

Advance in Monte Carlo Simulations and Robustness Study and their Implications for the Dispute in Philosophy of Mathematics

Chong Ho Yu

Abstract

Both Carnap and Quine made significant contributions to the philosophy of mathematics despite their diverged views. Carnap endorsed the dichotomy between analytic and synthetic knowledge and classified certain mathematical questions as internal questions appealing to logic and convention. On the contrary, Quine was opposed to the analytic-synthetic distinction and promoted a holistic view of scientific inquiry. The purpose of this paper is to argue that in light of the recent advancement of experimental mathematics such as Monte Carlo simulations, limiting mathematical inquiry to the domain of logic is unjustified. Robustness studies implemented in Monte Carlo Studies demonstrate that mathematics is on par with other experimental-based sciences.

Introduction

Carnap and Quine made tremendous contributions to numerous areas of modern philosophy, including the philosophy of mathematics. Carnap endorsed the dichotomy between analytic and synthetic knowledge and classified certain mathematical questions as internal questions appealing to logic and convention. In addition, he regarded the ontological question about the reality of mathematical objects as a pseudo-question. On the contrary, Quine made an ontological commitment to mathematical entities by asserting that mathematical objects are on par with physical objects. This assertion is tied to his belief that there is no first philosophy prior to natural science. In addition, Quine was opposed to the analytic-synthetic distinction and promoted a holistic view of scientific inquiry. On one hand, Quine recognized that there are differences between logic/mathematics and physical sciences. On the other hand, Quine maintained that it is a

mistake to hold a dualistic view. For Quine logic and mathematics are essentially empirically-based and they are subject to revision according to new evidence. The purpose of this paper is to argue that in light of the recent advancement of experimental mathematics such as Monte Carlo simulations, limiting mathematical inquiry to the domain of logic is unjustified. Robustness studies implemented in Monte Carlo Studies demonstrate that mathematics is on a par with other experimental-based sciences.

Quine (1966/1976) wrote, “Carnap more than anyone else was the embodiment of logical positivism, logical empiricism, the Vienna circle” (p. 40). To discuss Carnap’s philosophy of mathematics, it is essential to illustrate the ideas of the Vienna circle, as well as how members of the Vienna circle adopted and rejected other ideas. In the following, the theories of Frege, Russell, Whitehead and Gödel will be briefly introduced. These are by no means the only ones who are related to the formulation of Carnap’s and the Vienna Circle’s notions. Nonetheless, since this article concentrates on the argument against the logical view of mathematics endorsed by Carnap, discussion of Frege, Russell, Whitehead and Gödel is germane to the topic.

Different Views on The Philosophy Of Mathematics

The Vienna Circle

Logical positivism, which originated with the Vienna circle, embraced verificationism as the criterion for obtaining meaningful knowledge. The verification criterion is not just a demand for evidence. Verification does not mean that, with other things being equal, a proposition that can be

verified is of vastly greater significance than one that cannot. Rather, the verification thesis is much more restrictive than the above. According to logical positivism, a statement is meaningless if verification is not possible or the criteria for verification are not clear (Ayer, 1936; Schlick, 1959). To be specific, the verification principle is not an account of the relative importance of propositions, but a definition of meaning. Meaning and verifiability are almost interchangeable (Werkmeister, 1937). The principle of verification was used by the Vienna Circle as a tool to counteract metaphysics by enforcing adherence to empiricism. However, one may then ask how we can substantiate mathematical knowledge when mathematics is considered by many to be a form of knowledge that cannot be verified by sensory input. Following the strict criterion of verificationism, the analytic philosopher Ayer (1946) has said that mathematics is nonsense. In his view, mathematics says nothing about the world. What it can accomplish is to enlighten us how to manipulate symbols.

Russell and Whitehead

In order to make sense out of mathematics, logical positivists adopted a view of mathematics in the Frege-Russell-Whitehead tradition. This tradition took care of logic and mathematics, and thus left a separate epistemological problem of non-logical and non-mathematical discourse (Isaacson, 2000). According to Frege (1884/1960), logical and mathematical truths are true by virtue of the nature of thought. This notion is further expanded by Russell, and also by collaboration between Russell and Whitehead.

In Russell's view (1919), in order to uncover the underlying structures of mathematical objects, mathematics should be reduced to a more basic element, namely, logic. Thus, his approach is termed logical atomism. Russell's philosophy of mathematics is mainly concerned with geometry. At the time of Russell, the existence of geometric objects and the epistemology of geometry could not be answered by empiricists. In geometry a line can be broken down infinitely to a smaller line. We can neither see nor feel a mathematical line or a mathematical point. Thus, it seems that geometric objects are not objects of empirical perception (sense experience). If this is true, how could conceptions of such objects and their properties be derived from experience as an empiricist would require? Russell's answer is that although geometric objects are theoretical objects, we can still understand geometric structures by applying logic to the study of relationships among those objects: "What matters in mathematics, and to a very great extent in physical science, is not the intrinsic nature of our terms, but the logical nature of their inter-relations" (1919, p.59).

Whitehead and Russell's work on "Principia Mathematica" (1910/1950) is a bold attempt to develop a fully self-sufficient mathematical system through logical relationships. For Russell and Whitehead, mathematics is a purely formal science. The existence of mathematical objects is conditional upon structures. If a certain structure or system exists, then there also exist some other structures or systems whose existence follows logically from the existence of the former. In their view, mathematics could be reduced to logical relationships within the logical system without external references. The Frege-Russell-Whitehead tradition is considered the logical approach to mathematics. This approach is said to be a solution to infinite regress or circular proof.

Gödel

However, the proposal by Whitehead and Russell is seriously challenged by Gödel. Gödel (1944, 1961) proposed that a complete and consistent mathematical system is inherently impossible, and within any consistent mathematical system there are propositions that cannot be proved or disproved on the basis of the axioms within that system. Thus, the consequences drawn from mathematical axioms have meaning only in a hypothetical sense. In addition, mathematical propositions cannot be proved by using combinations of symbols without introducing more abstract elements. In Gödel's sense, logicism in mathematics does not solve the problem of infinite regress or circular proof.

In rejecting the logical approach, Gödel took an "intuitionistic" position to mathematics. Unlike Russell, who asserted mathematical structures exist in terms of relationships, Gödel maintained that it is not a question of whether there are some real objects "out there". Rather, our sequences of acts construct our perceptions of so-called "reality" (Tieszen, 1992). According to Gödel, "despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory... I don't see any reason why we should have less confidence in this kind of perception, i.e. in mathematical intuition, than in sense perception" (cited in Lindstrom, 2000, p.123). Indeed, there are followers of Gödel's even in the late 20th century. Jaffe and Quinn (1993) observed that there is "a trend toward basing mathematics on intuitive reasoning without proof" (p.1).

Carnap

Carnap disliked ontology and metaphysics. For Carnap intuition is a kind of mysterious and unreliable access to matters of independent fact. Creath (1990a, 1990b) argued that anti-intuition is one of the primary motives of Carnap's philosophy. Carnap was firmly opposed to the Platonic tradition of accepting "truths" based upon "supposed direct metaphysical insight or grasp of objects or features of things independent of ourselves but inaccessible to ordinary sensory observation." (p.

4) Creath (1900b) pointed out,

Carnap's proposal, then, is to treat the basic axioms of mathematics, of logic, and of the theory of knowledge itself, as well as the sundry other special sentences, as an implicit definition of the terms they contain. The upshot of this is that simultaneously the basic terms are understood with enough meaning for the purpose of mathematics, logic and so on, and the basic claims thereof need no further justification, for we have so chosen our language as to make these particular claims true... On Carnap's proposal the basic claims are in some sense truths of their own making. It is not that we make objects and features thereof, rather we construct our language in such a way that those claims are true. (p. 6)

Following Poincaré and Hilbert's assertion that the axioms of mathematics can be constructed as implicit definitions of the terms they contain, Carnap viewed numbers as logical objects and rejected the intuitionist approach to mathematics. Although Gödel's theorem brought arguably insurmountable difficulties to the Russell-Whitehead project, Carnap still adopted Russell's logico-analytic method of philosophy, including philosophy of mathematics. By working on logical syntax, Carnap attempted to make philosophy into a normal science in a logical, but not empirical, sense (Wang, 1986). Carnap accepted Russell and Whitehead's thesis that mathematics

can be reduced to logic. Further, Carnap asserted that logic is based on convention and thus it is true by convention. In his essay entitled “Foundations of logic and mathematics” (1971, originally published in 1939), Carnap clearly explained his position on logic and convention:

It is important to be aware of the conventional components in the construction of a language system. This view leads to an unprejudiced investigation of the various forms of new logical systems which differ more or less from the customary form (e.g. the intuitionist logic constructed by Brouwer and Heyting, the systems of logic of modalities as constructed by Lewis and others, the systems of plurivalued logic as constructed by Lukasiewicz and Taski, etc.), and it encourages the construction of further new forms. The task is not to decide which of the different systems is the right logic, but to examine their formal properties and the possibilities for their interpretation and application in science. (pp. 170-171)

The preceding approach is called linguistic conventionalism, in which things can make sense with reference to particular linguistic frameworks. Once we learn the rules of a certain logical and mathematical framework, we have everything we need for knowledge of the required mathematical propositions. In this sense, like the Russell-Whitehead approach, a linguistic framework is a self-contained system.

As mentioned earlier, the verification criterion of logical positivism might face certain difficulties in the context of mathematical proof. Carnap supported a distinction between synthetic and analytical knowledge as a way to delimit the range of application of the verification principle (Isaacson, 2000). To be specific, Carnap (1956) distinguished analytic knowledge from synthetic knowledge, and also internal questions from external questions. An external question is concerned

with the existence or reality of the system of entities as a whole. A typical example is, “Is there a white piece of paper on my desk?” This question can be answered by empirical investigation. A question like “Do normal distributions exist?” is also an external question, but for Carnap, it is a pseudo-question that cannot be meaningfully answered at all.

On the other hand, an internal question is about the existence of certain entities within a given framework. Mathematical truths, such as $1+1=2$, or a set theoretic truth, are tautology in the sense that they are verified by meanings within a given frame of reference; any revision may lead to a change of meanings. In Carnap’s view it is meaningful to ask a question like “Is there a theoretical sampling t-distribution in the Fisherian significance testing?” In other words, to be real in logic and mathematics is to be an element of the system. Logic and mathematics do not rely on empirical substantiation, because they are empty of empirical content.

Quine

Unlike Carnap, Quine did not reject the ontological question of the realness of mathematical entities. Instead, for Quine the existence of mathematical entities should be justified in the way that one justifies the postulation of theoretical entities in physics (Quine, 1957). However, this notion is misunderstood by some mathematicians such as Hersh (1997), and thus needs clarification. Hersh argued that physics depends on machines that accept only finite decimals. No computer can use real numbers that are written in infinite decimals; the microprocessor would be trapped in an infinite process. For example, pi (3.14159...) exists conceptually, but not physically and

computationally. While electrons and protons are measurable and accessible, mathematical objects are not. Thus, Hersh was opposed to Quine's ontological position. Hersh was confused here because he was equating measurability and representation to existence. In the realist sense, the existence of an object does not require that it be known and measured by humans in an exact and precise manner. While the numeric representation of pi does not exist, one could not assert that π also does not exist. Actually, the ontological commitment made by Quine, in which mathematical objects are considered on par with physical objects, is strongly related to his holistic view of epistemology. While Quine asserted that logic/mathematics and physical sciences are different in many aspects, drawing a sharp distinction between them, such as placing logic/mathematics in the analytic camp and putting physical science on the synthetic side, is erroneous. In his well-known paper "Two dogmas of empiricism," Quine (1951) bluntly rejected not only this dualism, but also reductionism, which will be discussed next.

Quine (1966/1976, originally published in 1936) challenged Carnap's notion that mathematics is reduced to logic and that logic is true by convention. Quine asserted that logic cannot be reduced to convention, because to derive anything from conventions one still needs logic. Carnap viewed logical and mathematical syntax as a formalization of meaning, but for Quine a formal system must be a formalization of some already meaningful discourse. Moreover, in rejecting the analytic-synthetic dichotomy, Quine rejected the notion that mathematics and logical truths are true definitions and we can construct a logical language through the selection of meaning. A definition is only a form of notation to express one term in form of others. Nothing of significance can follow

from a definition itself. For example, in the regression equation, $y=a+bx+e$, where y is the outcome variable, x is the regressor/predictor, a is the slope, b is the beta weight, and e is the error term, these symbols cannot not help us to find truths; they are nothing more than a shorthand to express a wordy and complicated relationship. For Quine, meaning is a phenomenon of human agency. There is no meaning apart from what we can learn from interaction with the human community. In this sense, logical truths are not purely analytical; rather, constructing logic can be viewed as a type of empirical inquiry (Isaacson, 2000).

Quine (1951) asserted that there are no purely internal questions. Our commitment to a certain framework is never absolute, and no issue is entirely isolated from pragmatic concerns about the possible effects of the revisions of the framework. In Putnam's (1995) interpretation, Quine's doctrine implies that even so-called logical truths are subject to revision. This doctrine of revisibility is strongly tied to the holistic theme in Quine's philosophy. To be specific, the issue of what logic to accept is a matter of what logic, as a part of our actual science, fits the truth that we are establishing in the science that we engaged in (Isaacson, 2000). Logics are open to revision in light of new experience, background knowledge, and a web of theories. According to Quine's holism, mathematics, like logic, has to be viewed not by itself, but as a part of all-embracing conceptual scheme. In this sense, even so-called mathematical truths are subject to revision, too.

It is essential to further discuss two Quineian notions: revisability of terms and holism, because viewing these Quineian notions as opposition to Carnapian views is a mistake. According to

Friedman (2002), criticism of Carnap by Quine is based on Quine's "misleading" assumption that analytic statements are said to be unrevisable. However, Carnap did not equate analyticity to unrevisability. It is true that in Carnap's linguistic conventionalism logical and mathematical principles play a constitutive role. Nevertheless, even if we stay within the same framework, terms can be revised but their meanings would be changed. Further, we could move from one framework to another, which contains a different set of principles. Consequently, terms are revised in the process of framework migration.

According to Creath (1991), the holist view that Quine embraced in Quine's earlier career might be called radical holism. In Quine's view it is the totality of our beliefs which meets experience or not at all. French scientist Duhem was cited in defense of this holism, but Duhem's argument was not that extreme. In the Duhemian thesis, scientists do not test a single theory; instead, the test involves a web of hypotheses such as auxiliary assumptions associated with the main hypothesis. On the other hand, radical holism states that in theory testing the matter is concerned with whether the totality of our beliefs meets the experience. Creath (1991) criticized that if that is the case, then all our beliefs are equally well confirmed by experience and also are equally disposed to give up as another.

In Quine's later career (1990/1992), he modified his holist position to a moderate one, in which we test theories against a critical mass rather than a totality. A critical mass is a big enough subset of science to imply what to expect from some observation or experiment. The size of this critical mass

will vary from case to case. According to Friedman (2002), Carnap explicitly embraced certain portions of holism such as the Duhemian thesis. For Carnap, a linguistic framework is wholly predicated on the idea that logical principles, just like empirical ones, can be revised in light of a web of empirical science. In this sense, the philosophies of Quine and Carnap share the common ground based on the Duhemian thesis.

According to Pyle (1999), Quine viewed moderate holism as an answer to certain questions in philosophy of mathematics, which are central to Carnapian philosophy. Carnap asserted that mathematics is analytic and thus mathematics can be meaningful without empirical context. Moderate holism's answer is that mathematics absorbs the shared empirical content of the critical masses to which it contributes. In addition, Carnap's analytic position to mathematics makes mathematical truth necessary rather than contingent. Moderate holism's answer is that when a critical mass of sentences jointly implies a false prediction, we could choose what component sentence to revoke. On the other hand, we employ a maxim of "minimum mutilation" (conservatism) to guide our revision, and this accounts for mathematical necessity. Nevertheless, Carnap might not have objections to this, because as mentioned before, Carnap accepted revision of beliefs in light of empirical science. Indeed, moderate holism, as the guiding principle of mathematical and other scientific inquiries, is more reasonable and practical than radical holism.

Discussion

Carnap's views on logic and mathematics, such as distinguishing between analytic-synthetic

knowledge, reducing mathematics to logic and basing logic on convention, are problematic. Indeed, Quine has deeper insight than Carnap because he asserted that logic and mathematics are based on empirical input in the human community; and thus they are subject to revision.

Statistical Theories and Empirical Evidence

There are many examples of mathematical theories that have been substantively revised in light of new evidence. How the newer Item Response Theory amends Classical True Score Theory is a good example. In the article “New rules of measurement,” prominent statisticians Embretson and Reise (2000) explained why the conventional rules of measurement are inadequate and proposed another set of new rules, which are theoretically better and empirically substantiated. For example, the conventional theory states that the standard error of measurement applies to all scores in a particular population, but Embretson found that the standard error of measurement differs across scores but generalizes across populations.

In addition, R. A. Fisher criticized Neyman’s statistical theory because Fisher asserted that mathematical abstraction to the neglect of scientific applications was useless. He mocked that Neyman was misled by algebraic symbolism (Howie, 2002). Interestingly enough, on some occasion Fisher was also confined by mathematical abstraction and algebraic symbolism. In the theory of maximum likelihood estimation, Fisher suggested that as sample size increases, the estimated parameter gets closer to closer to the true parameter (Eliason, 1993). But in the actual world, the data quality may decrease as the sample size increases. To be specific, when

measurement instruments are exposed to the public, the pass rate would rise regardless of the examinee's ability. In this case the estimation might be farther away from the true parameter! Statisticians could not blindly trust the mathematical properties postulated in the Fisherian theorems.

Someone may argue that the preceding examples have too much "application," that they are concerned with the relation between a measurement theory and observations, not a "pure" relation among mathematical entities. Nevertheless, on some occasions, even the functional relationship among mathematical entities is not totally immune from empirical influence. For example, the Logit function, by definition, is the natural log of the odd ratio, which is the ratio between the success rate and the failure rate. However, in the context that the rate of failure is the focal interest of the model, the odd ratio can be reversed.

Putting statistical findings in the arena of "applied mathematics" seems to be an acceptable approach to dismissing the argument that mathematics is subject to revisions. Actually, the distinction between pure and applied mathematics is another form of dualism that attempts to place certain mathematics in the logical domain. In the following I argue that there is no sharp demarcation point between them, and mathematics, like the physical sciences, is subject to empirical verification. Empirically verifying mathematical theories does not mean using a mapping approach to draw correspondence between mathematical and physical objects. Counting two apples on the right hand side and two on the left is not a proof that $2+2=4$. Instead, empirical

verification in mathematics is implemented in computer-based Monte Carlo simulations, in which “behaviors” of numbers and equations are investigated.

Distinction Between Pure and Applied Mathematics

Conventionally speaking, mathematics is divided into pure mathematics and applied mathematics. There is a widespread belief that some branches of mathematics, such as statistics, orient toward application and thus are considered applied mathematics. Interestingly enough, in discussion of the philosophy of mathematics, philosophers tend to cite examples from “theoretical mathematics” such as geometry and algebra, but not “applied” mathematics such as statistics. Although I hesitate to totally tear down the demarcation between pure and applied mathematics, I doubt whether being so-called “pure” or “applied” is the “property” or “essence” of the discipline. As a matter of fact, geometry could be applied to architecture and civil engineering, while statistics can be studied without any reference to empirical measurement. To be specific, a t-test can be asked in an applied manner, such as “Does the IQ mean of Chinese people in Phoenix significantly higher than that of Japanese people in Phoenix?” However, a t-test-related question can be reframed as “Is the mean of set A higher than that of set B given that the Alpha level is 0.5, the power level is 0.75, both sets have equal variances and numbers in each set are normally distributed?” A research question could be directed to the t-test itself: “Would the actual Type I error rate equal the assumed Type I error rate when the Welch’s t-test is applied to a non-normal sample of 30?”

A mathematician can study the last two preceding questions without assigning numbers to any

measurement scale or formulating a hypothesis related to mental constructs, social events, or physical objects. He/she could generate numbers in computer to conduct a mathematical experiment. There is another widespread belief that computer-based experimental mathematics is applied mathematics while traditional mathematics is pure. A century ago our ancestors who had no computers relied on paper and pencil to construct theorems, equations, and procedures. Afterwards, they plugged in some numbers for verification. Today these tasks are performed in a more precise and efficient fashion with the aid of computers. However, it is strange to say that mathematics using pencil and paper is pure mathematics while that employing computers is applied.

In brief, I argue that the line between pure and applied mathematics is blurred. Conventional criteria for this demarcation are highly questionable; the subject matter (geometry or statistics) and the tool (pencil or computer) cannot establish the nature of mathematics (pure or applied). In the following discussion I will discuss how mathematicians use Monte Carlo simulations to support my argument that mathematics is not purely logical but rather has empirical elements. Next, I will use an example of a robustness study to demonstrate how traditional claims on certain statistical theories are revised by findings in simulations.

Computer-Based Experimental Mathematics

With the advancement of high-powered computers, computer simulation is often employed by mathematicians and statisticians as a research methodology. This school is termed "experimental

mathematics" and a journal entitled "Journal of Experimental Mathematics" is specifically devoted to this inquiry (Bailey & Borwein, 2001). Chaitin (1998), a supporter of experimental mathematics, asserted that it is a mistake to regard mathematical axioms as self-evident truths; rather the behaviors of numbers should be verified by computer-based experiments. It is important to differentiate the goal of controlled experiments in psychology, sociology, and engineering from that of experimental mathematics. In the former, the objective is to draw conclusions about mental constructs and physical objects, such as the treatment effectiveness of a counseling program or the efficiency of a microprocessor. In these inquiries, mathematical theories are the frame of reference for making inferences. But in the latter, the research question is directed to the mathematical theories themselves.

Both of them are considered "experiments" because conventional experimental criteria, such as random sampling, random assignment of group membership, manipulation of experimental variables, and control of non-experimental variables, are applied (Cook & Campbell, 1979). Interestingly enough, in terms of the degree of fulfillment of these experimental criteria, experimental mathematics has even more experimental elements than controlled experiments in the social sciences. Consider random sampling first. In social sciences, it is difficult, if not impossible, to collect true random samples. Usually the sample obtained by social scientists is just a convenient sample. For example, a researcher at Arizona State University may recruit participants in the Greater Phoenix area, but he/she rarely obtains subjects from Los Angeles, New York, Dallas, etc., let alone Hong Kong, Beijing, or Seoul. In terms of controlling extraneous

variables or conditions that might have an impact on dependent variables, again social sciences face inherent limitations. Human subjects carry multiple dimensions such as personality, family background, religious beliefs, cultural context, etc. It is definitely impossible that the experimenter could isolate or control all other sources of influences outside the experimental setting. On the other hand, computer-based experiments achieve random sampling by using a random number generator. It is argued that some random number generators are not truly random, but the technology has become more and more sophisticated. Actually, even a slightly flawed random number generator could yield a more random sample than one collected in the human community. Also, computer-based experimental mathematics does not suffer the problem of lacking experimental control, because numbers and equations do not have psychological, social, political, religious or cultural dimensions. In brief, the preceding argument is to establish the notion that experimental mathematics is experimental in every traditional sense.

Monte Carlo Simulations and Robustness Study

Traditional parametric tests, such as t-test and ANOVA, require certain parametric assumptions. Typical parametric assumptions are homogeneity of variances, which means the spread of distributions in each group do not significantly differ from each other, and normal distributions, which means the shape of the sample distribution is like a bell-curve. Traditional statistical theories state that the t-test is robust against mild violations of these assumptions; the Satterthwaite t-test is even more resistant against assumption violations; and the F-test in ANOVA is also robust if the sample size is large (please note that in these theories the sample can be composed of observations

from humans or a set of numbers without any measurement unit). The test of homogeneity of variance is one the preliminary tests for examining whether assumption violations occur. Since conventional theories state that the preceding tests are robust, Box (1953) mocked the idea of testing the variances prior to applying an F-test: "To make a preliminary test on variances is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port!" (p.333).

However, in recent years statisticians have been skeptical of the conventional theories. Different statisticians have proposed their own theories to counteract the problem of assumption violations (Yu, 2002). For instance,

(1) Some researchers construct non-parametric procedures to evade the problem of parametric test assumptions. As the name implies, non-parametric tests do not require parametric assumptions because interval data are converted to rank-ordered data. Examples of non-parametric tests are the Wilcoxon signed rank test and the Mann-Whitney-Wilcoxon test. Some version of non-parametric method is known as order statistics for its focus on using rank-ordered data. A typical example of it is Cliff's statistics (Cliff, 1996).

(2) To address the violation problem, some statisticians introduce robust calculations such as Trimmed means and Winsorized means. The trimmed mean approach is to exclude outliers in the two tails of the distribution while the Winsorized mean method "pulls" extreme cases toward

the center of the distribution. The Winsorized method is based upon Winsor's principle: All observed distributions are Gaussian in the middle. Other robust procedures such as robust regression involve differential weighting to different observations. In the trimmed mean approach outliers are given a zero weighting while robust regression may assign a "lighter" count, say 0.5, to outliers. Cliff (1996), who endorsed order statistics, was skeptical of the differential weighting of robust procedures. He argued that data analysis should follow the principle of "one observation, one vote." Mallows and Tukey (1982) also argued against Winsor's principle. In their view, since this approach pays too much attention to the very center of the distribution, it is highly misleading. Instead, Tukey (1986) strongly recommended using data re-expression procedures, which will be discussed next.

(3) In data re-expression, linear or non-linear equations are applied to the data. When the normality assumption is violated, the distribution could be normalized through re-expression. If the variances of two groups are unequal, certain transformation techniques can be used to stabilize the variances. In the case of non-linearity, this technique can be applied to linearize the data. However, Cliff (1996) argued that data transformation confines the conclusion to the arbitrary version of the variables.

(4) Resampling techniques such as the randomization exact test, jackknife, and bootstrap are proposed by some other statisticians as a counter measure against parametric assumption violations (Diaconis & Efron, 1983; Edgington, 1995; Efron & Tibshirani, 1993; Ludbrook &

Dudley, 1998). Robust procedures recognize the threat of parametric assumption violations and make adjustments to work around the problem. Data re-expression converts data in order to conform to the parametric assumptions. Resampling is very different from the above remedies, for it is not under the framework of theoretical distributions imposed by classical parametric procedures. For example, in bootstrapping, the sample is duplicated many times and treated as a virtual population. Then samples are drawn from this virtual population to construct an empirical sampling distribution. In short, the resampling school replaces theoretical distributions with empirical distributions. In reaction against resampling, Stephen E. Fienberg criticized that "you're trying to get something for nothing. You use the same numbers over and over again until you get an answer that you can't get any other way. In order to do that, you have to assume something, and you may live to regret that hidden assumption later on" (cited in Peterson, 1991, p. 57).

It is obvious that statisticians such as Winsor, Tukey, Cliff, and Fienberg do not agree with each other on the assumption violation and robustness reinforcement issues. If different mathematical systems, as Russell and Whitehead suggested, are self-contained systems, and if mathematics, as Carnap maintained, is reduced to logic that is based on different conventions, these disputes would never come to a conclusive closure. Within the system of Winsor's school, the Gaussian distribution is the ideal and all other associated theorems tend to support Winsor's principle. Within the Tukey's convention, the logic of re-expression fits well with the notions of distribution normalization, variance stabilization, and trend linearization.

It is important to note that these disputes are not about how well those statistical theories could be applied to particular subject matters such as psychology and physics. Rather, these statistical questions could be asked without reference to measurement, and this is the core argument of the school of data re-expression. For example, researchers who argue against data re-expression complain that it would be absurd to obtain a measurement of people's IQ and then transform the data like [new variable = $1/(\text{square root of IQ})$]. They argue that we could conclude that the average IQ of the Chinese people in Phoenix is significantly higher than that of the Japanese, but it makes no sense to say anything about the difference in terms of $1/(\text{square root of IQ})$. However, researchers supporting data re-expression argue that the so-called IQ is just a way of obtaining certain numbers, just like using meters or feet to express height. Numbers can be manipulated in their own right without being mapped onto physical measurement units. In a sense non-parametric statistics and order statistics are forms of data re-expression. For example, when we obtain a vector of scores such as [15, 13, 11, 8, 6], we can order the scalars within the vector as [1, 2, 3, 4, 5]. This "transformation" no doubt alters the measurement and, indeed, loses the precision of the original measurement. Nevertheless, these examples demonstrate that statistical questions can be studied regardless of the measurement units, or even without any measurements. Monte Carlo simulation is a typical example of studying statistics without measurement.

As you may notice in the section regarding bootstrapping, statisticians do not even need empirical data obtained from observations to conduct a test; they could "duplicate" data by manipulating existing observations. In bootstrapping, number generation is still based on empirical observations,

whereas in Monte Carlo simulations all numbers could be generated in computer only. In recent years, robustness studies using Monte Carlo simulations have been widely employed to evaluate the soundness of mathematical procedures in terms of their departure from idealization and robustness against assumption violations. In Monte Carlo simulations, mathematicians make up strange data (e.g. extremely unequal variances, non-normality) to observe how well those mathematical procedures are robust against the violations. Box is right that we cannot row a boat to test the condition for an ocean liner. But using computers to simulate multi-million cases under hundreds scenarios is really the other way around—now we are testing the weather condition with an ocean liner to tell us whether rowing a boat is safe. Through computer simulations we learn that traditional claims concerning the robustness of certain procedures are either invalid or require additional constraints.

There are numerous Monte Carlo studies in the field of statistics. A recent thorough Monte Carlo study (Thompson, Green, Stockford, Chen, & Lo, 2002; Stockford, Thompson, Lo, Chen, Green, & Yu, 2001) demonstrates how experimental mathematics could refute, or at least challenge, the conventional claims in statistical theories. This study investigates the Type I error rate and statistical power of the various statistical procedures. The Type I error rate is the probability of falsely rejecting the null hypothesis, whereas the statistical power is the probability of correctly rejecting the null hypothesis. In this study, statistical procedures under investigation include the conventional independent-samples t-test, the Satterthwaite independent-samples t-test, the Mann-Whitney-Wilcoxon test (non-parametric test), the test for the difference in trimmed means

(robust procedure), and the bootstrap test of the difference in trimmed means (resampling and robust methods). Four factors were manipulated to create 180 conditions: form of the population distribution, variance heterogeneity, sample size, and mean differences. Manipulation of these factors is entirely under the control of the experimenters. No other non-experimental factors could sneak into the computer and affect the conditions. The researchers concluded that the conventional t-test, the Satterthwaite t-test, and the Mann-Whitney-Wilcoxon test produce either poor Type I error rates or loss of power when the assumptions underlying them are violated. The tests of trimmed means and the bootstrap test appear to have fewer difficulties across the range of conditions evaluated. This experimental study indicates that the robustness claims by two versions of the t-test and one of the non-parametric procedures are invalid. On the other hand, one of the robust methods and one of the resampling methods are proved to be true in terms of robustness. Although the scope of this study is narrowed to one of each statistical school, the same approach can be applied to various versions of parametric tests, non-parametric tests, robust procedures, data re-expression methods, and resampling.

Conclusion

The above findings are not achieved by the methods suggested by Russell and Carnap, such as the study of logical relationships, truth by definitions or truth by convention. Rather, the claims result from experimental study. When Quine introduced his philosophical theory on logic and mathematics, computer technology and the Monte Carlo method were not available. Nonetheless, his insight is highly compatible with recent development in experimental mathematics. I strongly

believe that if researchers put aside the analytic-synthetic distinction by adopting Quine's moderate holistic view to scientific inquiry, many disputes could come to a conclusive closure. Indeed, a holistic approach has been beneficial to mathematical inquiry. Although R. A. Fisher was a statistician, he was also versed in biology and agriculture science, and indeed most of his theorems were derived from such empirical fields. Winsor's principle is based on the Gaussian distribution, but Gauss discovered the Gaussian distribution through astronomical observations. Survival analysis or the hazard model is the fruit of medical and sociological research. As discussed before, Embretson and Reise, as psychologists, used the psychometric approach to revise traditional measurement theories. The example of robustness study demonstrates how social scientists employed Monte Carlo studies to challenge traditional claims in mathematics. As Quine's holism proposed, logic, mathematics, observation, and a web of scientific theories are strongly linked to each other.

REFERENCES

- Ayer, A. J. (1936). 'The Principle of Verifiability'. *Mind* (New Series), 45, 199-203.
- Ayer, A. J. (1946). *Language, Truth, and Logic* (2nd ed). London: V. Gollancz.
- Bailey, D. H., & Borwein, J. M. (2001). 'Experimental Mathematics: Recent Developments And Future Outlook'. In B. Engquist, & W. Schmid, (Eds.). *Mathematics Unlimited: 2001 And Beyond* (pp. 51-65). New York: Springer.
- Box, G. E. P. (1953). *Non-Normality And Tests On Variances*. *Biometrika*, 40, 318-335
- Carnap, R. (1971). 'Foundations of Logic And Mathematics'. In O. Neurath, R. Carnap, & C. Morris, (Eds.). *Foundations of The Unity of Science, Toward An International Encyclopedia of Unified Science* (pp. 139-212). Chicago: University of Chicago Press.
- Carnap, R. (1956). *Meaning And Necessity: A Study In Semantics And Modal*. Chicago, IL: University of Chicago Press.
- Cliff, N. (1996). *Ordinal Methods For Behavioral Data Analysis*. Mahwah, NJ: Erlbaum.
- Chaitin, G. J. (1998). *The Limits Of Mathematics: A Course On Information Theory And The Limits Of Formal Reasoning*. Singapore: Springer-Verlag.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design And Analysis Issues For Field Settings*. Boston, MA: Houghton Mifflin Company.
- Creath, R. (1990a). *Carnap, Quine, and The Rejection of Intuition*. In Robert B. Barrett & Roger F. Gibson (Eds.), *Perspectives on Quine* (pp. 55-66). Cambridge, MA: Basil Blackwell.
- Creath, R. (ed.). (1990b). *Dear Carnap, Dean Van: The Quine-Carnap Correspondence and Related Work*. Berkeley, CA: University of California Press.
- Creath, R. (1991). *Every Dogma Has Its Day*. *Erkenntnis*, 35, 347-389.

- Diaconis, P., and B. Efron. (1983). *Computer-Intensive Methods in Statistics*. Scientific American, May, 116-130.
- Edgington, E. S. (1995). *Randomization Tests*. New York: M. Dekker.
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to The Bootstrap*. New York: Chapman & Hall.
- Eliason, S. R. (1993). *Maximum Likelihood Estimation: Logic and Practice*. Newbury Park: Sage.
- Embretson, S. E., & Reise, S. (2000). *Item Response Theory For Psychologists*. Mahwah, NJ: LEA.
- Frege, G. (1884/1960). *The Foundations of Arithmetic* (2nd ed.). New York: Harper.
- Friedman, M. (2002). *Kant, Kuhn, and The Rationality Of Science*. *Philosophy of Science*, 69, 171-190.
- Gödel, K. (1944). 'Russell's Mathematical Logic'. In Paul A. Schilpp, (Ed.). *The Philosophy of Bertrand Russell* (pp.125-153). Chicago: Northwestern University.
- Gödel, K. (1961). *Collected Works, Volume III*. Oxford: Oxford University Press.
- Hersh, R. (1997). *What is Mathematics, Really?* Oxford: Oxford University Press.
- Howie, D. (2002). *Interpreting Probability: Controversies and Developments In The Early Twentieth Century*. Cambridge, UK: Cambridge University Press.
- Isaacson, D. (2000). 'Carnap, Quine, and Logical Truth'. In D. Follesdal (Ed.). *Philosophy Of Quine: General, Reviews, and Analytic/Synthetic* (pp. 360-391). New York: Garland Publishing.
- Jaffe, A., & Quinn, F. (1993). "*Theoretical Mathematics*": *Toward A Cultural Synthesis of Mathematics And Theoretical Physics*. *American Mathematics Society*, 28, 1-13.
- Ludbrook, J. & Dudley, H. (1998). 'Why Permutation Tests Are Superior To T And F Tests In Biomedical Research'. *American Statistician*, 52, 127-132.
- Lindstrom, P. (2000). 'Quasi-Realism in Mathematics'. *Monist*, 83, 122-149.
- Mallows, C. L., & Tukey, J. W. (1982). 'An Overview of Techniques of Data Analysis, Emphasizing Its Exploratory Aspects'. In J. T. de Oliveira & B. Epstein (Eds.), *Some Recent Advances in Statistics* (pp.

111-172). London: Academic Press.

Peterson, I. (July 27, 1991). 'Pick a Sample'. *Science News*, 140, 56-58.

Putname, H. (1995). 'Mathematical Necessity Reconsidered'. In P. Leonardi & M. Santambrogio, (Eds.). *On Quine: New Essays* (pp. 267-282). Cambridge: Cambridge University Press.

Pyle, A. (ed.). (1999). *Key Philosophers in Conversation: The Cogito Interviews*. New York: Routledge.

Quine, W. V. (1951). *Two Dogmas of Empiricism*. *Philosophical Review*, 60, 20-43.

Quine, W. V. (1957). 'The Scope and Language of Science'. *British Journal for the Philosophical Science*, 8, 1-17.

Quine, W. V. (1966/1976). *The Ways of Paradox, and Other Essays*. Cambridge, MA: Harvard University Press.

Quine, W. V. (1990/1992) *Pursuit of Truth* (2nd ed.). Cambridge, MA: Harvard University Press.

Russell, B. (1919). *Introduction to Mathematical Philosophy*. London: Allen & Unwin.

Schlick, M. (1959). 'Positivism and Realism'. In A. J. Ayer (Ed.), *Logical Positivism* (pp. 82-107). New York: Free Press.

Stockford, S., Thompson, M., Lo, W. J., Chen, Y. H., Green, S., & Yu, C. H. (2001 October). 'Confronting The Statistical Assumptions: New Alternatives For Comparing Groups'. Paper presented at the Annual Meeting of Arizona Educational Researcher Organization, Tempe, AZ.

Thompson, M. S., Green, S. B., Stockford, S. M., Chen, Y., & Lo, W. (2002 April). 'The .05 level: The probability that the independent-samples t test should be applied?' Paper presented at the Annual Meeting of the American Education Researcher Association, New Orleans, LA.

Tieszen, R. (1992). 'Kurt Gödel and Phenomenology'. *Philosophy of Science*, 59, 176-194.

Tukey, J. W. (1986). *The Collected Works of John W. Tukey (Volume IV): Philosophy and Principles Of Data Analysis 1965-1986*. Monterey, CA: Wadsworth & Brooks/Cole.

Wang, H. (1986). *Beyond Analytic Philosophy: Doing Justice To What We Know*. Cambridge, MA: MIT Press.

Werkmeister, W. H. (1937). 'Seven Theses of Logical Positivism Critically Examined Part I'. *The Philosophical Review*, 46, 276-297.

Whitehead, A. N., & Russell, B. (1910/1950). *Principia Mathematica* (2nd ed.). Cambridge, UK: Cambridge University Press.

Yu, C. H. (2002). 'An Overview Of Remedial Tools For Violations Of Parametric Test Assumptions in The SAS System.' *Proceedings of 2002 Western Users of SAS Software Conference*, 172-178.

Copyright © 2004 Minerva

All rights are reserved, but fair and good faith use with full attribution may be made of this work for educational or scholarly purposes.

Chong Ho Yu has a PhD in Measurement, Statistics, and Methodological Studies from Arizona State University, USA. Currently he is pursuing a second doctorate in Philosophy at the same institution. He is also a Psychometrician at Cisco Ssystems/Aries Technology, USA. His research interests include philosophical foundations of research methodology, and relationship between science and religion.

Email: chonghoyu@yahoo.com