Toward a General Theory of Observational Causal Inference

C. Sterling Portwood[*]

Abstract

Neither frequentist nor Bayesian paradigms can be used to draw causal inferences because the word cause is not a part of their derivations. Nevertheless, the physical/experimental sciences have been overwhelmingly successful, by intuitively extending these classical paradigms with unstated, but simple and generally acceptable assumptions. The nonexperimental sciences have had the opposite experience, because their assumptions required for extension are voluminous, complicated, and anything but generally acceptable. Further, Young and Karr (*Significance*, 2011) showed empirically that 100 percent of the causal findings from 52 randomly selected, observational extension studies were incorrect. I summarize a new inquiring paradigm for causal inference called causal statistics and layout the assumptions foundational to its derivation and application. The outcome of the derivation is a series of simultaneous causal expressions (not equations) redolent of Pearl's do–calculus form. Causal statistics infers causal connections in samples. Classical statistics transports these results to populations. They are complementary. Much work remains to fully develop this new causal inquiring paradigm, but this research points the way.

Key Words: causal inference, derivation, observational, nonexperimental, do–calculus, assumptions

1. Causality Is the Most Fundamental, the Most Important, and the Most Difficult Concept in Science

We have evidence in cuneiform tablets from Mesopotamia and papyri from Egypt that back to at least 5000 years ago and forward to Aristotle (2400 years ago), people generally had a clear, intuitive understanding of what cause meant, but they believed that the spirits, the gods, etc. caused just about everything; sickness, good fortune, the outcomes of wars, etc.

About 600 BC (2600 years ago) the first philosopher in the Greek tradition, Thales, eschewed supernatural explanations for natural phenomena, in favor of a causal explanation of nature. He believed that natural events had natural causes. That insight was the beginning of science and it carried the physical sciences to unimaginable heights, physically beyond the heliopause and mentally back to the beginning, 13.8 billion years ago.

Causality is a unique kind of concept; it is a metaphysical notion; like the concepts of gods, morality, and the presence of spirits or souls. These things are beyond our sense perceptions. They cannot be observed, measured, or calculated.

Consider correlation versus causality. Correlation is a typical variable. It can be observed, measured, and calculated. Under certain circumstances, causal connections can be inferred from correlations; based on assumptions, data, judiciously chosen prior information (e.g., prior research results), etc.

But if you can't see, feel, or measure causality; then how do we know that it is so important to science? Well, in a sense we don't know that; we can't know it with certainty, because there are alternatives that cannot be ruled out with indubitableness.

It cannot be proved that Thales was correct or, stated in the alternative; there are different explanations for the operation of the universe that cannot be proved incorrect. The three most common metaphysical mechanisms posited as controls for the behavior of the universe are as follows: 1) God or the gods are responsible for and control the operation of

---

[*] Center for Interdisciplinary Science, 2509 Ala Wai Blvd Suite 205, Honolulu, HI 96815

the universe; 2) the universe is not controlled, the apparently well–ordered universe which we see is simply a highly improbable, random occurrence, where no connections between natural phenomena actually exist; 3) Thales had it right, the operation of the universe is based on causal laws, mechanistically directing the world's behavior.

None of these three alternative hypotheses concerning the operation of the universe can be proven either correct or incorrect, but which seems to you most reasonable and useful? These alternative explanations are observationally equivalent. Our sense perceptions cannot distinguish among these alternative foundations and our logical analyses cannot, with certainty, prove one of them correct and the others wrong. The best that we can do is to argue to human intuition.

Before Thales, all scribes seemed to accept some version of the gods–control–all explanation for the operation of the universe. Even today a large number of humans still accept one or another form of the hypothesis that gods, spirits, etc. control and direct the activities the universe. That was easier to accept 1) when so much was unexplained by science and 2) when cathedrals, mosques, etc. were the largest, most impressive buildings within their experiences.

But today the outcome of that kind of reasoning has been reversed. Causality–based science (physical science, at least) and secular building construction has outstripped religious progress by orders of magnitude. This is not a proof of foundational validity, but it appeals to the intuition, consistent with the argument that foundations are validated by their outcomes.

Progress via the social sciences versus progress via religion is a much closer call. Jumping somewhat ahead in this paper, the reason for the arguable parity between the productivity of the social sciences and the productivity of religion, is the vast retardation in the development of the social sciences. That vast retardation in the development of the social sciences is largely due to their inability to draw causal inferences from their research studies and that inability is in turn due to the very limited ability of social science researchers to conduct experimental studies. The social sciences are largely limited to nonexperimental/observational research. We will get to all of that later.

Concerning the foundational hypothesis that the well–ordered universe is simply a freak random occurrence, where no connections between natural phenomena really exist; all would concede that such an event would be wildly improbable. In fact the improbability increases with every second that passes in which the universe doesn't collapse into an enormous and amorphous, stochastic soup. Nevertheless it must be admitted that this random fundamental hypothesis is theoretically possible, but with odds fantastically beyond the most high–stakes lottery. From a practical point of view, I would consider it pragmatically impossible and just too fantastical to be believed.

Looking back, the physical sciences, which are based on Thales' assumption, have taken mankind far further than any other area of human endeavor and the random hypothesis, combined with the well–ordered universe we observed, imply a vanishingly improbable occurrence. Therefore, this leaves Thales causal insight as the last metaphysical foundation reasonably standing.

So, even though no one of the three hypotheses can be proved correct or incorrect and each person is logically free to make his/her own choice, my own intuition is that it would be most practical to assume that Thales was correct and the operation of the universe is based on natural causal laws. I will assume that for the rest of this paper.

So science is based on the metaphysical concept of causality and on data. Data sets are the connection to the universe that distinguish among alternative, theoretically possible scientific hypotheses. This odd add mixture of causality and empirical evidence is probably the reason that technology was adopted by the human species over 500,000 years ago, but science was initiated only 2600 years ago and, even then, the Greeks were very weak at bringing data into the mix and, in the next epoch, the Romans showed little interest in causality. It may seem simple today, but in those days the pieces of science and how they fit together were not obvious. In fact it wasn't until the Renaissance that humans generally started to fit the pieces together properly.

In a somewhat oversimplified view, it is the job of empirical science (both the physical sciences and the social sciences) 1) to perform studies and collect data, 2) to draw causal inferences, and from that, 3) to build causal theories within the science; so that research consumers can intervene and not just predict.

In 1953, Einstein, in a letter to a colleague, wrote (paraphrased here for clarity) that the phenomenal success of the physical (i.e., experimental) sciences is based on two great achievements: 1) mathematical axiomization, as in the derivation of Euclidean geometry, and 2) causal inferences, arrived at through experimentation and leading to causal theories.

In this paper, we shall deal with both of these great achievements in the context of both new and old statistical paradigms and the sciences, especially the social sciences.

Thales and Einstein were thinking only about the physical sciences. But Thales' insight is equally foundational and important for the social sciences. Yet the social sciences have shown not one 10,000th of the success of the physical sciences. Why is this? Think about it . . .

The general answer is that there are a number of reasons why, but the most important reason is the inability of social science researchers to experiment. In nonexperimental (also called observational) research, causal inferences are vastly more difficult. Social scientists are therefore very limited in their ability to build causal theories and social science research consumers are similarly limited in their ability to intervene appropriately. They are largely relegated to prediction, only.

2. Can Classical Statistics Be Used to Draw Causal Inferences?

Summary of Section 2: Neither frequentist nor Bayesian statistical paradigms can be used to draw causal inferences because the word cause is not a part of their derivations. These inquiring systems cannot reach conclusions about a concept that they have no knowledge of, like the mathematical discourse of Euclidean geometry cannot by itself draw conclusions about the apparent color produced by overlaying florescent blue and yellow, congruent triangles. The answer is that the apparent color would be green, but that result could only be attained from geometry, if geometry were extended to incorporate the color wheel or if a new color geometry were developed.

Classical statistics can be considered to be an axiomatic mathematical discourse, as is Euclidean geometry. For purposes of this presentation frequentist and Bayesian statistics will be considered the components of classical statistics. Fiducial statistics will be disregarded here because of its limited, current use.

Andrey Kolmogorov (1903–1987) derived probability theory beginning with six axioms which later mathematicians reduced to three. Building on the pure mathematical discourse of probability theory, every time you teach your basic statistics class, you do a simplified derivation of classical statistics. The resulting pure mathematical discipline of statistics can then be transformed into the two alternative applied mathematical branches of frequentist and Bayesian statistics by inserting two different definitions for the technical element of probability.

An understanding of the relationships between Euclidean and non–Euclidean geometries will be helpful in comprehending the situation among Frequentist statistics, Bayesian statistics, and other potential statistical paradigms. Does Euclidean geometry applied to everything? No, it only applies to flat planes. If you wanted to determine the number of degrees in a triangle inscribed on a sphere the total number of degrees in the three angles would be greater than 180. Similarly, if you wanted to determine the number of degrees in a triangle inscribed on the inside surface of a sphere, like the inside of a world globe, the number of degrees would be less than 180. Various non–Euclidean geometries apply to these types of situations, but Euclidean geometry does not.

A non–Euclidean geometry is one in which one or more fundamental aspects, often a postulate, fundamental to Euclidean geometry, is altered. For example, one might alter the fifth postulate in Euclidean geometry, i.e., the famous parallel postulate, so that parallel lines do intersect, which does not happen on flat planes, but is something that happens on both inside and outside surfaces of a sphere. Then, using one of these modified axiom sets, new and different theorems can be derived. These altered theorems then apply to different parts of the world (e.g., on convex surfaces) from the world of flat planes, for which Euclidean geometry is applicable.

The point being that it is normal and typical for a given branch of applied mathematics, including its provable theorems, to be applicable to specific phenomena or specific aspects of the real world and not to others.

The question at issue here is whether or not classical statistics can be employed to make causal inferences from either experimental or nonexperimental studies. The applied mathematical discourses of frequentist and Bayesian statistics are perfectly capable of handling, for example, sample means or correlations and inferring their values to populations, but, by themselves, mathematically incapable of drawing causal inferences from samples.

This is because in the derivations of these classical inquiring paradigms, the word ”cause” is never mentioned—not in the axioms, not in the technical terms or primitives, and not in the metalanguage (e.g., English). Therefore causes and causal inference cannot be a part of or a result of either one of these classical statistical paradigms, leading to the logically necessary conclusion that any statistical inquiring system which could validly draw statistical causal inferences must be a paradigm distinct from the classical statistical disciplines, i.e., an extension of classical statistics or a whole new paradigm.

Do you find it interesting that causality is the most fundamental and most important concept in science and yet the classical statistical paradigms cannot handle it? I find it shocking and almost incomprehensible! That’s a huge void in the firmament of statistics.

This is reminiscent of an old story which I will adapt for the situation. A statistician and a wheelchair–bound social science researcher were ambulating down a path in the rain forest and came upon a mango tree with ripe mangoes at the very top. The statistician went directly to the tree and started pulling off leaves. He then ate some of them and gave some to be researcher to eat. Puzzled, the social science researcher asked the statistician, You know the leaves have little food value and taste terrible, why not climb to the top and pick the mangoes to eat. The statistician responded, That would be far better, but the leaves are so much easier to get and there is no risk of falling.

Causal knowledge is far better, but correlations are much easier to get, there is no risk of failure, and parenthetically they are accepted for tenure.

3. So, How Have the Physical Sciences Been so Successful at Drawing Causal Inferences?

Summary of Section 3: The physical/experimental sciences have been overwhelmingly successful, by intuitively extending the classical paradigms by implicitly inserting causality and injecting unstated, but simple and generally acceptable, assumptions.

Thales began science by claiming that natural phenomena had natural causes, as opposed to being caused by the gods. The physical/experimental sciences have been overwhelmingly successful, by intuitively applying Thales' insight. Yet few if any of the scientists (or statisticians for that matter) were conscious of the full explication of the mathematics or statistics underlying their efforts. They were tacitly extending the classical statistics paradigms into a different paradigm that might be appropriately called experimental causal statistics. Their implicit processes and assumptions were typically not questioned because they were so intuitively appealing and generally acceptable.

The fact that scientists and statisticians are not cognizant of some of the logical foundations of their research activities may sound incredible, but it's really not. Both Euclid and Einstein, in their derivations, made intuitive and implicit assumptions that they were not consciously aware of at the time. In fact in Euclid's case it was 21 centuries before anyone became aware of his implicit assumptions.

To understand these intuitive and subconscious structures, one must at least allude to them explicitly. Analytically, one would begin with classical statistics and extend it with additional mathematical processes. Cause would be inserted into the extended axiomization as a technical term; its definition or interpretation would be specified using the metalanguage (e.g., English); a modern statement of Thales' insight would be input as an axiom; and other needed assumptions or axioms concerning the experimental design would be inserted.

The updated Thales axiom would note that all events have natural (i.e., non–mystical) causes and/or that all real correlations are in some way a result of causal mechanisms. This would lay the foundation for progressing from correlation (which is observable, measurable, and calculable) to the inference of causation (which is none of those things).

A typical and generally acceptable experimental assumption would be that the experimenters' action in manipulating the putative cause is not correlated with other potential causes of the putative effect. An example counter to this assumption would be the experimenter who arises early every morning and concludes that his Snap, Crackle, and Pop breakfast food causes the cock to crow, when in fact the sunrise is the causal variable.

Through this explicit, lengthened derivation, it is clear that the resulting applied mathematical discourse is no longer a classical statistical paradigm. It is an expanded, more capable and far–reaching statistical inquiring system, arrived at via continued axiomization, beginning from the classical statistics discourse. This resulting paradigm, that I call experimental causal statistics, is capable of logically manipulating and relating correlation, causation, and experimental data and, in particular, of drawing causal inferences from experimental data. It gives an explicit, logical foundation for the research which physical/experimental scientists have been carrying out intuitively and implicitly over the past few hundred years.

On the other hand, I don't mean to imply that I have above actually performed the extended derivation. I have only been explicit about describing how it would be done. The actual extended axiomization would, by itself, be a full paper and the proofs of important theorems could be another paper or so.

Now, turning to nonexperimental/observational causal inference, that is a whole different kettle of fish.

4. What about Causal Inference in the Social Sciences?

Summary of Section 4: The social sciences are largely limited to conducting nonexperimental studies. The social and other nonexperimental sciences have had the opposite experience to that of the physical/experimental sciences. The assumptions required for extension *of classical statistics to draw causal inferences from social/observational studies, are voluminous, complicated, and anything but generally acceptable. Therefore the development of a new observational causal inquiring paradigm is the only reasonable path forward.*

One would hope that for the social and other observational sciences, one could do a similar thing, i.e., extend the classical statistics, to that which was done for the physical/experimental sciences implicitly and in this paper alluded to explicitly. Unfortunately that turns out not to be the case.

Obtaining knowledge in the social/observational sciences is far more difficult at every level than obtaining knowledge in the physical/experimental sciences. There are far more variables involved in almost any social science situation. Typically the variables are far less precise; for example confidence may originate from intellectual performance or athletic performance. But, of all the difficulties faced by the social sciences, the most devastating is their general inability to experiment.

This inability to experiment makes drawing causal inferences almost impossible. Certainly the comparatively simple extension to classical statistics that worked for the physical sciences has not and will not work for the social sciences, epidemiology, and other nonexperimental disciplines. For social scientists to have any general chance of drawing causal connections from their observational studies, a completely new causal inquiring paradigm would have to be developed, derived.

The difficulties have dissuaded most statisticians and social scientists from even attempting observational causal inference, although many have chosen, inappropriately, to come down in the middle by using ambiguous synonyms like leads to, results in, yields, etc. All this does is add to the confusion.

A few of the best statisticians have, during the last 100 years, perceived the need and tried to appropriately fill it. They attempted to do largely as the physical scientists did, i.e., intuitively and more or less implicitly extend classical statistics to draw causal connections from nonexperimental data.

Unfortunately these efforts, although receiving a certain amount of acclaim in their time, have, in the end, been generally unsuccessful. We know that these efforts have been unsuccessful because, if any one of them was truly capable of, in a generalized and reasonably applicable way, drawing observational causal inferences, the response would've been sensational. Any paradigm truly capable of that would revolutionize the social and other nonexperimental sciences and that hasn't happened.

Further and surprisingly, there is now empirical evidence that these approaches have not worked. In 2011, Stanley Young and Alan Karr analyzed 52, effectively randomly selected, published studies which had reported success in their efforts to draw nonexperimental causal inferences through the utilization an extension of classical statistics. The findings of Young and Karr were presented in *Significance* under the title, Deming, Data and Observational Studies: a Process Out Of Control and Needing Fixing.

In the lead up to the paper, the editors wrote:

> Any claim coming from an observational study is most likely to be wrong. Startling, but true. Coffee causes pancreatic cancer. Type A personality causes heart attacks. Transfat is a killer. Women who eat breakfast cereal give birth to more boys. All these claims come from observational studies; yet when the studies are carefully examined, the claimed links appear to be incorrect. What is going wrong? Some have suggested that the scientific method is failing, that nature itself is playing tricks on us. But it is our way of studying nature that is broken and that urgently needs mending . . . .

Specifically, at the end of their analysis, they concluded that not even one of the 52 discovered nonexperimental causal conclusions was borne out by experimental replication.

Further, to add insult to injury, in five of the cases, experimental findings yielded significant causal inferences in the opposite direction to those found by intuitive and implicit extension of classical statistics.

The natural question at this point is, why would an extension to classical statistics work so well for the physical sciences be so powerless when applied to the social and other nonexperimental sciences? The answer is that the experimenter acts as a touchstone, over repeated runs limiting the causing variables to one, the one manipulated by the experimenter. In the nonexperimental case there is no such touchstone; any variables, even those not considered by the study, can be the cause or causes of a change in the putative effect variable. This is analogous to trying to solve an algebraic equation of 10 unknowns with only one equation. There are logical ways around this, but they are complicated and require questionable assumptions. It's not impossible it's only almost impossible.

This is a shocking and almost inconceivable finding and is testatory of the noxiousness of the causal–theory–construction difficulties faced by social, health, and other observational scientists. As Young and Karr mildly, but pointedly, note in their title, this needs fixing.

To their credit they proposed a fix, but even if their suggestions were scrupulously followed, the outcome would be beneficial, but far from a complete fix. I dealt with this issue some time ago and the next Section is a report on the development and the results of that attempted fix.

5. A Proposed Fix

Summary of Section 5: This Section reports on the development of a new statistical paradigm for inferring causal connections from observational studies. It outlines the philosophical and technical investigations, the logic which led to the conclusion that an axiomization was required, and the derivation of the new inquiring paradigm. The derivation and accompanying information and explanations are about 300 pages long, necessitating the summarization reported in this Section.

Imagine that your goal was to develop a new statistical paradigm for making nonexperimental causal inferences. Just spend a moment thinking about how you would approach and carry out that effort . . .

I faced this precise problem. My initial approach was to sit down at a blank sheet of paper and start writing, writing anything that came into my head. I found myself writing questions and then trying to answer them. At first I wrote about five questions for every answer that I generated. Often I would produce three or four questions before I could finish the previous question and consider an answer. Toward the end, I began to catch up, i.e., more answers than new questions.

That blank paper exercise extended to 900 pages and two man years. During that time and in a process that sounds far more organize then it really was, I investigated the history, philosophy, science, statistics, and practical application of causality and related topics. The first step focused largely on what others had said and done concerning the philosophy and science of causality.

The first known writings and their references to even older texts going back to 5000 years ago, paint a clear picture of ancient people's understanding and usage of the concept of causality. People had a good intuitive understanding of what causality meant, but they were confused about what caused what. Until Thales' insight in about 600 BC, people believed that the gods caused pretty much everything.

Then about 2400 years ago Aristotle (384 BC–322 BC) presented and defined four different types of causality and that began a free–for–all of confusion among subsequent philosophers. Many of these philosophers attempted to draw the definition of causality from logic; i.e., some combination of necessary and/or sufficient conditions, if/then relationships, implications (all related structures from logic); not because the definitions fitted the need, but because logic was what philosophers of the time knew. Somewhat inexplicably, the idea of applying something new and different because that's what the situation demanded, seemed to have been pretty much lost on them.

Fairly deep into my analysis, I came to the realization that not even one of these philosophers had produced an adequate definition of cause. At that point I realized that, given the poor quality of the work product on causal definitions, any hope that the old philosophers had solved the other, more difficult problems concerning the concept, was likely a pipedream. Even so, I completed the study of the work that others had done, in an effort to gleaning what I could that might be fundamental to scientific thought. There were a very few tidbits here and there and Hume had some interesting insights, but on the whole I felt that many had labored mightily, only to bring forth the proverbial mouse, i.e., that the output as regarded causality and causal philosophy of the best minds of the last 24 centuries was remarkably sparse.

Even more surprisingly and contrary to my article of faith concerning the operation of the intellectual world, things didn't get better with time; they seemed to get progressively worse. The chaos reached a crescendo in the 20th century when 1) some scientists declared that physics, and later all of science, had no need of the concept of causality; 2) many quantum physicists concluded that causality was incompatible with quantum mechanics (especially entanglement); and 3) no less a luminary than Bertrand Russell declared, The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm. Later many of these scholars realize the error of their thoughts and a few, including Russell, even recanted. But as anyone who reads newspapers knows, the corrections to a front–page story are many times below the fold and more often buried on the inside pages.

After completing my analysis of the old Masters, I concluded that the work product was grossly insufficient to ground the development of a generalized observational causal inquiring paradigm, its application to data, or the coherent utilization of causal results obtained therefrom. Without answers to the crucial philosophical and technical questions about causality, these activities could not be carried out with the necessary understanding or with confidence. I was disappointed and disillusioned. I despaired at the thought of ever being able to straighten out the messes, both in causal philosophy and in observational causal inference.

Nevertheless, on one sunny California morning, I woke up and the path forward (step two) was obvious, the mental die was cast. I would have to ape Descartes and reject everything that I thought I knew and a lot that I was pretty sure I didn't know. I would have to go back to the beginning and build a science and statistics oriented causal philosophy and understanding from scratch.

I had a multitude of questions about causality, almost none of which had been considered or at least adequately resolved by philosophers over the past 2400 years. The answers to these questions would be foundational to the observational causal inquiring system which I hope to develop. To give a feeling for the undertaking, the following is a sample of about 25 percent of these questions:

- What is causality? What is the origin and evolution of the concept?
- Is the definition of cause fixed by the universe, like the definition of micro momentum, or is it a concept whose definition is unconstrained and hence open to human choice?
- Is causality a generalization, a type of theory, a law, a principal, a property, a rule, etc.?
- Is causality an emergent phenomenon? Can causal connections be proved?

- Does the universe know about the causal connections inferred by social scientists?
- What is the relationship between tautologies and causality or definitions and causality?
- What is the origin and nature of probabilistic causality?
- What is the relationship between the cause of a specific event and of a type of event?
- What about the fact that in the social sciences the path is seldom considered?
- Is causality an event or is it a generalized force acting on variables of all types?
- What if the composition of a variable differs from situation to situation?
- Are all real correlations ultimately causal in origin? Is causality inherently local?
- What is the proper mathematical form to describe causal relationships at the macro level?
- What assumptions are implicitly made in drawing causal inferences?
- Why are causal inferences so much more difficult in nonexperimental studies?
- Is it causal when an apple is pulled on by gravity, but the stem does not break?
- What about when a firing squad kills a prisoner? Who caused his death?
- Is F = ma an example of bidirectional and/or functionally symmetric causality?
- Do physics, quantum mechanics, and science in general really not need causality?
- Is there a qualitative difference between causality in the physical and social sciences?
- Can causality move at speeds faster than the speed of light within a quantum waveform?
- What about gravity from a free electron being felt on the other side of the universe?
- Is causality built into the universe or is it a creation only of human minds? Etc., etc., etc.

In the same timeframe I began step three, i.e., an analysis of the relevant efforts of the outstanding statisticians who, over the previous decades, had butted heads with the problem of observational causal inference. Seemingly, the most appropriate formulation utilized by these confident and capable statisticians was a set of reciprocal simultaneous linear equations, as follows:

$$\begin{aligned} Y_1 &= \alpha_1 + \gamma_{12}Y_2 + \gamma_{13}Y_3 + \cdots + \gamma_{1M}Y_M + \beta_{11}X_1 + \cdots + \beta_{1K}X_{1K} + e_1 \\ Y_2 &= \alpha_2 + \gamma_{21}Y_1 + \gamma_{23}Y_3 + \cdots + \gamma_{2M}Y_M + \beta_{21}X_1 + \cdots + \beta_{2K}X_{2K} + e_2 \\ &\vdots \\ Y_M &= \alpha_M + \gamma_{M1}Y_1 + \gamma_{M2}Y_2 + \cdots + \beta_{M1}X_1 + \cdots + \beta_{MK}X_{MK} + e_M \end{aligned} \tag{1}$$

Such intuitive specification, i.e., without proof, can lead to major problems, the most obvious of which is that the appropriate formulation might be different than the one intuited. Another set of problems can arise because the assumptions and simplifications inherent in these equations are largely unknown, making the structural accuracy of any result obtained also unknown. Of course I had no way to know that Young and Karr would far later come along and prove empirically that there was at least something rotten in the state of observational causal inference.

At the end of the aforementioned three–step process, many of the definitional, foundational, philosophical, statistical, and procedural issues had been resolved. Nevertheless the analysis had reached no determination concerning the appropriateness of the preferred statistical formulation, i.e., form (1), along with any others that had been intuited by statisticians. That left both the correct of the foundational formulation for an observational causal inquiring system and the assumptions which had hitherto been intuitively and implicitly made, uncertain.

In an effort to overcome these uncertainties, I concluded that the way forward was to carry out an axiomization, designed to begin with the fundamental laws of the universe and derive the nature of the macro causes and macro objects that populate our observable world. This would enable the determination of 1) the appropriate mathematical form for representing causal connections in the macro world, 2) the statistical formulation from which to begin an inquiry into observational causal inference, and 3) the required assumptions upon which nonexperimental macro causal inferences are based. These findings would then lead to a host of secondary, consequential important understandings; for example the relationship of the derived observational causal inference paradigm to the classical statistics paradigms, the likely future direction and importance of the field of panoptic statistics, etc. But, more practically and more importantly, the correct foundational statistical formulation and explication of the inherent assumptions would prove the validity of the derived observational causal inference tool and instill great confidence in its application and results. On the other hand, knowing the foundational assumptions, would make the results easier to critique, but, if you think about it, that is consistent with good science.

To begin the derivation process, I needed to look into the universe to discover the laws by which the universe acts. For this we need to go to the micro levels; below anything we can see, smaller than molecules, beneath the chemistry of atoms, to fundamental particles. The universe seems to operate at this level or below because all activity at larger scales seem to be consequences or aggregations of the activities at the fundamental scale.

Electrons are fundamental particles, but protons and neutrons are not. Protons and neutrons are composed of three quarks each, which are elementary particles. Protons are made up of two up quarks and one down quark and neutrons: two down quarks and one up quark. There are four fundamental forces that act between fundamental particles; the electromagnetic force, the strong nuclear force, the weak nuclear force, and gravitation.

The atom is held together by electrostatic attraction between the electrons and the protons. This electrostatic force is carried from positive charge to negative charge and vice a versa by virtual photons, another elemental particle. Each group of three quarks is held together by the attraction of the strong nuclear force, that force being carried from one quark to another by virtual gluons which are also fundamental particles. Virtual particles owe their short–term existence to the probabilistic and rule violating nature of the Heisenberg uncertainty principle.

The nucleus is rife with repulsive electromagnetic force, but it is held together by the secondary effects of the strong nuclear force, carried by gluons which leak out of one nucleon and into another. Nevertheless, for large nuclei, for example uranium and above, the disparity between the repulsive electromagnetic forces and the attractive secondary strong forces diminish and eventually reverse. That's why large nuclei are unstable and even larger nuclei cannot exist.

Below fundamental particles and forces is the theory of string theory, maybe. String theory hypothesizes that elemental particles are nothing more than vibrating strings which are trillions of times smaller than say a proton. The frequency with which a string vibrates determines what fundamental particle it forms. But at this point all of string theory is only theory and speculation; there is no empirical evidence to support it and strings are so small that many think there can never be empirical proof. So we will start the derivation from the lowest level for which there is empirical support, i.e., the fundamental level.

Social scientists are interested in the macro world, i.e., the world of highly aggregated variables and multiple aggregated causal pathways, so why would the derivation need to begin with the fundamental particles and causal forces of the universe? The answer is, because fundamental interactions are at the level that the universe knows about, i.e., the level at which the universe decides how to behave and the aggregation of these fundamental activities is the ultimate source of behaviors in the macro world.

A macro variable is simply a summation, an emergence, of the actions of trillions of fundamental objects and variables. Similarly macro causal connections are simply probabilistic regularities resulting from the behaviors of trillions of fundamental variables acting in accordance with the four fundamental forces. This is like a mudslide which is an aggregation of the behaviors of billions of sand and dust particles. Humans only comprehend the macro phenomena of a mudslide, but the mudslide itself acts only in accordance with the accumulated interactions of the individual sand and dust particles.

The universe knows nothing about macro objects and macro behaviors, like mudslides. It does its thing at the fundamental level and whether the outcome in the macro world is a hydrogen bomb or a gesture of kindness, is unknown to the inanimate universe.

So, to mathematically represent the world at the macro level of our perceptions and to understand the definitions and assumptions made in gaining that mathematical representation, we need to start with a mathematical representation of activity at the fundamental level and build up by aggregation and by appropriate simplifying definitions and assumptions, to a tractable mathematical representation of the macro variables and macro operations in our environment.

The derivation (which can be seen at www.causalstatistics.org) begins in a manner analogous to the beginning of the derivation of Euclidean geometry. Cause is inserted into the axiomization as a technical term and given a science oriented definition using the metalanguage. Then a set of foundational axioms (i.e., assumptions), designated as axiom set number one, is posited.

Axiom set number one is a highly updated version of Thales' foundational proclamation in 600 BC, i.e., natural events have natural causes. The most basic assumptions in the first axiom set are that 1) the operation of the universe is based on natural laws; 2) these natural laws are micro causal, i.e., causal laws operating between the fundamental particles of the universe; 3) these micro causal laws are local in time and space, meaning that there is no causation without temporal and spatial contact; 4) these natural, micro causal laws do not change over time or space.

The derivation then turns to the mathematical representation of a micro causal impulse traveling from one fundamental particle, through a micro chain of fundamental particles, to an ending elemental particle. Then, assuming the micro causal chain is long, the law of large numbers can be called on, by way of a very reasonable assumption, to remove the micro chain from the mathematical representation and described the causal connection in terms of a functional relationship between the beginning and ending fundamental particles only.

This exemplifies the beginning of a second assumption set, designed to simplify the mathematics of the derivation, while maintaining an acceptable level of accuracy. The purpose of this second axiom set is to 1) facilitate the derivation, 2) to keep the mathematical expressions to a convenient size, and 3) to mold the resulting inquiring system into a tractable form, readily usable by observational researchers.

Next the initial and final particles are expanded to macro objects and variables and the number of micro causal pathways are increased and logically grouped, ultimately resulting in macro causal chains. The mathematical formulation is expanded to handle this more complicated situation and again reasonable assumptions are made and added to assumption set number two to reduce and simplify the mathematical form, for example remove macro causal chains from the mathematical formulation.

As size moves from the world of subatomic physics toward the macro variables of sociology, everything changes. At the fundamental scale all objects of a given type, say electrons, are identical to each other. Causality is probabilistic, but due only to the Heisenberg uncertainty principle. As the derivation proceeds, fundamental variables are collected into macro variables and cascades of micro causal chains are aggregated into macro causal connections, with reasonable simplifying assumptions being expressed and inserted into the derivation and into assumptions set 2 all along the way.

This second set of axioms allows the derivation to progress without excessive complication. Some assumptions are utilized to get through a particularly difficult part of the derivation and then later relaxed, at a more convenient stage of the derivation. These fluctuating assumptions blink on and off like fireflies at dusk. Even with these simplifying axioms and techniques, the derivation was still quite complicated. In fact I really didn't know if it could be done until I reach the end.

From a non–mathematical point of view, as size increases, the nature of the observable macro world, including causal connections themselves, changes. The probabilistic effect of the Heisenberg uncertainty principle vanishes due to the law of large numbers and other stochastic sources come to the fore. A macro cause no longer operates in accordance with a precise law of the universe because a macro cause is a conglomeration. Macro variables and macro causal chains of the same types are each different from one case to the next. Macro objects can be as similar as baseballs or as different one from another as confidence flowing from athletic performance versus confidence stemming from intellectual performance versus confidence which is genetically determined.

Therefore an inferred macro cause is a causal regularity averaged over the different instances for a given macro variable and/or a given macro causal chain or set of chains. If one person with the confidence score of 60 is substituted with another person with the confidence score of 60, the causal connection to success in the same situation may be quite different for the two people. In making such a macro causal inference, we are not trying to discover some basic law of the universe (although maybe psychology), we are simply trying to identify a macro causal regularity between two fluctuating macro variables.

In the first major stop at the macro level, the derivation arrives at a discreet macro formulation for a single causal event or case. The linear formulation of this mathematical structure is as follows:

$$\begin{aligned}
Y_1 &\leftarrow \alpha_1 + \gamma_{12}Y_2 + \gamma_{13}Y_3 + \cdots + \gamma_{1M}Y_M + \beta_{11}X_1 + \cdots + \beta_{1K}X_{1K} \\
Y_2 &\leftarrow \alpha_2 + \gamma_{21}Y_1 + \gamma_{23}Y_3 + \cdots + \gamma_{2M}Y_M + \beta_{21}X_1 + \cdots + \beta_{2K}X_{2K} \\
&\vdots \\
Y_M &\leftarrow \alpha_M + \gamma_{M1}Y_1 + \gamma_{M2}Y_2 + \cdots + \beta_{M1}X_1 + \cdots + \beta_{MK}X_{MK}
\end{aligned} \tag{2}$$

The Y's are endogenous macro values for the specific case and the X's are exogenous macro values for the same case. The arrows represent the direction of causality. This mathematical representation, formulation (2), of our macro world describes one case, one event, with perfect accuracy, if and only if all of the assumptions made along the way are perfectly accurate. Of course not all of these assumptions are perfectly accurate, but that eases our derivation at this point and the assumption of perfect accuracy can later be relaxed.

Our interest is in representing a set of macro causal regularities for several cases and not for a single event, so we need to generalize formulation (2) to a class of events, i.e., to a set of a given type of instance. The linear formulation for a class of events is as follows:

$$\begin{aligned}
Y_1 &\leftarrow \alpha_1 + \gamma_{12}Y_2 + \gamma_{13}Y_3 + \cdots + \gamma_{1M}Y_M + \beta_{11}X_1 + \cdots + \beta_{1K}X_{1K} \\
Y_2 &\leftarrow \alpha_2 + \gamma_{21}Y_1 + \gamma_{23}Y_3 + \cdots + \gamma_{2M}Y_M + \beta_{21}X_1 + \cdots + \beta_{2K}X_{2K} \\
&\vdots \\
Y_M &\leftarrow \alpha_M + \gamma_{M1}Y_1 + \gamma_{M2}Y_2 + \cdots + \beta_{M1}X_1 + \cdots + \beta_{MK}X_{MK}
\end{aligned} \tag{3}$$

Now the Y's are endogenous macro variables and the X's are exogenous macro variables.

This of course leads to greater potential for error, in that the macro events, objects, and variables within the case will differ from each other at sub macro levels, usually leading to significantly different outcomes. These differences would introduce additional error into the system, but again by using one of those derivation assumptions which blink on and off like fireflies, this type of error is temporarily assumed away to facilitate development and comprehension of mathematical formulation (3).

Now we have a discreet macro causal representation for a given class of events and we need to modify it into statistical form, i.e., a form which admits and allows for error. First we need to blink off the assumptions that allowed no error and consequently add the error terms to the formulation. As errors enter into the system due to dissatisfaction of a few of the assumptions in axiom set two, the deviations from the assumed exactitude of mathematical form (3), will be soaked up by these added error terms. Hence the linear mathematical form for the statistical causal inquiring system can be represented as follows:

$$\begin{aligned} Y_1 &\leftarrow \alpha_1 + \gamma_{12}Y_2 + \gamma_{13}Y_3 + \cdots + \gamma_{1M}Y_M + \beta_{11}X_1 + \cdots + \beta_{1K}X_{1K} + e_1 \\ Y_2 &\leftarrow \alpha_2 + \gamma_{21}Y_1 + \gamma_{23}Y_3 + \cdots + \gamma_{2M}Y_M + \beta_{21}X_1 + \cdots + \beta_{2K}X_{2K} + e_2 \qquad (4) \\ &\vdots \\ Y_M &\leftarrow \alpha_M + \gamma_{M1}Y_1 + \gamma_{M2}Y_2 + \cdots + \beta_{M1}X_1 + \cdots + \beta_{MK}X_{MK} + e_M \end{aligned}$$

where the Y's are endogenous random macro variables and the X's are exogenous random macro variables. The added random error terms are represented by e's. Formulation (4) is the linear form of observational causal statistics. QED.

As I noted earlier, during the work on the 300 page axiomization (which can be viewed at www.causalstatistics.org), it was not at all clear to me that the goal of deriving an observational causal inquiring system was at all doable. In fact there were a number of points at which I concluded that it was likely impossible. Nevertheless, in the end I mathematically traveled from quarks to the macro variable of confidence. I came out of a long tunnel into the light of a tractable, yet surprising, formulation. I had hoped and expected that the derivation would arrive at a set of simultaneous equations and it did, almost, but not quite. The arrows are not equal signs.

The arrows indicate that these are simultaneous causal expressions, not equations. These arrowed expressions are unidirectional, as is causality. A change on the right–hand or tail side of the first expression would result in a change to the Y1 on the left–hand or arrowhead side of the expression. But a manipulation of the arrowhead side, Y1, would not modify the various random variables on the right–hand side of that expression, at least not in accordance with that first expression's functional form, because causal expressions don't work that way. They are only valid in one direction. Moving against the arrow is like dividing by zero; it's not allowed or, if you prefer, indeterminate. Manipulation of Y1 could affect the other Y's, but only in accordance with the other causal expressions (but not the first), each functioning in the proper direction.

6. Interpretation and Analysis of Results

When I first saw this unexpected formulation, I didn't understand it and that caused me a great deal of angst. But once I figured out what it meant, why it was there, and that and why it was better than the simultaneous equations I had expected I realized that the derivation was smarter than I was. Formulation (4) was more instinctively satisfying than simultaneous equations would have been, so I was unreservedly able to appreciate and embrace the new, arrowed formulation.

Looked at today, these arrowed expressions are evocative of Judea Pearl's famous do–calculus, used to predict the effects of causal manipulations, but I used no brainpower to obtain them. The arrowed expressions simply fell out of the derivation and into my lap, well before Judea's introduction of do–calculus.

In Einstein's axiomization of relativity, he did not end up at a new, unknown formulation. He came up with the Lorentz transforms, which had been developed empirically 17 years earlier on the basis of data from the Michelson–Morley experiment. Nevertheless, Einstein's derivation was a great accomplishment, not because it produced new formulations, but because it explained and justified the extant Lorentz equations and also because it delineated the assumptions/axioms required to arrive at these equations. One of the discovered assumptions was a doozy that was counter intuitive and a shock to everyone, viz, that the measured speed of light is the same no matter how fast the observer is moving. A derivation is like a box of chocolates . . . .

The instant derivation of observational causal statistics had a different outcome. None of the necessary assumptions were particularly remarkable, but the final formulation was different than expected and, upon analysis and reflection, more intuitively satisfying than the anticipated formulation.

Frequentist statistics, Bayesian statistics, experimental causal statistics, and observational causal statistics can all be considered branches of applied (as opposed to pure) mathematics, derived via the axiomatic method. Experimental causal statistics is an axiomatic extension of classical statistics and nonexperimental causal statistics is a sui generis, new, derived causal inquiring paradigm.

Now, instead of the generalized field of statistics being composed of two classical inquiring paradigms, I consider panoptic statistics to be composed of 3.5 statistical branches. I added 0.5 for experimental causal statistics, which is a significant extension of the classical fields, and I added 1.0 for observational causal statistics.

What is the relationship between the old and the new discourses? Are they in conflict with each other, unrelated, complementary, or what? Basically they are some form of complementary. The new 0.5 paradigm is an extension of classical statistics and therefore contains both causal inference abilities and the classical inference capabilities, as constituent parts. Therefore these inference engines work together in experimental causal statistics to infer causal connections within a sample and then to further infer the sample causal connections to a population. Therefore, the two parts of experimental causal statistics are complementary.

This is true for both physical science types of experiments and for Fisher/Cochran types of experiments. I subdivide these highly overlapping statistical extensions into 0.5a for the physical sciences and 0.5b for Fisher/Cochran type experiments in which potentially confounding variables are typically controlled through randomization.

The new 1.0 paradigm, i.e., observational causal statistics, is a nonexperimental causal inference tool. One could perform a nonexperimental study on a sample, make all the required assumptions, perform the analysis using the causal inquiring engine, and arrived at causal inferences for the sample. Then classical statistics would be needed to infer the sample causal connections to the population. Therefore classical statistics and observational causal statistics are complementary.

Actually they are more than complementary. Classical statistics without causal statistics can say nothing about causality, the most important conception in the sciences. Observational causal statistics can infer causal connections for a sample, but cannot transport those causal relationships to populations. They need each other, so they are complementary, and, beyond that, they can accomplish things together that neither can accomplish on its own, so they are synergistic.

On the other side of the coin, to the best of my knowledge no one has yet developed a procedure for estimating the parameters in these simultaneous causal expressions. As a matter of fact, there are many issues concerning observational causal statistics that need further research and development before the inquiring paradigm could be considered operational. I would characterize the development so far as being foundational and the current state as pointing the way forward, but not as ready for prime time. In fact I see the observational causal paradigm as being the fountainhead of The Statisticians Full Employment Act of 2014 and I look forward to a time when 25–50 percent of the articles in research journals touch on this inquiring discourse. We are probably at least 10 years away from observational causal statistics becoming fully operational and 30 to 50 years away from it becoming the sine qua non in social sciences research.

7. The Future of Statistics

I challenge you to think of something more important to the social sciences than the ability to draw causal inferences from nonexperimental data. I think that it is clear that there is nothing more important for those fields and I believe observational causal statistics is the key that fits that lock, albeit a very sticky lock due to the inherent difficulties of the problem. Our universe is as it is and to learn about it we can't do less than to touch all the logical bases, no matter how difficult that may be. Even so I predict that the nonexperimental causal inquiring paradigm, when embraced by the overall statistics community, will begin a long, but ultimately meteoric climb in usefulness, utilization, and importance.

Initially, the observational study designs will likely remain largely unchanged. But soon researchers will see the advantage of serial studies, where the initial study will be reformulated and rerun, considering more potentially causing variables. The big advantages of this type of redo is that the assumptions concerning the core variables of greatest interest can be weakened and that assumptions in general can be pushed out and away from those focal variables. After that, social science researchers should see the light and take the shortcut, i.e., design larger studies from the get–go, studies with 15 to 100 or more variables.

Eventually, nonexperimental causal statistics will usher in unforeseen magnitudes and types of research designs. After causal connections are inferred from some level of initial nonexperimental study, these causal relationships will be used in causal intervention projects (which, from a methodology point of view, can be viewed as experiments or at least quasi–experiments), designed to improve the human condition. The outcomes from this project should be monitored from both an evaluation point of view and a general research point of view. These data and results will then be combined with those of the initial study and the whole reanalyzed together to reach new and modified findings, sort of analogous to a Bayesian revision based on new information.

A further stage research design would be the beginning of a chain study. After the aforementioned combined data analysis, a new and larger empirical causal study is conducted with its design informed by findings from the combined analysis, mentioned above. Then the causal inference process begins anew, in the expanded follow–on study.

I believe that, because of the need and the power, observational causal statistics will eventually be considered the most important of the paradigms within panoptic statistics, but not at the expense of the other discourses. The swelling tide of value within global statistics, brought on by causal inquiry, will lift all boats.

”The development and utilization of observational causal statistics will eventually (within 100 years or so) be as important to the social/non–experimental sciences as the codification and utilization of the scientific method was to the physical/experimental sciences.”