

# Theory and Evidence in Economics

Julian Reiss, Durham University

CHES Working Paper No. 2019-06

[Produced as part of the Knowledge for Use (K4U) Research Project]

Durham University

July 2019



The K4U project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 667526 K4U) The above content reflects only the author's view and that the ERC is not responsible for any use that may be made of the information it contains

# Theory and Evidence in Economics

Julian Reiss

Julian Reiss  
Department of Philosophy  
Durham University  
50 Old Elvet, Durham DH1 3HN  
julian.reiss<at>durham.ac.uk

## 1. Introduction

Economics is a science of contradictions. On the one hand, ‘Economics is awash in data, and the vast majority of published articles are empirical’ (Hoover 2009: 202). On the other hand, economists have the reputation not to take the empirical findings too seriously (Krugman 2009):

Unfortunately, this romanticized and sanitized vision of the economy led most economists to ignore all the things that can go wrong. They turned a blind eye to the limitations of human rationality that often lead to bubbles and busts; to the problems of institutions that run amok; to the imperfections of markets — especially financial markets — that can cause the economy’s operating system to undergo sudden, unpredictable crashes; and to the dangers created when regulators don’t believe in regulation.

Of course, the two beliefs — that most work in economics is empirical and that economists tend to ignore empirical evidence — do not directly contradict each other. But why would these economists engage in the collection and analysis of evidence just to be ignored by their colleagues? We will see that the role theory plays in the collection and analysis of evidence is key to understanding differences in attitude.

Specifically, this chapter will argue that the relationship between theory (the neoclassical version of which Krugman derides as ‘this romanticised and sanitised vision of the economy’) and evidence is highly contested in economics and has been since its inception. I will first revisit a number of episodes in the history of economics in order to illustrate that the relationship between theory and evidence has been at the centre of the methodological debate from the heyday of classical economics onwards. I will then argue that there is no consensus on the issue in contemporary economics either. The final section will present some speculation about where this debate might go in the future.

## 2. Theory and Evidence in the History of Economic Thought

Adam Smith is frequently given the epithet ‘father of political economy’ (e.g., Haakonssen 2006: 4). His methodology of studying the phenomena of political economy was highly eclectic (e.g., Viner 1968: 327), however, and hardly resembles contemporary mainstream approaches. If we were to look for a father of modern economic *methodology*, we would have to look among the Ricardians — the disciples, defenders, and disseminators of David Ricardo’s economics and its methodology. Ricardo was the first systematic proponent of the *deductive* approach in political economy — the building of abstract models, aided by rigid and somewhat artificial definitions, and syllogistic reasoning (Sowell 1974). In order to arrive at principles of high generality, he isolated in thought what he considered the main causal factors from disturbing factors, held the latter constant and thus hoped to establish the unchanging elements common to all economies: ‘I put these immediate and temporary effects quite aside, and fixed my whole attention on the permanent state of things which will result from them’ (quoted from Sowell 1974: 113-4).

It is important to note that Ricardo's deductive methodology has little to do with the hypothetico-deductive approach with which contemporary philosophers are familiar. In particular, Ricardo had much to say about the generation of the principles that would form the basis for deductions. In order to determine the principles to describe the distribution of profits among the different classes in society (landowners, farmers/capitalists, and labourers), he built a 'model farm' as a stand-in for the whole economy out of numerous conceptual and empirical elements, including definitions, the labour theory of value, the tendency law of the profit rate to fall, and the experimental farming tradition (Morgan 2012: 79). Ricardo's 'model farm' was thus partly *a priori* and partly empirical, because the numbers used in his numerical cases were typical of his place and time, Ricardo's numerical accountings mirrored the reports of experimental farming, and the examples he chose all related to lively discussions in the political economy of his day (Morgan 2012: 76-8). In Ricardo's (and later Ricardian) economics, evidence thus played a twofold role: first, in the elaboration of the fundamental principles, and second, in the testing of the consequences of these principles.<sup>1</sup> John Stuart Mill later formalised and defended the Ricardian approach as inevitable. His main argument built on the unavailability of controlled experimentation in economics (Mill 1844).<sup>2</sup>

The Ricardians' most vocal methodological critics were the Cambridge Inductivists, especially Richard Jones and William Whewell.<sup>3</sup> Jones lamented especially the hasty and broad generalisations made from a narrow evidential base. He proposed instead that economists should pay much closer attention to the evidence right from the start, and slowly and gradually generalise from what they learn (Jones 1831: xxxix-xl):

During this process, the too hasty erection of whole systems, a frail thirst for the premature exhibition of commanding generalities, will probably continue to be the sources of error most to be guarded against. It is, assuredly not by indulging and encouraging such errors that the boundary of human knowledge in this direction will be successfully or safely approached. The portions of truth which can in the first instance be safely attained, must necessarily be narrow principles, grounded upon a limited field of experience, cautiously and patiently worked out. Wider generalities of more scientific simplicity, can only be approached after these intermediate truths have been mastered. This is the appointed course of true and permanent science. To spring at once from partial and broken observations to the most general axioms; to dart from a state of ignorance and confusion upon the fundamental and ultimate elements of systematic knowledge, without touching the ground during the intermediate flight: this is the course of a rash theorist, and not of a philosopher; and those who have often tracked that course, must know but too well, that the very simplicity and commanding aspect of propositions so attained, is much oftener a warning of the insecurity of their application, than any evidence of their truth.

---

<sup>1</sup> In fact, Ricardo's principles didn't do so well in predicting returns, prices, rents and profits (Blaug 1958: 182). I will come back to this point further below.

<sup>2</sup> This too is a point to which I'll return.

<sup>3</sup> The Cambridge inductivists were a group of scholars who had read Francis Bacon's *Novum Organum* together as undergraduates and aimed to remake science in the image of this work. Apart from Jones and Whewell, Charles Babbage, the inventor of the first mechanical computer, and John Herschel, a notable astronomer, were members of the group. The economist Thomas Malthus is also sometimes associated with the group (Hollander 2017).

The debate between the Ricardians and the Cambridge Inductivists exemplifies two fundamentally different approaches to learning from evidence in economics. The Ricardians build fundamental principles from a number of *a priori* elements and casual observations and deduce from these principles (jointly with facts about the concrete systems the principles are applied to and simplifying assumptions) predictions that can be tested against evidence, as well as policy recommendations.<sup>4</sup> Importantly, the fundamental principles were thought to be universally applicable. The Cambridge Inductivists proposed instead to begin with a broader and much more systematic investigation of the evidence, limiting inferences to go not much beyond what has been observed and thereby allowing principles to be time- or geographically specific. At any rate, the concepts and principles the inductivists established remained tied to specific nations, regions, cultures, and epochs.

Though schools that would implement and defend the inductivist approach to studying socio-economic phenomena flared up repeatedly in the history of economics — most famous was the German Historical School that defended inductivism in the *Methodenstreit* with the leader of the Austrian School of Economics, Carl Menger (see for instance Schumpeter 1954/2006: 782-3; Reiss 2008: Ch. 1) — the deductive approach remained dominant. It will prove instructive to examine a debate about reasoning with evidence that emerged after the Second World War. The debate was informed by the birth of the new discipline of econometrics. In 1946, Arthur Burns and Wesley Clair Mitchell, both associated with American Institutionalists (whose work was in turn strongly influenced by the inductivist German Historical School), published a book about business cycles that set out to approach its topic in a manner as free of *a priori* theory as possible (Burns and Mitchell 1946: 8; emphasis original):

Systematic factual research is often thought of as belonging to the stage of 'inductive verification', and 'inductive verification' as a step to be taken *after* a 'theory' has been excogitated. Of course no writer has ever attempted to devise a hypothesis concerning business cycles entirely apart from the facts to be explained. But theorists have not infrequently been handicapped by a sadly incomplete, sometimes by a badly twisted, knowledge of the facts. Numerous writers have invented plausible explanations of business cycles before they have tried to ascertain what consensus actually prevails among the cyclical movements of different economic activities...

When this order of inquiry is followed—explanation preceding thorough knowledge of what is to be explained—the results are likely to be unhappy. (1) The theorist often stops before his work is finished, leaving 'inductive verification' to others, who may or may not take on the job. (2) When anyone tries to 'verify' a hypothesis about business cycles, he often finds that it rests on assumptions purposely chosen to simplify the situation that is analyzed. In that case evidence drawn from the actual world has a problematical relation to the simpler world of the theorist's imagination; the hypothesis propounded may be logically impeccable, but it cannot be confirmed or refuted by an appeal to facts. (3) Granted that the hypothesis concerns actual experience, the worker who tries to verify it must examine the processes on which it centers attention; but unless he examines other processes as well, the test will be superficial.

In a review of their book, Dutch-American econometrician Tjalling Koopmans took up Burns and Mitchell's inductivism (Koopmans 1947). Koopmans distinguished what he called a 'Kepler stage'

---

<sup>4</sup> The derivation of policy recommendations from fundamental principles and simplifying assumptions has been termed the 'Ricardian Vice' (Schumpeter 1954/2006: 448).

of inquiry from a 'Newton stage', the former aiming to establish empirical regularities and the latter, fundamental laws. He argued that using the concepts and hypotheses of economic theory as a part in the processes of observation and measurement would lead to a better understanding of the phenomenon of business cycles. His reasons were the following.

My *first argument*, then, is that even for the purpose of systematic and large scale observation of such a many-sided phenomenon, theoretical preconceptions about its nature cannot be dispensed with, and the authors do so only to the detriment of the analysis. (*ibid.*: 163; emphasis original)

This, then, is my *second argument* against the empiricist position: Without resort to theory, in the sense indicated, conclusions relevant to the guidance of economic policies cannot be drawn. (*ibid.*: 167; emphasis original)

[A]ny rigorous testing of hypotheses according to modern methods of statistical inference requires a specification of the form of the joint probability distribution of the variables. [...]he the extraction of [useful] information from the data requires that, in addition to the hypotheses subject to test, certain basic economic hypotheses are formulated as distributional assumptions, which often are not themselves subject to statistical testing from the same data. Of course, the validity of information so obtained is logically conditional upon the validity of the statistically unverifiable aspects of these basic hypotheses. The greater wealth, definiteness, rigor, and relevance to specific questions of such conditional information, as compared with any information extractable without hypotheses of the kind indicated, provides the *third argument* against the purely empirical approach. (*ibid.*: 170; emphasis original).

The first argument foreshadows a number of points made by Carl Hempel in his criticism of what he called 'naive inductivism' (Hempel 1966: 11-13). Hempel argued that 'theory-free' collection of evidence could never get off the ground because unless scientists wanted to continue collecting evidence until the end of the world, they needed an idea of what the *relevant* evidence was. Relevance, Hempel continued, was not determined by the scientific problem alone but by the scientific problem in conjunction with tentative answers to the scientific problem, that is, with a hypothesis. Evidence is therefore always *relative to a hypothesis*. Moreover, even if evidence could be gathered, without hypotheses it is not clear how to classify, aggregate and interpret the evidence (Hempel 1966: 13).

Koopmans' third argument relates to the fact that modern statistical hypothesis testing proceeds against a backdrop of assumptions about the data-generating process, for instance about the functional form in which variables are related, the distribution of the error term (that describes the behaviour of omitted variables) and the causal relations among the measured variables as well as the omitted variables. This information cannot be derived from an observation of the data sets studied and must therefore come from prior knowledge. According to Koopmans, economic theory can and ought to play this role.

The second argument will prove particularly significant for the contemporary debate. As discussed above, one aspect of the inductivist approach is that great caution was exercised in taking inferences well beyond the evidence at hand. Specifically, what was learned in one epoch or region or context was not assumed to travel to other epochs, regions or contexts easily. But this caution makes policy recommendations difficult. On the one hand, policies are rarely applied to

situations that are exactly like the situations from which the evidence is drawn. On the other, policies often create entirely new situations. For both reasons inferential prudence will undermine the ability to make post-policy predictions.

Proponents of the deductivist approach maintain that they get a handle on this problem by aiming to measure the fundamental behavioural parameters in systems of equations. Behavioural regularities change gradually through changing habits and tastes, urbanisation and industrialisation, incremental technological change; it may change abruptly through important innovations, economic policies and large-scale political events (Koopmans 1947: 167). By measuring the fundamental behavioural parameters, Koopmans thought the effects of such changes could be made predictable.

Rutledge Vining responded to Koopmans on behalf of the inductivist approach. He agreed that economic theory *could* help in causal inference but pointed out that economics didn't have a theory yet that was good enough to perform well in this role. Therefore, economists have to examine the facts in order to come up with theoretical hypotheses that, after they've passed the empirical test, can be used for further inferences (Vining 1949). His point was, in other words, that though helpful if available, the economists of the time simply didn't have theory that was credible enough to play significant roles in the collection, classification, aggregation, and interpretation of evidence and inferences from it at their disposal.

### 3. The Evidence Wars in Contemporary Economics

In 1983 the econometrician Edward Leamer published an influential article titled 'Let's Take the Con Out of Econometrics' (Leamer 1983). In this article, Leamer argued that the assumptions econometricians needed to make to be able to identify parameters of interest were 'whimsical' and that therefore the economics community 'consequently and properly withholds belief' (*ibid.*: 43).<sup>5</sup> The problem remained with econometrics in the intervening years. On the one hand, theory-free inference from the evidence does get off the ground. On the other, there is no credible theory and so different factions within economics use different sets of background assumptions and do not trust each other.

The year 2010 finally brought an alleged 'Credibility Revolution' (Angrist and Pischke 2010). Angrist and Pischke argued that *randomised trials* provided a benchmark for applied econometrics (*ibid.*: 12). In order to assess the significance of this claim, I need to backtrack a little bit.

As mentioned above, John Stuart Mill argued that the inductive method was not fruitful in economics because of the absence of controlled experimentation. The classical economists, of

---

<sup>5</sup> He proposed in response that econometricians test the robustness of results to specification changes. His approach, called extreme bounds analysis (e.g., Leamer 1995) has been taken up in econometrics, but it is not generally thought to solve the problem, in part because results are rarely robust to specification changes (see for instance Sala-i-Martin 1997).

which Mill was the last prominent representative before the marginal revolution brought about neoclassical economics in the 1870s, were mainly interested in macro questions such as the distribution of income among the different social classes, the effects of population growth on incomes, and the effects of free trade. Indeed one cannot perform experiments that address these kinds of questions directly — we cannot construct, and do not find, two economies that are exactly identical except with respect to trade policy and its effects. Under the assumption that experimentation is the benchmark for inductive inference, but that relevant experiments could not be performed in economics, Mill's rejection of induction was consonant with widely held beliefs at his time.

But the times have changed. After the marginal revolution it was possible to address questions about the behaviour of individuals (consumers, firms), and thus micro economics was created. Microeconomic questions *can* be addressed experimentally.<sup>6</sup> The other important event was the emergence of the randomised trial.<sup>7</sup> A randomised trial is essentially a probabilistic version of Mill's controlled experiment. Whereas in Mill's experiment we want two situations that are exactly identical except with respect to some factor of interest and its potential effects, in a randomised trial, we have two groups of experimental subjects that are identical with respect to the probabilistic distribution of factors that may make a difference to an outcome of interest. When there are many such factors and not all factors can be controlled or even measured, a randomised trial may constitute a valuable alternative to the controlled experiment.

Experimental economics goes back to attempts to estimate demand functions experimentally in the 1930s (Moscati 2007) but the field began to boom in earnest in the 1980s and 1990s (Guala 2008). According to Alvin Roth, experimental economists pursue at least three purposes (Roth 1986). They:

- *speak to theorists* (i.e., they test formal economic theories and potentially modify them in response);
- *search for facts* (i.e., they collect data on interesting phenomena and important institutions);  
and
- *whisper into the ears of princes* (i.e., they provide input into the policy-making process).

Experiments can thus be a part of a more deductive theory-testing approach and a more inductive fact-searching approach, and indeed it has been argued that 'experimental economics suggests that the entire approach to thinking about the appropriate mix of induction and deduction needs to be rethought' (Colander et al. 2004: 494). We will see momentarily that the relationship between theory and evidence is as controversial today as it was nearly two centuries ago.

---

<sup>6</sup> Whether fruitfully or not remains controversial, however. I will come back to this below.

<sup>7</sup> Randomised trials originated in psychology in the late 19th century and were widely used in agriculture in the first third of the 20th century (Hacking 1988), before conquering the field of medical research (Marks 2000: Ch. 5).



Angrist and Pischke point to a number of reasons why today's econometric estimations are considered more credible than those of previous generations of econometricians. Their main reason is, however, better research design (Angrist and Pischke 2010: 12; emphasis original):

In applied micro fields such as development, education, environmental economics, health, labor, and public finance, researchers seek real experiments where feasible, and useful natural experiments if real experiments seem (at least for a time) infeasible. In either case, a hallmark of contemporary applied microeconometrics is a conceptual framework that highlights specific sources of variation. These studies can be said to be *design based* in that they give the research design underlying any sort of study the attention it would command in a real experiment.

The econometric methods that feature most prominently in quasi-experimental studies are instrumental variables, regression discontinuity methods, and differences-in-differences-style policy analysis. [...] The best of today's design-based studies make a strong institutional case, backed up with empirical evidence, for the variation thought to generate a useful natural experiment.

The final sentence of the quote is key. The best studies are grounded in institutional background knowledge and empirical evidence. Theory is ostensibly absent. Not surprisingly, after what I have argued above, I maintain that the econometricians who follow a design-based approach are today's analogues of the inductivists of the past.

When a new inductivist school emerges, we don't have to wait long for deductivist criticism to follow suit. One of the most vocal critics of the design-based approach is Angus Deaton. He took up both the randomised trials that his opponents accept as benchmark for credible inferences as well as the quasi-experimental techniques that are designed to mimic randomised trials (Deaton 2010a, Deaton 2010b, Deaton and Cartwright 2017).

It is useful to distinguish between the internal and the external validity of an experimental result (e.g., Reiss 2008: Ch. 5). The result is internally valid whenever it is true of the experimental system studied. It is externally valid whenever it is true of some other, as yet unstudied system of interest.<sup>8</sup> Deaton argues that there are no reasons to believe good design makes estimates of parameters of interest automatically more credible. Some of the problems can successfully be addressed — by using theory.

*Threats to internal validity.* In social science applications, it is not typically possible to conduct a double-blind trial, which may create various biases. Among other things, subjects may refuse the treatment they are assigned to or get their preferred treatment elsewhere, and this behaviour may be correlated with other prognostic factors.<sup>9</sup> Sometimes allocation according to alphabet ('alphabetisation') is used as a substitute of randomisation, but alphabetisation does not guarantee that treatment allocation is independent of prognostic factors, even in the limit.

---

<sup>8</sup> In a medical setting, a trial result is said to be internally valid if it is true of the patients in the trial; it is externally valid if it is true, for instance, of all sufferers of a certain disease or some other group of patients not in the trial.

<sup>9</sup> Deaton here refers to James Heckman as maintaining that 'the deviations from assignment are almost certainly purposeful, at least in part' (Deaton 2010a: 445; cf. Heckman 1992).

*Threats to external validity.* Experimental populations are rarely randomly drawn from the general population so that it is unlikely that the experimental result can be generalised. Risk-averse subjects might, for instance, not agree to be part of an experiment because doing so amounts to playing the lottery. The experiment itself might also create artefacts so that the applicability of results outside the trial becomes problematic. There might be ‘general equilibrium effects’, i.e., changes in effect due to the fact that everyone is treated instead of a small sample.

*Limited informativeness.* Even an ideal randomised trial only guarantees (in the limit) the correctness of the *average treatment effect*, which can be estimated by the difference in mean outcomes between the experimental and the control group. The trial on its own does not identify other features of the distribution. Thus, without further assumptions, quantities such as the difference in median outcome (which is likely to be of interest to policy makers) cannot be computed. Nor can it be determined whether the difference in means is significant.<sup>10</sup>

Many of these problems can be solved or at least ameliorated, but only by making strong assumptions that come from theory. If we have a theory about why subjects take up a treatment, we can predict whether he or she will accept or refuse a given treatment status and thus correct for the associated bias. In order to perform a significance test, a regression can be run on the trial data. However, regression models require strong background assumptions, the justification of which relies on theory. In the context of considerations about external validity, theory is needed for at least two reasons, according to Deaton. First, from an economist’s point of view, an average treatment result is uninteresting because he or she is interested in the *welfare consequences* of policies. In order to assess these, a theoretical model must be built. Second, as there are reasons to believe that populations outside of the experiment differ systematically from the experimental population, we need a theoretical model that predicts how the treatment effect varies with characteristics of the individuals in these populations.

Analogous considerations apply to quasi-experimental techniques. These are valid only under strong causal assumptions. A valid instrumental variable, for instance, (a) causes an independent variable; (b) does not cause the dependent variable except through the independent variable; and (c) is not caused by or does not have causes in common with other causes of the dependent variable (Reiss 2005). Deaton argues that assumptions such as these cannot be made in a credible manner unless there is a theoretical model within which a variable can be shown to be a valid instrument (Deaton 2010a: 428). Summarising his stance on the design-based movement in econometrics, he writes (*ibid.*: 450):

---

<sup>10</sup> There are also ethical problems related to this point. A positive difference in means can mask the fact that the treatment, though effective on average, has harmful effects for some subpopulations. The trial, again, does not identify the subpopulations for which the treatment was effective. Policy makers might not want to implement policies that are harmful to some groups or target specific groups such as the least well off (Khosrowi 2019).

In the end, there is no substitute for careful evaluation of the chain of evidence and reasoning by people who have the experience and expertise in the field. The demand that experiments [and quasi-experimental techniques] be theory-driven is, of course, no guarantee of success, though the lack of it is close to a guarantee of failure.

The problem is, of course, that lack of trust in theory is what gave rise to the development and popularity of the design-based approach to econometrics in the first place. When it comes to reasoning from evidence in contemporary economics, we seem to be stuck between a rock and a hard place.

#### 4. Quo Vadis, Evidence in Economics?

Can the dilemma be overcome? Perhaps it would be presumptuous for me to try to resolve a nearly 200-year old debate that has engaged the brightest minds of their time on both sides, but let me make a proposal. The proposal has two main ingredients: middle-range theory and inferential contextualism.

*Middle-range theory.* To some extent, the choice between deductivist and inductivist approaches to evidential reasoning in economics is based on a false dilemma. When deductivists talk about theory, they typically mean 'grand theory' such as the principles of neoclassical economics, those of the Austrian School or those of Post-Keynesian economics. What characterises these principles is their universal nature: they are intended to apply to all individuals and all economies at all places and all times. Perhaps rightly, inductivists are sceptical of such universal theoretical frameworks. But by demanding a 'theory-free' approach, they throw out the baby with the bathwater. The reason is that inductions cannot get off the ground without *some* theory. A simple example will illustrate.

Suppose a detective wants to find out who is behind a series of killings and investigates the hypothesis that 'Jones was the serial killer'. What evidence is relevant to the hypothesis? According to the hypothetico-deductive model of confirmation, any observable statement that is deductively entailed by the hypothesis (Hempel 1966). But the hypothesis does not entail anything on its own. What is needed are behavioural generalisations such as 'Serial killers are often psychopaths' and the many generalisations about psychopaths psychology provides. Given these and other generalisations, an investigator can form expectations about what clues to look for.<sup>11</sup>

A generalisation such as 'serial killers are often psychopaths' is a middle-range theory. It is intended to apply to a very narrow class of individuals, it explicitly allows of exceptions, it may be qualified in all sorts of ways, and its content may change over time. Nevertheless, it goes well

---

<sup>11</sup> None of these will make the inference from hypothesis and background generalisations deductive, however. One reason is that few of the generalisations that drive the inference are free of exceptions.

beyond the existing evidence, it is explanatory, and it can play the role of an enabler and justifier of an inductive inference.<sup>12</sup>

Even though economists do not talk in these terms, the discipline is full of middle-range theories: people respond to incentives; they tend to think harder about decisions when the stakes are higher; they are subject to status quo bias and the endowment effect; recessions are caused by, among other things, increasing interest rates, stock market crashes, a decline in aggregate demand and price controls; significant increases in government spending (causally) contribute to GDP growth. More limited in scope, such theories are better placed to drive inductive inferences because they are more accessible to empirical scrutiny and development and are therefore more likely to command widespread consensus.<sup>13</sup>

*Inferential contextualism.* The inferential contextualist denies the existence of secure foundations, i.e., claims regarded as certain, independently of context, and therefore not subject to critical scrutiny. He also denies the existence of universal standards for justifying epistemic success. Inferences proceed against a backdrop of postulates that include both factual assumptions — such as middle-range theories — but also ethical, methodological, and conceptual norms, a case for which must be made against the goals and purposes and the available background knowledge that characterise the concrete context of the inference (Reiss 2015). John Norton uses the slogan ‘All induction is local’ to describe this feature of inferential contextualism (Norton 2003).

Inferential contextualism enables the iterative improvement of claims to knowledge at different levels. One inquiry make take some middle-range theory as given for the context of that inquiry and use it to power an inductive inference about a claim concerning a particular matter of fact. This new factual claim can then be used to confirm or disconfirm related middle-level theories in separate inquiries. The goal of this process is to develop a robust body of reliable middle-range theories that is broad enough to power all or most desired inferences, but also secure enough so that few inferences go astray.

I believe that the proposal sketched here can help to make learning from evidence in economics more cumulative, but it does come at a cost, both for deductivists and inductivists. Deductivists have to give up their attachment to ‘grand theory’ that is responsible, for instance, for the demand that claims about the behaviour of macroeconomic aggregates be built on solid microfoundations. Middle-range theory straddles the micro/macro divide, and the proposal made here assumes that some robust macro theories of the middle-range exist independently of whether or not a microfoundation can be supplied. Following inferential contextualism, the inductivist has to give up the idea that there are any secure foundational claims about

---

<sup>12</sup> The term ‘middle-range theories’ was originally introduced by Robert Merton and defined as ‘theories that lie between the minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behavior, social organization and social change’ (Merton 1968: 68).

<sup>13</sup> For a discussion of middle-range theories in the context of social policy, see Cartwright (forthcoming).

observables, facts, evidence, data or what have you. He also has to give up the idea that there are universal and mechanical rules for learning from evidence. All inductive inference is local, contextual, and is powered by items of substantial background knowledge, i.e., middle-range theories.

Whether a proposal such as this can be successful will depend on whether economists are able to build a substantial stock of middle-range theories that most researchers working in the profession can accept. The level of controversy that surrounds almost any claim to knowledge in economics may be cause for pessimism, but let us be patient and allow time to tell.

## Bibliography

- Angrist, Joshua and Jörn-Steffen Pischke 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives* **24**(2): 3-30.
- Blaug, Mark 1958. *Ricardian Economics*. New Haven (CT), Yale University Press.
- Burns, Arthur and Wesley Clair Mitchell 1946. *Measuring Business Cycles*. New York (NY), National Bureau of Economic Research.
- Cartwright, Nancy (forthcoming). Lullius Lectures 2018: Mid-level theory: Without it what could anyone do? In C. Martínez Vidal and C. Saborido (Eds.), Nancy Cartwright's Philosophy of Science. Special Issue of *Theoria*.
- Colander, David, Richard Holt and Barkley Rosser Jr. 2004. "The Changing Face of Mainstream Economics." *Review of Political Economy* **16**(4): 485-499.
- Deaton, Angus 2010a. "Instruments, Randomization, and Learning About Development." *Journal of Economic Literature* **48**(2): 424-455.
- Deaton, Angus 2010b. "Understanding the Mechanisms of Economic Development." *Journal of Economic Perspectives* **24**(3): 3-16.
- Deaton, Angus and Nancy Cartwright 2017. Understanding and Misunderstanding Randomized Controlled Trials. Working Paper. National Bureau of Economic Research. Cambridge (MA), NBER.
- Guala, Francesco 2008. Experimental Economics, History Of. *The New Palgrave Dictionary of Economics*. Vol. 3. Steven N. Durlauf and Lawrence E. Blume, Eds. London, Palgrave-Macmillan: 152-156.
- Haakonssen, Knud 2006. Introduction: The Coherence of Smith's Thought. *The Cambridge Companion to Adam Smith*. Knud Haakonssen, Ed. Cambridge, Cambridge University Press: 1-21.
- Hacking, Ian 1988. "Telepathy: Origins of Randomization in Experimental Design." *Isis* **79**(3): 427-451.
- Heckman, James 1992. Randomization and Social Policy Evaluation. *Evaluating Welfare and Training Programs*. Charles F. Manski and Irwin Garfinkel, Eds. Boston (MA), Harvard University Press: 201-230.
- Hempel, Carl 1966. *The Philosophy of Natural Science*. Upper Saddle River (NJ), Prentice-Hall.
- Hollander, Samuel 2017. Thomas Robert Malthus (1766–1834). *The Palgrave Companion to Cambridge Economics*. Robert Cord, Ed. London, Palgrave Macmillan: 233-256.
- Hoover, Kevin 2009. "How Can Economics Be an Inductive Science?" *Economics and Philosophy* **25**(2): 202-206.
- Jones, Richard 1831. *An Essay on the Distribution of Wealth and the Sources of Taxation*. London, John Murray.
- Khosrowi, Donal 2019. "Trade-Offs between Epistemic and Moral Values in Evidence-Based Policy." *Economics and Philosophy* **35**(1): 49-71.

- Koopmans, Tjalling 1947. "Measurement without Theory." *Review of Economic Statistics* **29**(3): 161-171.
- Krugman, Paul 2009. How Did Economists Get It So Wrong? New York Times.
- Leamer, Edward 1983. "Let's Take the Con out of Econometrics." *American Economic Review* **73**(1): 31-43.
- Leamer, Edward 1995. "Sensitivity Analyses Would Help." *American Economic Review* **75**(3): 308-313.
- Marks, Harry 2000. *The Progress of Experiment: Science and Therapeutic Reform in the United States, 1900–1990*. Cambridge, Cambridge University Press.
- Merton, Robert 1968. *Social Theory and Social Structure*. New York (NY), Free Press.
- Mill, John Stuart 1844. *On the Definition of Political Economy; and on the Method of Investigation Proper to It. Essays on Some Unsettled Questions of Political Economy*. London, Parker: 120-164.
- Morgan, Mary 2012. *The World in the Model: How Economists Work and Think*. New York (NY), Cambridge University Press.
- Moscatti, Ivan 2007. "Early Experiments in Consumer Demand Theory: 1930-1970." *History of Political Economy* **39**(3): 359-401.
- Norton, John 2003. "A Material Theory of Induction." *Philosophy of Science* **70**(4): 647-670.
- Reiss, Julian 2005. "Causal Instrumental Variables and Interventions." *Philosophy of Science* **72**(PSA 2004): 964-976.
- Reiss, Julian 2008. *Error in Economics: Towards a More Evidence-Based Methodology*. London, Routledge.
- Reiss, Julian 2015. "A Pragmatist Theory of Evidence." *Philosophy of Science* **82**(3): 341-362.
- Roth, Alvin E. 1986. "Laboratory Experimentation in Economics." *Economics and Philosophy* **2**(02): 245-273.
- Sala-i-Martin 1997. "I Just Ran Two Million Regressions." *American Economic Review* **87**(2): 178-183.
- Schumpeter, Joseph 1954/2006. *History of Economic Analysis*. Taylor & Francis e-Library edition. New York (NY), Routledge.
- Sowell, Thomas 1974. *Classical Economics Reconsidered*. Princeton (NJ), Princeton University Press.
- Viner, Jacob 1968. Smith, Adam. *International Encyclopedia of the Social Sciences*. Vol. 14. David Sills, Ed. New York (NY), Macmillan and the Free Press.
- Vining, Rutledge 1949. "Methodological Issues in Quantitative Economics." *Review of Economics and Statistics* **31**(2): 77-86.