experimental research on the colors of sunlight. Because he had lost his only notes, he could not remember what kind of absorbing materials he used, nor could he recall how to combine them in his replication. His 1847 paper was entirely based on his recollection of events that had happened more than a decade previously. What Airy presented in 1847 was not a real repetition but a reconstruction. Airy openly admitted that this was a "partial imperfection". But he tried to persuade his readers that his reconstructed repetition based upon personal recollection was as reliable as the original. On this point, Airy did not give any substantial argument. Instead, he simply made a rhetorical statement, claiming that "upon the method, upon the results, and upon the reasonings, my recollection is as perfect as it was on the days on which the experiments were made". The only evidence Airy gave to support this statement was the order of presentation. He said that "I shall give my statement [in the 1847 paper] in the same order in which I gave it (in the year 1833, I believe) to the Cambridge Philosophical Society" (1847, p. 73). Although the order of presentation was not essential here, Airy hoped that by emphasizing this apparent identity of the original repetition and the reconstruction his readers would believe in the reliability of his recollection.

Not surprisingly, Airy's paper prompted a strong reaction from Brewster. In the same year, 1847, Brewster published a reply to Airy in the *Philosophical Magazine*, in which he both attacked Airy's repetition and defended his original experiment.

According to Brewster, the most formidable problem in Airy's account was that he built everything on his recollection. Brewster wrote that "no apology can be made for those who, with the means and the leisure for repeating their experiments, bring forward their recollections to discredit or to overturn the researches of others who have laboured patiently and successfully in the same field of scientific research" (1847a, p. 157). Brewster continued, stating that it was particularly wrong for Airy to base his experiment appraisal on his own recollection, because Airy himself confessed that he had no memory of colors.

According to Brewster, his opponent was "bound to repeat the identical experiments which he challenges, with similar apparatus and similar materials". If discrepancies were found in the process of these identical repetitions, the challenger should "inquire into the causes by which such discrepancies have arisen", and "establish his own views by new and effective experiments". After the causes were identified, the challenger also needed to justify his claim publicly, "to publish his researches in vindication of his charge against a fellow-labourer in science" (ibid., p. 155). These were the principles concerning the nature of the repeated experiment, and equally central for properly applying the reproducibility rule.

However, Brewster did not exactly follow these principles in his appraisal of Airy's repeated experiment. Brewster decided not to further repeat Airy's experiment, because Airy in his 1847 paper only reconstructed his repetition. If Airy's work had been a real repetition, Brewster could have designed another repetition with an identical experimental setting. But, facing a reconstructed repetition, Brewster was unable to figure out what exactly Airy's experimental arrangement was. Airy himself could not remember every detail, and even forgot what kind of absorbing materials he actually used. No matter what results Brewster came up with, Airy could easily deny Brewster's challenge by claiming that Brewster's repetition was not exactly identical with his own. A direct application of the reproducibility rule to Airy's experiment became impossible.

Instead of performing new experiments, Brewster simply applied these principles to criticize Airy's work. Brewster focused particularly on the identity of his original experiment and Airy's repetition. One salient difference in Airy's repetition was the position of the absorbing material. Instead of putting the plate of colored glass between the prism and the observer, as in Brewster's experiment, Airy placed it between the light source and the prism. He had argued that this different arrangement could make simultaneous comparison of the modified and unmodified spectrum, which was an essential improvement in observation accuracy because the eye had no memory of color. But Brewster pointed out that "it may be true that *the eye* has no memory of any kind, and therefore not for colours; but *I have a memory* for colours" (ibid., p. 154; original emphasis). A simultaneous comparison was not necessarily an improvement, and the different position of the absorbing material in Airy's repetition was not justified.

Brewster also pinpointed another difference in Airy's repetition. Instead of observing the spectrum directly with the eye, Airy used a piece of paper as a screen to receive the spectrum. This technique, however, would create some distortions, according to Brewster. He believed that, while Airy was viewing the paper screen, "his retina was influenced by all the various colours which shone in his modified and unmodified spectrum" (ibid., p. 58). Brewster's allegation was quite reasonable. In the mid 1840s, the technique of using a paper screen to receive images in optical experiments was replaced by direct observation of the image by the eye, thus eliminating distortions caused by reflection at the surface of paper. This technique of direct observation had been widely appreciated by first-rank researchers in optics, including Fresnel, and was actually employed in Brewster's experiments. By exposing the problems associated with Airy's technique of observing the spectrum, Brewster implied that the use of a paper screen was responsible for Airy's negative results.

Airy did not respond to Brewster's criticisms. Later, Brewster published several papers (1847b, 1848) to justify his original experiment on the change of colors in the solar spectrum; but publicly Airy kept silent. This suggests that Airy may have realized some of the problems in his repetition or may have accepted some of Brewster's arguments. Because Airy did not object to Brewster's principles concerning the appropriate conditions of repetition, it would be difficult for him to make substantial counterattacks unless he had new experimental findings. In the early 1850s, the majority of the optical community agreed that the changes of colors in the solar spectrum caused by absorbing materials were physical facts. Helmholtz and later Maxwell, for instance, admitted that absorption could cause change of colors, although they had different interpretations of how these changes might happen (Brewster, 1965, Vol. 1, pp. 123–26; Larmor, 1971, Vol. 2, p. 21). The debate between Brewster and Airy finally ended and a consensus among the majority of the optical community was reached in Brewster's favor.

3. METHODOLOGICAL RULES AND SOCIAL CONVENTIONS

The debate between Brewster and Airy on the change of colors in the solar spectrum demonstrates that the process of experimental repetition, or the application of the reproducibility rule, was much more complicated than we might expect. Contrary to some existing accounts of science, the reproducibility rule did not have a mandatory status in this instance of experiment appraisal. In this historical episode, although both Brewster and Airy agreed that repetition was the essential method for testing experimental findings, neither felt compelled to obey the reproducibility rule. This was so because the reproducibility rule itself did not specify the appropriate conditions for its application. Instead, it was the actors who decided whether the reproducibility rule was applicable in a particular case. A good example was Brewster's decision not to test Airy's work by a physical repetition. After the rule was known as in principle applicable, actors still needed to determine whether applying the rule to the particular case was appropriate. If not, they could terminate the application half way, just as Airy had done in the early 1830s.

These complexities in applying the reproducibility rule lead to some serious questions. If scientists are free to decide whether, when, and how to apply the rule, how could they achieve objective conclusions in their appraisals of experiments? Or, in our historical episode, how could Airy's replication finally be rejected on a unanimous basis? More important, if the reproducibility rule is not mandatory and if its applications are not defined by the rule itself, what is the basis for its applications? These are questions about the foundation for applying the reproducibility rule.

The answer to these questions, provided by the traditional philosophy of science, is to identify or to construct some other rules that provide the guidelines for the applications of the reproducibility rule. According to Popper, for example, all scientific activities are governed by rules. He claims that "[j]ust as chess might be defined by the rules proper to it, so empirical science may be defined by means of its methodological rules" (1968, pp. 82–83). Hence, we should be able to construct a group of rules that specifies the appropriate conditions for the applications of the reproducibility rule, with the expectation that these newly constructed rules will become the bedrock for its applications. In the debate between Brewster and Airy, for example, the appropriate conditions for applying the reproducibility rule could be specified by a group of supplementary rules about the appropriateness of repeated experiments. These supplementary rules stipulate that a proper replication should, say, be identical with the original, and be conducted by qualified experimenters.

The logical interpretation of rule-following involves a serious theoretical difficulty. This is the problem of the "vicious regress" that it may create (Barker, 1988, p. 100). For the sake of argument, we may express the reproducibility rule as follows: If one conducts experiments whose results are reproducible, one produces experimental knowledge. Now, if we want to introduce a new rule for the application of the reproducibility rule, the new rule should first specify what a reproducible result looks like. We may express this new rule in a specific form, say, the following conditional sentence:

> *Rule 1:* If the same result is obtained through an appropriately repeated experiment, this experimental result is reproducible.

Now, the question of applying the rule of reproducibility becomes the question of specifying the conditions for an appropriately repeated experiment. If we insist on using rules to solve this problem, we should expect a second rule to determine the appropriateness of a repeated experiment. We also can express this rule in a specific conditional sentence:

> *Rule 2:* If the experimental setting in a repeated experiment is identical with the one in the original experiment, this repeated experiment is appropriate.

However, the introduction of the second rule does not solve the problem. Before we can apply the second rule to any case, we need to know how we can judge the identity of two objects, or two actions, that happen in different times and different places. We need a third rule to specify the conditions of identity. No matter what the third rule says, we can expect that a new rule is needed for specifying the conditions of its application. This is a "vicious regress"; an infinite number of rules is needed for the application of the reproducibility rule and for any actual experiment appraisal. This, however, is an absolutely implausible picture for scientists' daily appraisal practices.

In his sociological studies of experimentation, Harry Collins also identifies the problem of vicious regress involved in a purely logical interpretation of rule-following, but with a different analysis. According to Collins, because of the skill-like nature of experimentation, the appropriateness of a replication, including the competence of the experimenter and the integrity of the experiment, can only be ascertained by examining its results. But appropriate results, in turn, can only be recognized if produced by competently performed replication. In his discussion of the controversy over the experiment about gravitational radiation, Collins writes:

> What the correct outcome is depends upon whether there are gravity waves hitting the Earth in detectable fluxes. To find this out we must build a good gravity wave detector and have a look. But we won't know if we have built a good detector until we have tried it and obtained the correct outcome! But we don't know what the correct outcome is until . . . and so on, *ad infinitum*. (1985, p. 84)

This is a vicious circle, which Collins labels the "experimenter's regress".

To solve this problem, Collins appeals to agreements that scientists reach in a process of negotiation, and to the constraints imposed by certain social conventions. Inspired by Wittgenstein's ideas on rule-following (1958, §§185–241), Collins and some other sociologists have developed a radically relativistic position on experiment appraisal. Collins, for example, argues that even a simple rule of arithmetic such as 'add a 2 and then another 2 and then another and so on' "doesn't fully specify what we are to do . . . because that instruction can be followed by writing '82, 822, 8222, 82222' or '28, 282, 2282, 22822', or 8^2, etc. Each of these amounts to 'adding a 2' in some sense" (1985, p. 13). From these different interpretations of the rule of 'adding a 2', a consensus can only be reached through negotiations, or through the social conventions that compel people to accept certain actions as right and others as wrong.[1] Similarly, David Bloor claims that "the application of formal principles is always a potential subject for informal negotiation" (1991, p. 133). In his study of the replications of experiments about gravitational radiation, Collins (1981) shows that the applications of the reproducibility rule and the process of replication first occur without well-defined standards. Scientists subsequently negotiate the conditions that specify proper experiment replication. These negotiations among scientists break the vicious regress and guarantee the success of replications. Collins's solution to the vicious regress, however, leads to a radical relativist and subjectivist interpretation of experiment appraisal. According to this interpretation, the negotiations among scientists and the relevant social conventions determine the standards for appropriate experiment replications and consequently the results of experiment appraisal, but natural or objective phenomena have no role at all in the appraisal processes. This is also an absolutely implausible picture for scientists' daily appraisal practices.[2]

In our historical case, we simply cannot find any trace of the negotiations among scientists or the constraints of social conventions that are supposed to bring about the final consensus. We just do not find evidence of negotiation between Brewster and Airy, or any other relevant actors, about what an appropriate experiment replication should look like. Evidence also indicates that, in the physical community of the early nineteenth century, there were no social conventions against what Airy did in his experiment replication. Reinterpreting experimental results and even building reinterpretations upon personal recollection were common among physicists in their experiment appraisals during this period. For example, between 1822 and 1825, John Herschel, one of the most distinguished scientists in the early nineteenth century, had conducted an experiment studying the connections among electricity, magnetism, and light. This experiment was not successful – Herschel did not find the expected connection between magnetism and light. Herschel initially attributed his failure to the weak battery he used. But more than two decades later, when he learned about Faraday's experiment on the magneto-optic effect in 1845, Herschel reinterpreted his own experimental results, although he no longer possessed any written record of his original experiment, and barely remembered when exactly it was done. On the basis of his personal recollection, Herschel reinterpreted the null result of his experiment as a piece of positive evidence to support Faraday's findings, showing a particular property of the magneto-optic connection. His reinterpretation was accepted by the community, because Herschel was able to persuade other physicists to undertake further experiments along his line (Gooding, 1989, pp. 67–70).

Collins's and Bloor's social interpretation of rule-following also has several theoretical difficulties. First, their appeal to social conventions encounters a similar problem of infinite regress as long as social conventions are understood as some kind of universal statements. A social convention or a social norm does not specify its application conditions; no matter what a social convention or a social norm says, we can expect that a new convention or norm is needed for specifying the conditions of its application. Therefore, an infinite regress is again unavoidable.

Secondly, Collins's and Bloor's social interpretation involves an internal conflict. Undoubtedly, both Collins and Bloor have correctly caught Wittgenstein's insight that a rule does not determine the indefinite totality of its applications. But they then invoke social factors, such as negotiations among actors and social conventions from the relevant community, to explain how an actor can apply a rule correctly. However, this kind of social interpretation of rule-following entails an inconsistency between its basic assumption and its implications. Wittgenstein's view on rule-following, accepted by both Collins and Bloor, is set up to overthrow the quasi-causal picture of rule-following. According to Wittgenstein, following a rule is "a spontaneous decision. . . . [T]hat's how I act; ask for no reason!" (1978, VI, §24). According to Collins's and Bloor's interpretation, however, we are able to apply the rule only because we are guided by a process of negotiation, or compelled by relevant social conventions. Further, these negotiations and social conventions can be explained in terms of the social interests of the relevant groups. Thus, there is a causal mechanism connecting actors' applications of the rule, the negotiations they are involved in or the social conventions to which they are subjected, and their underlying social interests. This external causal mechanism explains the activities of applying rules. But this account directly contradicts Wittgenstein's view on rule-following. Because both Collins and Bloor regard Wittgenstein's account of rule-following as the starting point of their own analysis, their social explanations inevitably lead to an inconsistency.

4. THE BEDROCKOF THE REPRODUCIBILITY RULE

As some critics have indicated, a crucial move in Collins's and Bloor's relativist interpretation of rule-following is the isolation of the formulation of a rule from the practices it formulates (Lynch, 1992). Both Collins and Bloor treat a rule as a representation of an activity, and presuppose the independence of a rule and its extension. Once a rule is treated as a statement isolated from the practice it formulates, the relation between the rule and the practice it formulates becomes undetermined. No matter how many successful applications it had in the past, no rule can exclude the possibility of misapplications or misinterpretations in the future. This raises the question: How does a rule determine its applications? In order to answer this question, Collins and Bloor invoke social factors, like social negotiation and social convention, to explain how an actor can correctly extend a rule to cover new cases.

In his investigation of the nature of rules and their role in language, Wittgenstein makes it clear that following or obeying a rule is not based on any interpretation, explication or negotiation of the appropriate conditions for applying the rule. According to Wittgenstein, "obeying a rule is a practice" (1958, §202). The simplest example to illustrate this point is the case of following a signpost. According to Wittgenstein, a signpost is just like a rule specifying which direction we should follow. In fact, every signpost indicates two directions, from the post to the tip and from the tip to the post. But in daily life we do not need any new rules or any interpretation to tell us which direction we should follow. "[A] person goes by a sign-post only in so far as there exists a regular use of sign-post, a custom" (ibid., §198). We obey a rule because there are regular practices or customs of following the rule in that way. We are trained to do so, and through such training we firmly believe that what we do is simply the way it should be done.

According to Wittgenstein, not only is obeying a rule a practice, but, more importantly, practices are prior to the formulations and followings of rules. He says:

> [A] rule can lead me to an action only in the same sense as can any direction in words, for example, an order. And if people did not agree in their actions according to rules, and could not come to terms with one another, that would be as if they could not come together about the sense of orders or descriptions. It would be a 'confusion of tongues', and one could say that although all of them accompanied their actions with the uttering of sounds, nevertheless there was no language. (Wittgenstein as quoted in Malcolm, 1989, p. 8)

Here, Wittgenstein insists that there can be no rule without a setting of agreement, which is based upon a framework of orderly activities, or, a framework of regular practices. The order of activities or practices already exists when a rule is formulated, and a rule cannot determine anything except within such a framework of orderly activities or regular practices. In this sense, a rule is not a representation (or a universal generalization), but an expression of the orderly activity or regular practice in which it occurs (Lynch, 1992). Also, in this sense, a rule is a part of a regular practice, and it is wrong to separate a rule from the regular practice it prescribes.

On the basis of Wittgenstein's account of rule following, Baker and Hacker argue that the relativist interpretation of rule-following, such as the one advocated by Collins and Bloor, involves a fundamental misconception of the relation between a rule and its applications. Baker and Hacker indicate that a rule cannot be separate from the practices it formulates. The relation between a rule and its applications is internal. Here, an 'internal relation' is a relation between two entities unable to be further decomposed or analyzed into a pair of relations with some independent third entity. There is no third entity mediating between a rule and its applications.[3] Instead, to understand a rule is to be able to apply the rule correctly. Baker and Hacker claim:

> It seems as if there are two independent things, the rule and its applications. In fact, they are two sides of the same coin. One can, of course, say that the rule determines such-and-such as its correct application at this point. But then, not as an external properly of the rule, and so not as an *explanation* of why *this* is a correct application of the rule. And so too, the question: "*How* does the rule determine this as its application?" makes no more sense than "how does this side of the coin determine the other side as its obverse?". (1984, p. 96; original emphasis)

Hence, Collins and Bloor are guilty not only of invoking social factors to mediate between a rule and its applications, but also of raising the wrong question about how a rule determines its applications.

Baker and Hacker further argue that, if the relationship between a rule and its applications is internal, a relativist interpretation of rule-following, such as the one proposed by Collins and Bloor, commits two other mistakes. First, according to this relativist interpretation, the identity of a rule is isolated from its applications, as though a rule is one thing, its applications are others, and only an agent's independent interpretation links them together. However, "[i]t is widely held to be a conceptual truth that to understand a proposition is to know what would be the case if it were true. The parallel for rules is at least as plausible, namely that to understand a rule is to know what would count as acting in accord with it" (ibid., p. 101). Thus, the internal relation between a rule and its applications makes it impossible that a rule be understood without knowing how it should be applied.

Second, the relativist interpretation of rule-following confuses the relation between a rule and the actions it prescribes with the relation between a hypothesis and the events it explains. It is assumed that a rule and ils applications, or its extension, must be represented within the framework of hypothetico-deductive explanation. In particular, the rule must be formulated as a universal generalization, and the descriptions of its extension must be instantiations of this generalization. Yet the internal relationship excludes the need of deductive reasoning to connect a rule and its extension. "The generality essential to a rule need not be made explicit in its expression, but is manifested in the applications of the rule formulation. Rules, unlike hypothetico-deductive explanations, need not have the form of universal generalizations" (ibid., p. 102).

Wittgenstein's idea of the internal relation between a rule and its applications indeed sheds light on the nature of rules and the activities of the rule-following. To apply this insight to our historical case, there are two immediate implications. First, the rule of reproducibility is neither a representation nor a universal generalization; so it cannot be expressed in the form of a hypothetical imperative. Second, the rule of reproducibility cannot be separated from the practice of experiment replication; so neither social negotiations nor social conventions can be inserted as a medium between the rule and its applications. According to Wittgenstein's insight, the rule of reproducibility is embedded in the practice of experiment replication, or, more generally, in the practice of repeating (copying).

Repeating or copying is a regular practice of human beings. To repeat or copy something is to produce a new object identical with the original. This process involves objects (both the new and the original), actions (the operation of reproduction), and sometimes language (the descriptions of the relationship of the two objects) if more than one person is involved.

After the mid seventeenth century, in particular after Boyle's works, experiment replication became a regular practice in scientific communities.[4] Similar to the general practice of repeating or copying, experiment replication aims to produce a new experiment identical to the original. Brewster's descriptions (in his 1847 paper) of what a good replicated experiment should look like reflected the contemporary practice of experiment replication in the early nineteenth century. As stated by Brewster, experiment replication embraced objects, actions, and language. The objects involved in experiment replication include both experimental apparatus and experimental materials. The actions involved are experimental operations, both replications and further explorations if discrepancies are found. And language is needed to describe the result in the form of experimental reports.

At first glance, it seems as if Brewster was proposing a set of methodological rules or describing a group of social conventions for what a good or a legitimate experiment replication should look like. But this interpretation has a formidable problem: even Brewster himself did not strictly follow his own descriptions in his evaluation of Airy's experimental works. To understand better the nature of Brewster's descriptions, it is necessary to review briefly some recent developments in the theory of categorization in cognitive psychology.

According to the classical theory of categorization, every category should have a group of necessary and sufficient criteria specifying the boundary of its membership, and all category members should equally satisfy these criteria and thereby be logically equivalent. Within this theoretical framework, Brewster's descriptions are understood as a set of necessary and sufficient conditions specifying the category 'experiment replication'. According to Wittgenstein, however, categorical judgments become problematic if one is concerned with boundaries; in the normal course of life, we are able to make categorical judgments on the basis of clear cases without knowing anything about boundaries. For example, although we cannot give a definition, or draw a clear boundary for the category "games", we are able to grasp this concept by studying the clear cases of it and the "family resemblance" relationships between the members (Wittgenstein, 1958, §§65–81).

Wittgenstein's insight has gained support from the studies in contemporary cognitive psychology. By conducting a series of psychological experiments, Rosch and Mervis in the 1970s demonstrated the existence of an internal structure, or a graded structure, in categories. Instead of being equivalent, the members of a category vary in exemplifying their category. Some members are especially good or typical examples; the very best called "prototypes". Other members are only moderately typical, and even atypical (Rosch and Mervis, 1975; Rosch, 1978). Further studies in cognitive psychology indicate that the generation of a prototype for a category involves two factors: the cultural stereotype which scientists have adopted and the knowledge base about the properties of the category members. In particular, for a goal-driven category like 'experiment replication', its knowledge base specifies the properties that would be maximally expedient to achieving a goal the category was constructed to serve. During the process of prototype generation, the cultural stereotype people have adopted activates information from the category's knowledge base selectively to represent the category as a whole (Barsalou and Sewell, 1984; Barsalou, 1987).[5]

A graded structure also existed in the category of experiment replication. What Brewster's descriptions represented were not a set of methodological rules or social conventions, but an ideal example, or a prototype, of experiment replication. This replicated experiment contained identical apparatus and materials, proper operations, and proper experimental reports. In fact, the majority of members within the category of experiment replication are merely typical or moderately typical, at a certain distance from this prototype. For example, the requirement of identity is seldom completely satisfied in the daily practices of experiment replication, because scientists are notably uninterested in simply repeating each other's experiments and always make adjustments in their replications. Airy's replication was even more distant from the prototype of experiment replication. First, there were several crucial differences between Airy's experimental setting and Brewster's original; second, Airy in 1847 did not physically reproduce the original experiment, but only presented a reconstruction based upon his personal recollection; third, he did not have a written record of the 1832 experiment he really conducted. All these differences indicate that Airy's work was an atypical example of experiment replication.

It is interesting to ask why Airy's replication was atypical and unacceptable, while Herschel's, which also based on recollection and involved a radical reinterpretation of the experimental result, was regarded as a moderately typical example and widely accepted. One thing to be kept in mind is the difference between their goals – Airy's replication was used to challenge Brewster's findings, while Herschel's was used to further support Faraday's. For those replications which aim to challenge other experiments, the identity between the repetition and the original is one of the most salient features. To challenge an experiment, the less identical the the replication is, the lower the degree of disconfirmation it can produce if it produces a different result. However, for those replications used to support other experiments, identity is not a salient feature. To support an experiment, the more different the replication is, the higher the degree of affirmation it can produce if it produces a similar result. Hence, Airy's and Herschel's replications were not compared to the same prototype. Due to different knowledge bases, there were different prototypes for the replications aiming at disconfirmation and for the replications aiming at confirmation.

If we understand rule-following as a part of practice, and if we consider the graded structure of the concept of experiment replication, we will have a different answer to the questions of why Airy's work had been rejected, and of how consensuses can be achieved in experiment appraisal. From the alternative perspective I have presented, Airy's work was rejected not because it violated any methodological rules, or any social conventions. Instead, Airy's work was rejected because it did not have significant similarities to the ideal example, or the prototype, of experiment replication defined by the existing practice and conceptualized by the members of the relevant scientific communities. In other words, Airy's work was not accepted by the majority of the scientific community because they regarded it as an atypical example of experiment replication. As members of the same scientific community, scientists shared their training, their practices, and "form of life" in many ways. Their understanding of the practice of experiment replication, or their idea of the prototype of experiment replication, was not simply their private judgment, but reflected the training, the practices, and the "form of life" they shared. Hence, although actors may apply the reproducibility rule differently, and they are able to decide whether, when and how to apply the rule, their applications will converge to a consensus on the ground of the existing regular practice, as in the case of Brewster and Airy. In the words of Wittgenstein, this consensus "is not agreement in opinion but in form of life" (1958, §242). In this way, the results of experiment replication are firmly built on a nonsubjective ground: the existing practice of certain scientific communities.

5. CONCLUSION

Wittgenstein's account of rule-following has drawn attention from many philosophers, but few have examined its implications to the practice of science.[6] In an attempt to apply Wittgenstein's account of rule-following to experiment appraisal, I have argued that the reproducibility rule is embedded in the practice of experiment replication, or, more generally, in the practice of repeating (copying), and that there are different degrees of typicality in the concrete practices of experiment replication, or in the correct applications of the reproducibility rule. Given this alternative understanding of the foundation for the reproducibility rule, I have not interpreted the consensus in the Brewster–Airy debate as stemming from the constraints of any methodological rules or any social conventions, but as a result regulated by the existing practice – the degree of typicality was crucial in determining which application was acceptable and which not.

Our alternative interpretation of the reproducibility rule also provides a new understanding of the foundation of experiment appraisal. According to the traditional philosophy of science, such as the account provided by Popper, experiment appraisal is a purely cognitive affair – only logical reasoning is needed. But many historical and empirical studies have indicated that social factors are intrinsically involved in experiment appraisal. Recently, some sociologists have developed a new theory of experiment appraisal, which takes social factors into account. Unfortunately, they go to another extreme – in their theory, experiment appraisal becomes a purely social event.

However, by understanding the reproducibility rule as embedded in practice and by considering the existence of a graded structure of experiment replication, we are able to explain experiment appraisal in terms of both cognitive and social factors. One way to develop such an integrated perspective is to examine the factors that affect the generation of the prototype of experiment replication. As indicated by Barsalou and Sewell's research in cognitive psychology, the generation of a prototype for a given category involves the interactions between cultural stereotype and knowledge base – one mainly reflects the social characters of experimenters, the other, the objective features of experimental elements. On the one hand, people with different cultural or social backgrounds may have different ideas about what the typical example of a category should look like. On the other hand, the content of the knowledge base is relatively independent of the particular cultural stereotype that people adopt, and people with different cultural backgrounds may still have a certain degree of overlap in their knowledge bases about a given category. In this way, the result of experiment appraisal relies on both social and cognitive factors. Experiment appraisal would not be a self-execution of purely logical reasoning. At the same time, although actors have certain freedom in applying the reproducibility rule, their opinions are still deeply shaped by the natural world.

ACKNOWLEDGEMENT

I would like to thank Peter Barker, Mike Bishop, and Joseph Pitt for useful suggestions and criticisms on previous drafts.

NOTES

[1] Here "social conventions" are used strictly in a sociological sense, referring to rules based on general agreements among individuals.

[2] It is also very difficult for this subjectivist interpretation to explain how a consensus can be reached in experiment appraisal among individuals with different interests.

[3] According to Wittgenstein, one example of such an internal relation is that between a desire and its fulfillment. For more discussion, see Baker and Hacker (1984, pp. 106–15).

[4] For details of Boyle's work on experiment replication, see Shapin and Schaffer (1985).

[5] One direct outcome of this prototype generation process is that people who have different cultural backgrounds often have different opinions about how typical a certain object is of its category. In the category of bird, for example, American college students generally agree that robins and eagles are very typical, pigeons moderately typical, and ostriches atypical, while Chinese students generally agree that swans and peacocks are typical, ostriches moderately typical, and bats atypical (Barsalou and Sewell, 1984). I have argued elsewhere (1990) that this phenomenon has an important implication on the issue of incommensurability.

[6] For recent works on Wittgenstein's influence on the philosophy of science, see Barker (1986, 1988) and Dreyfus and Dreyfus (1986).

REFERENCES

Airy, George: 1831, *Mathematical Tract on the Lunar Planetary Theories, the Figures of the Earth, Precession and Nutation, the Calculus of Variations, and the Undulatory Theory of Optics*, Deighton, Cambridge.

Airy, George: 1847, 'On Sir David Brewster's New Analysis of Solar Light', *Philosophical Magazine* **30**, 73–77.

Baker, G. P., and P. M. S. Hacker: 1984, *Scepticism, Rules and Language*, Basil Blackwell, Oxford.

Barker, Peter: 1986, 'Wittgenstein and Authority of Science', in W. Leinfellner and F. Wuketits (eds.), *The Task of Contemporary Phllosophy*, Hölder-Pichler-Tempsky, Vienna, pp. 265–67.

Barker, Peter: 1988, 'The Reflexivity Problem in the Psychology of Science', in B. Gholson et al. (eds.), *Psychology of Science*, Cambridge University Press, Cambridge, pp. 92–114.

Barsalou, Lawrence: 1987, 'The Instability of Graded Structure: Implication for the Nature of Concepts', in U. Neisser (ed.), *Concept and Conceptual Development: Ecological and Intellectual Factors in Categorization*, Cambridge University Press, Cambridge, pp. 101–40.

Barsalou, Lawrence, and Daniel Sewell: 1984, 'Constructing Representations of Categories from Different Points of View', Emory Cognition Project Report #2, Emory University, Atlanta.

Bloor, David: 1991, *Knowledge and Social Imagery*, University of Chicago Press, Chicago.

Brewster, David: 1834, 'On A New Analysis of Solar Light, Indicating Three Primary Colours, Forming Coincident Spectra of Equal Length', *Transactions of the Royal Society of Edinburgh* **12**, 123–36.

Brewster, David: 1837, Review of Whewell's *History of the Inductive Sciences*, *Edinburgh Review* **66**, 58–78.

Brewster, David: 1847a, 'Reply to the Astronomer Royal on the New Analysis of Solar Light', *Philosophical Magazine* **30**, 153–58.

Brewster, David: 1847b, 'Observations on the Analysis of the Spectrum by Absorption', *Philosophical Magazine* **30**, 461–63.

Brewster, David: 1848, 'Observations on the Elementary Colours of the Spectrum, in Reply to M. Melloni', *Philosophical Magazine* **32**, 489–94.

Brewster, David: 1965, *Memoirs of the Life, Writings and Discoveries of Sir Isaac Newton*, Johnson Reprint, New York.

Chen, Xiang: 1990, 'Local Incommensurability and Communicability', in A. Fine, M. Forbes, and L. Wessels (eds.), *PSA 1990*, Vol. I, Philosophy of Science Association, East Lansing, pp. 67–76.

Collins, Harry: 1981, 'Son of Seven Sexes: The Social Construction of A Physical Phenomenon', *Social Studies of Science* **11**, 33–62.

Collins, Harry: 1985, *Changing Order: Replication and Induction in Scientific Practice*, SAGE Publications, London.

Dreyfus, Hubert, and Stuart Dreyfus: 1986, *Mind over Machine*, Free Press, New York.

Franklin, Allan: 1986, *The Neglect of Experiment*, Cambridge University Press, Cambridge.

Franklin, Allan, and Colin Howson: 1988, 'It Probably Is a Valid Experimental Result: A Bayesian Approach to the Epistemology of Experiment', *Study in the History and Philosophy of Science* **19**, 419–27.

Gooding, David: 1989, 'History in the Laboratory: Can We Tell What Really Went On?', in F. James (ed.), *The Development of the Laboratory*, American Institute of Physics, New York, pp. 63–82.

Hones, Michael: 1990, 'Reproducibility as Methodological Imperative in Experimental Research', in A. Fine, M. Forbes, and L. Wessels (eds.), *PSA 1990*, Philosophy of Science Association, East Lansing, pp. 585–99.

Larmor, Joseph (ed.): 1971, *Memoir and Scientific Correspondence of the Late Sir George G. Stokes*, Johnson Reprint, New York.

Lynch, Michael 1992, 'Extending Wittgenstein: The Pivotal Move from Epistemology to the Sociology of Science', in A. Pickering (ed.), *Science as Practice and Culture*, University of Chicago Press, Chicago, pp. 215–65.

Malcolm, Norman: 1989, 'Wittgenstein on Language and Rules', *Philosophy* **64**, 5–28.

Mulkay, Michael, and G. Nigel Gilbert: 1981, 'Replication and Mere Replication', *Philosophy of Social Science* **16**, 21–37.

Newton, Isaac: 1952, *Opticks*, Dover Publications, New York.

Popper, Karl: 1968, *The Logic of Scientific Discovery*, Hutchinson, London.

Rosch, Eleanor: 1978, 'Principles of Categorization', in E. Rosch and B. Lloyd (eds.), *Cognition and Categorization*, Erlbaum, Hillsdale, pp. 27–48.

Rosch, Eleanor, and C. Mervis: 1975, 'Family Resemblances: Studies in the Internal Structure of Categories', *Cognitive Psychology* **7**, 573–605.

Shapin, Steven, and Simon Schaffer: 1985, *Leviathan and the Air-Pump*, Princeton University Press, Princeton.

Whewell, William: 1967, *History of the Inductive Sciences*, Frank Cass, London.

Wittgenstein, Ludwig: 1958, *Philosophical Investigation*, 2nd ed., G. E. M. Anscombe and R. Rhees, trans. G. E. M. Anscombe, Basil Blackwell, Oxford.

Wittgenstein, Ludwig: 1978, *Remarks on the Foundations of Mathematics*, 2nd ed., ed. G. E. M. Anscombe, R. Rhees, and G. H. von Wright, trans. G. E. M. Anscombe, Basil Blackwell, Oxford.

Zucherman, Harriet: 1977, 'Deviant Behaviour and Social Control in Science', in E. Sagarin (ed.), *Deviance and Social Change*, Sage Publications, Beverley Hills, pp. 87–138.