

Probability and Statistics in the Tinbergen-Keynes Debates

WILLIAM PEDEN

Lingnan University

I. INTRODUCTION

Erwin Dekker's *Jan Tinbergen (1903-1994) and the Rise of Economic Expertise* (2021) provides a rare treasure in the history of economics: from a single volume, we can obtain a sense of both Jan Tinbergen's worldview and historical context. The worldview was egalitarianism, optimistic about scientific expertise, and hopeful that internationalism could avoid the horrors of the First World War, which occurred in Tinbergen's formative years. His historical context was the early years of several parts of modern economics, including development economics, macroeconomics, environmental economics, econometrics, and welfare economics.

Dekker details Tinbergen's major influence across all of these areas. I shall consider just one instance of Tinbergen's wide-ranging impacts. He was a foundational figure in the development of macrodynamics: the study of systemic changes in an economy over time, such as business cycles. Tinbergen's *Statistical Testing of Business Cycle Theories* (1939a, 1939b) published by the League of Nations and his earlier study (Tinbergen, 1936) were the first large-scale and rigorous formal excursions into macrodynamics.

Like any highly original science, Tinbergen's research attracted criticisms, including from future Nobel Prize winners (Friedman 1940; Haavelmo 1943). By far the most famous critique was by John Maynard Keynes (1939, 1940). In brief, Keynes argued that (1) Tinbergen failed to identify and satisfy the conditions for using correlation analysis, (2) he used this technique badly, and (3) he selected an unpromising subject matter: business cycles. In short, Tinbergen was employing "black magic [...] a branch of statistical alchemy" (Keynes 1940, 156) with almost no significance for the tumultuous business cycle theory debates of the 1930s and 1940s. Keynes endorsed statistical reasoning in economics, but

only to estimate the parameters that are important in particular theories, not to test them.

Chapter 8 of Dekker's book adds to our understanding of the Tinbergen-Keynes debates, especially their origins within internal League of Nations discussions, which included Keynes. Dekker reveals that there was a persistent misunderstanding, because Keynes interpreted Tinbergen as stridently presenting a finished (or nearly finished) method to solve major macroeconomic debates. Meanwhile, Tinbergen took an optimistic but modest view of his research as a preliminary excursion into testing business cycle theories, yet his confident rhetoric in the 1939 volumes belied his private circumspection.

Furthermore, Dekker explains how Tinbergen's research for the League of Nations, as well as Keynes' response, took place within a wider socio-political context. At the time, economics was still strongly divided into separate schools. Most business cycle theories were associated with a particular country: there were various American, Austrian, British, French, German, Norwegian, and Swedish schools of thought. Tinbergen saw his macroeconometric project as an attempt to combine insights from each approach. Yet, as Dekker explains, Tinbergen's decision was not purely scientific. His internationalist approach to theory was parallel to his internationalist values: just as he hoped to break down national barriers in economics, so Tinbergen yearned to break down the barriers to international coordination and peace (165-172). Meanwhile, Dekker points out how Keynes' scepticism and rhetoric also seems to have been partly political. By the time of Keynes' critique, the technocratic and institutional approach favoured by the League of Nations had clearly failed to create a stable international order. The League of Nations would continue for six more years, but it was insignificant. In Keynes' view, the economic research that the League had produced through Tinbergen was also insignificant. For Keynes, the future was in the hands of greater leaders, partnered with men of intuitive brilliance, like himself (186-187).

In addition to this contextualisation of the Tinbergen-Keynes debates in the history of international politics, Dekker also provides us with their context in the development of macroeconomics. Tinbergen's research was a high point in a much larger econometric movement in the Inter-War period. Dekker points out the considerable irony that, in hindsight, subsequent syntheses of Tinbergen and Keynes would overshadow Keynes' objections. While Keynes was harshly pessimistic about Tinbergen's methods, many Keynesians (especially Americans) shared Tinbergen's

optimism. Hence, they sought combinations of (a) Keynes' macroeconomic perspective and (b) Tinbergen's ambitions to add statistical methods to macroeconomists' toolkit. With many intermediary vicissitudes, this synthesis survives and thrives in macroeconomics (163–164).

Dekker's biography focuses on the difficult challenge of explaining Tinbergen's worldview and wider historical context to a modern audience. One aspect of this worldview was Tinbergen's robust aversion to discussing philosophical issues explicitly (415). Even in his replies with Keynes, Tinbergen focused relentlessly on the technical challenges, such as whether Tinbergen had tested his estimates using multiple subperiods, rather than the philosophical issues. Consequently, Dekker understandably puts aside the latter in his discussion. Focusing on them could mislead us about Tinbergen's view of the controversy. As Tinbergen saw it, the controversy primarily consisted in technical debates arising from Keynes' misinterpretations and his ignorance of the relevant econometric literature (Magnus and Morgan 1987, 129–130).

Another reason that could be given for Dekker's focus is that many scholars have already explored the philosophical aspects of the Tinbergen-Keynes debates. For example, Anna Carabelli emphasises the importance of Keynes' *A Treatise on Probability* (1921) for understanding the sources of Keynes' criticisms (Carabelli 1988, chap. 10). She argues that Keynes' ontological assumptions made him sceptical that many statistical methods were applicable in economic contexts. The phenomena of economics are characterised by instabilities arising from human nature, which lacks the stability of natural materials that have deterministic features (like mechanical systems) or stable long-run probabilities (like radioactive half-lives). In contrast, Tinbergen had a physics background, and tended to assume—as a defeasible conjecture—that the statistical methods which had been so successful in physics would also be applicable in economics (Dekker 2021, sec. 4.3).¹ Similarly, Mary Morgan has explored the distinct roles that Keynes and Tinbergen saw for statistical inference in economics: Keynes granted that one could statistically estimate very local facts, such as an economy's price level or its fiscal multiplier over a brief period. Tinbergen advocated a further theoretical role of refuting macroeconomic theories or even indicating ways of synthesising them (Morgan 1990, 124–125). To summarise the main differences, Keynes and Tinbergen differed on:

¹ See Lawson (2003) for a concise discussion of Keynes' ontology and its role in both *A Treatise on Probability* and his explicitly economic works.

- (1) The applicability of methods from physics to economics;
- (2) the role of statistical inference in economics;
- (3) the adequacy of Tinbergen's chosen statistical methods, even for the modest roles envisaged by Keynes; and
- (4) whether the techniques for correlation analysis methods (a key part of Tinbergen's study) in economics were improving. In particular, Keynes was sceptical that these methods had fundamentally improved since the period he was actively researching the topic, which was the long development period of *A Treatise in Probability*, prior to 1921 (Dekker 2021, 186–187).

In this book symposium contribution, I shall seek to supplement the insights provided by Dekker by discussing some underexplored philosophical aspects of the Tinbergen-Keynes debates. Like much of the literature, Dekker focuses on the differences between Tinbergen and Keynes. I do not dispute that there were differences, but I shall discuss some important similarities. I shall also connect these debates to contemporary methodological issues. Additionally, the literature on the debates has tended to be pessimistic about (or uninterested in) the possibilities of integrating insights from Keynes into econometrics, while still retaining Tinbergen's ambitions for the subject.

In the first two sections, I discuss their shared views on the objectivity and nature of statistical evidence. I shall then explain how recent developments in statistical methods and the philosophy of science may help econometricians to address one of Keynes' fundamental criticisms—the problem of amalgamating evidence in macroeconometrics. Thus, it may be able to synthesise (at least partly) Tinbergen's ambitions with Keynes' cautions. Overall, my discussion will illustrate how, by examining the similarities as well as the insights of both economists, we can understand their individual perspectives better and their relation to modern debates.

II. OBJECTIVITY

A common view in contemporary philosophy of statistics is Subjective Bayesianism, according to which statistical reasoning is based on arbitrarily chosen probabilities (Ramsey 1990; Howson and Urbach 2006). In this section, I shall explain how Tinbergen and Keynes both differed from this subjectivist position regarding statistical research.

Tinbergen understands statistical practice as performing tests that will estimate relative frequencies (the rate at which some event, property, characteristic, event, and so on occurs in a class of things) with a degree of reliability we judge to be sufficiently high for some practical or

scientific problem. Frequentists quantify ‘reliability’ in terms of the long-run frequencies of error in a statistical inference.² Economic theories are tested through using them to guide the development of empirical models. For example, if a business cycle theory T implies that interest rates have an important effect on investment under certain conditions, but statistical evidence in estimating a model suggests that the effect is small, then that is evidence against T (Tinbergen 1940, 142–143). Thus, Tinbergen’s approach is an early example of frequentist econometrics. His revolutionary contribution was to demonstrate the possibility of testing *dynamic* macroeconomic (macrodynamic) models. These models are about macroeconomic changes over time, like business cycles. Moreover, he pioneered the approach of using a wide variety of tests (Morgan 1990, chap. 4).

In frequentism, the main subjective element is the choice of acceptable error rates—the long-run frequency of errors in a test. For example, in significance testing, given a particular model, rejection threshold (‘significance level’), and dataset, the evidential significance of a test is determined by objective factors, such as (a) the likelihood of the data given the model and (b) the methods used to sample the data. By ‘objective’, I mean that it is not determined by scientists’ arbitrary opinions.³

Which error rates should we choose? This decision is not determined by frequentist methodology, but it is also not arbitrary. For instance, confidence interval testing is incoherent if we set our standards too low. In particular, contradictory estimates can all exceed a confidence level that is set too low. A confidence *level* is $\gamma = (1 - \alpha)$, where α is the test conditions’ maximum long-term rate of random error for inferring a particular type of confidence *interval*.⁴ The confidence interval is an approximate

² These errors can be random or systematic. Random error is error due to chance, such as selecting an unrepresentative sample. There is random error even in an unbiased sampling process. It is distinct from systematic error, which is due to a problem with the testing set-up, such as a biased selection process. If we are observing bats in a forest and estimating their relative frequency per acre, we can mitigate the risks of random error with larger samples, but if we choose to make our observations during the day, then we will still make systematic errors, because our sampling process is biased against seeing bats.

³ See Mayo (2018) for a contemporary and sophisticated methodology of frequentist statistics.

⁴ In more detail, the maximum long-run rate of error for inferring a type of confidence interval is the approximate relative frequency at which we would make errors if we used such samples for inferences. If we made the inferences enough times, the proportion of our inferences that are mistaken would settle around this frequency α . When it settles, we have reached the long-run. Some error might be unavoidable (it depends on the facts about what we are studying) but we can mathematically determine a maximum level that the long-run error rate cannot exceed.

estimate of a parameter θ . If we set γ at 50% or lower, then it is possible that both a statistical hypothesis (such as ' θ is in the interval 0.8 to 0.9') and its negation (' θ is not in the interval 0.8 to 0.9') can have maximum error rates above γ . To use an extreme case, if $\gamma = 0$, then $\alpha = 1$, meaning that we are willing to make estimates with a long-term error rate of 100%. Thus, by this 'standard', we could infer both statistical hypotheses and any statistical hypothesis that disagrees with them, and our inferences would still meet our 'standards' for avoiding long-run errors. To avoid incoherence, statisticians must set γ high enough to avoid the possibility of incompatible hypotheses exceeding γ . Typical conventions for α include 0.05 and 0.01, but any value below 0.5 will avoid incoherence, because the maximum probability of an error in inferring a type of hypothesis H_1 is $1 - P_E(H_d)$, where $P_E(H_d)$ is the maximum possible error rate for inferring the disjunction of alternatives to H_1 . Thus, if $\alpha > 0.5$, then $P_E(H_d) < 0.5$. The relevant error for any alternative to H_1 will be less than $P_E(H_d)$. Hence, $\alpha > 0.5$ makes it impossible to estimate both a statistical hypothesis and a contrary hypothesis when using confidence interval estimation.

Keynes was far more active in complex philosophical controversies than Tinbergen, which makes Keynes' epistemology more comprehensive and yet also harder to interpret. Keynes developed a detailed theory of probability and statistics in *A Treatise on Probability* (1921). However, the extent to which his ideas had changed by 1939 is a matter of considerable controversy. Curiously, the Tinbergen-Keynes debates are rarely used as exegetical evidence in this controversy.⁵

Keynes (1921) interprets probability as an objective evidential relation between propositions. Thus, just as 'All men are mortal and Socrates is a man' has an objective relation of deductive implication towards 'Socrates is mortal', so '95-100% of men are mortal and Socrates is a man' has an objective evidential relation of partial support in favour of 'Socrates is mortal'. This partial support is not deductive, because the premise could be true and the conclusion false, but Keynes argued that it has a weaker, non-deductive form of logical implication. Keynes used this interpretation of probability to formally analyse a wide range of non-deductive types of reasoning, like induction and analogy. In the latter part of Keynes (1921), he also applied it to statistical methodology. Thus, in 1921, Keynes believed that evidential relations in statistics are objective.

⁵ There are some exceptions, such as Rowley and Hamouda (1987), Brady (1988), and Marchionatti (2010).

In 1922–1926, Keynes was criticised by his close friend, Frank P. Ramsey (1990). The exegetical controversies concern the extent to which Keynes altered his views in response to Ramsey. These controversies involve the interpretation of several difficult texts, including Keynes' *The General Theory of Unemployment, Interest and Money* (1936) and his obituary for Ramsey (Keynes 1972, 335–339). Very roughly, there are two popular positions in this dispute, which I shall call (a) the Continuity interpretation and (b) the Recantation interpretation.⁶

According to the Continuity interpretation, Keynes' views about probability and statistics were more or less unchanged by Ramsey's critique (O'Donnell 1989; Runde 2003, 52; Brady 2017; Davis 2019, 88). According to the Recantation interpretation, Ramsey convinced Keynes to modify his views fundamentally, including abandoning objectivity (Bateman 1987; Gillies 2006; Raffaelli 2006). To see how good scholars can disagree about on this topic, even though Keynes is rarely an obscure writer, consider this passage from his obituary for Ramsey:

Ramsey argues [against me] that probability is concerned not with objective relations between propositions but (in some sense) with degrees of belief, and he succeeds in showing that the calculus of probabilities simply amounts to a set of rules for ensuring that the system of degrees of belief which we hold shall be a *consistent* system. Thus the calculus of probabilities belongs to formal logic. But the basis of our degrees of belief—or the *a priori* probabilities [...]—is part of our human outfit, perhaps given to us merely by natural selection, analogous to our perceptions and our memories rather than to formal logic. So far as I yield to Ramsey—I think he is right. (Keynes 1972, 338–339)

The problem is that the scope of 'so far' is unclear (Fitzgibbons 1998, 163). It certainly involves the previous sentence: Keynes agrees with Ramsey that we learn prior probabilities from our innate nature, rather than from an intuitive grasp of platonic conceptual relations. How much more does it include? Does it include Ramsey's denial of probability's objectivity? Furthermore, arguably much of the paragraph was consistent with Keynes' earlier views anyway, such as the role of the probability calculus.⁷

The Tinbergen-Keynes debates help to resolve this controversy. Tinbergen claimed that econometric tests can prove a theory to be incorrect.

⁶ There is also the Subjectivist interpretation by Anna Carabelli (1988, 2021) and Donald Moggridge (1992, 623) who argue that Keynes was *always* similar to Ramsey and therefore had little, if anything, to concede.

⁷ Rules for inferring probabilities from other probabilities.

Keynes argued that, outside of special circumstances, this is wrong. However, he does not appeal to the subjectivity of evidential relations (Keynes 1939, 559–560). Such an appeal would have made Keynes' task easier. If we are free to adopt almost any probability distribution as our initial degrees of belief, then talk of 'proof' in statistics is completely out of place, even in an extended non-deductive sense. For example, there are probability distributions where a high proportion of Heads in a sample of coin tosses is evidence in favour of its bias towards Heads, but also probability distributions where it is irrelevant, and even some in which such evidence confirms that the coin is biased *against* Heads! On Ramsey's view, all of these probability distributions are equally permissible. Keynes does not even mention this shortcut, indicating that he still regarded evidential relations in statistics as objective. In general, throughout Keynes' criticism of Tinbergen, he apparently adopts an objective conception of evidence—it is the relation of a hypothesis to our evidence and background knowledge that determines a test's evidential force (or lack thereof).

Of course, Keynes may have been trying to rely on premises that his contemporary economists believed—Subjective Bayesianism was still very rare in 1939. However, Keynes enjoyed controversy and had no fear of advocating unpopular positions. The Tinbergen-Keynes debates were a perfect opportunity for him to explore the implications of a Ramseyan view, if he had adopted it. Thus, the texts in the Tinbergen-Keynes debates support the Continuity interpretation, though they do not end the debate. Overall, it seems that although Tinbergen and Keynes had different theories of statistical evidence (frequentist and non-frequentist) both agreed on the objectivity of evidential relations in econometrics.

It is worth emphasising that 'objectivity' in this sense still leaves a wide scope for rational disagreement in economic debates. That two people should agree on the evidential relations between a theory T and statistical evidence E does not imply that they will agree on whether we should believe T , believe that T is false, or be agnostic about T , even if they both believe E . As Keynes details in *A Treatise on Probability*, scientific evidence is a three-place relation between theories, evidence, and background information (assumptions, knowledge, and so on). For instance, disagreements about the econometric assumptions that economists make in a particular context can block agreements on the pertinence of some statistical evidence for a macroeconomic theory, even among rational economists. Thus, this sense of 'objectivity' does not have the absurd implications that controversies in economics are easily

resolved, nor that economic reasoning has no role for intuition and subjective judgement.

III. THE PROBLEM OF INDUCTION

Some commentators have linked Keynes' views on probability and statistics, including in the Tinbergen-Keynes debates, to the Problem of Induction (POI). I shall argue that this is misleading, given the traditional understanding of the POI, and that the usefulness of inductive reasoning in economics is actually a salient area of common ground for Tinbergen and Keynes. First, I shall explain the POI. Second, I shall argue that Keynes did not view it as a problem for economics. Third, I shall argue that Tinbergen, like Keynes, adopted a sophisticated form of inductivism about economics. Thus, on a major issue of dispute in the methodology of economics,⁸ Tinbergen and Keynes were on the same side.

Dating the origins of the POI is beyond this article's scope, but it was most influentially presented by David Hume ([1740, 1748] 1975). There are many interpretations of it, but the fundamental core is clear. Hume contends that inductive arguments are not deductively valid. Inductive arguments have (a) premises known directly or indirectly by observation and (b) conclusions about unobserved objects or properties. Given that (now) uncontroversial logical fact, it is not clear how they are otherwise justified. In other words, how the premises can support the conclusions? On the traditional interpretation of Hume, the POI is a sceptical paradox: we have an instinct to trust induction, but Hume thought that we cannot justify this instinct (Russell 1912, chap. 6; Stove 1973; Stroud 1977, chaps. III-IV; Stove 1982, chap. IV).

Sheila Dow interprets Keynes as raising the POI to criticise Tinbergen (Dow 2004, 552). She interprets Keynes as an inductive sceptic, in the sense that he regarded inductive reasoning as a convention (Dow 2009, 2010; see also Andrews 1999). There are weak versions of the POI where Dow is certainly correct. If one interprets Hume as just saying that induction is not deductively valid, and thus involves some special risks, then Keynes agreed with Hume. However, that position is not sceptical enough to be an objection to any but the most overconfident econometrician.

On stronger interpretations of the POI, Keynes' position against Tinbergen actually presupposes that the POI is *not* an issue for economics. Dow identifies a passage by Keynes as echoing themes from Hume:

⁸ For economists with opposing views, see, for example, Peter Boettke (1997), Murray Rothbard (2004), and perhaps Ludwig von Mises (1949; but see Linsbichler 2017).

The most important condition is that the environment in all relevant respects, other than the fluctuations in those factors of which we take particular account, should be uniform and homogeneous over a period of time. We cannot be sure that such conditions will persist in the future, even if we find them in the past. [...] [T]he main prima facie objection to the application of the method of multiple correlation to complex economic problems lies in the apparent lack of any adequate degree of uniformity in the environment. (Keynes 1939, 566–567)

Granted, this is *a* problem involving induction, but it is not Hume's POI. Hume raised a problem of knowing about the wider 'environment' beyond what we have observed, whereas Keynes is claiming that our observations of instability in economic phenomena are representative of that environment: we *know* that economic phenomena are such that Tinbergen's methods are unreliable beyond ephemeral time spans and particular economies.⁹ Consequently, Keynes' criticisms of Tinbergen suggest that he had a thoroughly inductivist view: induction can and does contribute to economic knowledge.

What about Tinbergen? Some of his remarks may suggest an anti-induction position. Tinbergen says that econometric testing cannot prove a theory to be true, but it can prove it to be false (Tinbergen 1939a, 12). Taken in isolation, this remark corresponds with Karl Popper's position about statistical inference (Popper [1980] 2002). Popper's approach to statistics was based on his belief that the POI is unsolvable. Against Keynes and others, Popper argued that science's logic is purely deductive. A key step in Popper's research programme to establish this deduction-only view of scientific evidence was to argue that truly *scientific* theories are falsifiable, in the sense that they could be logically inconsistent with some data. Statistics present a challenge to Popper's methodology, because no statistical estimate is ever deductively inconsistent with our data: even $T_1 =$ 'At least 99% of human births are male' is consistent with what we know, because we might just have observed extremely atypical subsequence thus far. Popper proposed that we set conventional rules to

⁹ See also Carabelli (1988, 185ff). Keynes tends to assume that heterogeneity and instability go together. If heterogeneity is something like the stochastic relevance of many conditions to a random variable's value and stability is something like a high standard deviation for that random variable, then these are mathematically distinct. However, Keynes might be right that, as an empirical fact, these are correlated in economics. The instability of economic—as opposed to many natural—phenomena was a crucial part of Keynes' methodology. It was why he regarded economics as a 'moral science' rather than as analogous to the natural sciences, especially physics (Keynes 1978, 296–300).

“regard” data that is extremely improbable given a theory as if it falsifies that theory (Popper [1980] 2002, 194).

However, Tinbergen’s views are not Popperian. First, it is important to note that there is no logical asymmetry between verification and falsification in a statistical context.¹⁰ Both rejecting T_1 and accepting a hypothesis like $T_2 =$ ‘Less than 99% of male births are male’ require non-deductive reasoning, because our evidence neither deductively entails nor refutes either hypothesis. Second, Tinbergen (1939a, 12) says that there are senses of “verification” that econometrics can achieve, though he does not specify them; he apparently means what contemporary philosophers call ‘confirmation’, which is evidential support such that the evidence may prove the hypothesis or (more typically) provides weaker evidence in favour of it.

Given this logical symmetry between verification and falsification, what rationale could Tinbergen have for distinguishing them? Dekker points out an intriguing subtlety in Tinbergen’s prospectus for econometrics. Tinbergen was pessimistic about econometric models’ capacities to make accurate predictions. Nonetheless, he was optimistic about the possibility of synthesising theories of the business cycle, by using econometrics to test the adequacy of rival theories’ claims, and then combining the insights of the remaining ideas (Dekker 2021, 152–153, 178). Thus, his ambitions for econometrics required the exclusion of false predicted relationships. Given such a refinement of the rivals, theoreticians could construct a synthesis from a narrower, more robust set of theories. Therefore, econometrics had a negative role in Tinbergen’s macrodynamic programme.¹¹ In contrast, Tinbergen did not think that econometrics alone can warrant the (provisional) acceptance of a business cycle theory for either practical or scientific purposes. Other theoretical reasoning is necessary. A theory must survive the filter of testing, but the overall evaluation of its plausibility is beyond econometrics.

Where did Tinbergen acquire these fairly sophisticated views on induction, falsification, and statistical inference? It is hard to say, because he tries to avoid philosophical discussions wherever possible. However,

¹⁰ In this sentence, by ‘logical’, I mean deductive logic. Standard statistical norms are sometimes called the ‘logic’ of statistics, but they are separate from what I mean. Many statisticians sharply distinguish verification and refutation, saying that statistically significant result is evidence against the null hypothesis H_0 , but failure to reject is not evidence for H_0 , and similar claims (Fisher 1974, 16). These norms are open to dispute, but they are compatible with what I say in this article. I thank an anonymous referee for raising this point.

¹¹ Aside from estimating parameters.

placing him in historical context provides some indications. Tinbergen first encountered scientific methods (at a tertiary education level) studying physics at the undergraduate and postgraduate levels (his PhD combined physics and economics) at Leiden University from 1921 to 1929. The 1920s were an exciting and uncertain period in the history of physics. The Newtonian theories were no longer tenable. The exemplar of an empirically successful and theoretically rigorous scientific theory had fallen just as quickly and dramatically as the great land empires of Germany, Russia, Turkey, and Austro-Hungary. Yet, despite some stunning successes, the revolutionary theories of Einstein and others were far from well-established, not least because the fall of Newtonianism had dramatically demonstrated the fallibility of science. In this period, the fallibility of induction was prominent in the minds of educated people. David Stove has noted this effect in the philosophy of science (Stove 1973, chap. 7) yet it also created a new sense of modesty in physics. Despite the comparative stability of physics since the fall of Newtonianism, this modesty survives to some extent even today. In this atmosphere, even a physicist as averse to philosophy as Tinbergen could easily pick up a relatively subtle methodology of induction: it is fallible, but it is also capable of providing some evidential support.¹² Since Tinbergen tends to assume that the methodology of physics is also applicable to economics, it is plausible that his practice in economics inherited this approach to inductive inference.

Therefore, while positive evidence and negative evidence are logically the same in statistics, it was negative evidence that was important for Tinbergen's aims. His focus on falsification is pragmatic, rather than Popper's epistemological motivation from the POI. Overall, inductivism in economics is common ground for Keynes and Tinbergen.

IV. THE DEBATES AND EVIDENCE SYNTHESIS: AN OPTIMISTIC NOTE

One of the deepest methodological issues that Keynes raised is now known as evidence synthesis—how to amalgamate information from multiple methods (Keynes 1939, 561). Tinbergen argued that, although there

¹² It is representative of Keynes' philosophical sophistication that he developed a modest theory of induction *prior* to the fall of Newtonianism: he developed the key ideas of *A Treatise on Probability* in the 1900s and early 1910s, before the main revolutions in physics. It is striking to contrast Keynes' stress on induction's fallibility against almost all of his prominent predecessors—John Stuart Mill, William Whewell, Immanuel Kant, and other less influential figures in 19th century philosophy of science. It is even more striking given that Hume's ideas about induction were not prominent, at least outside Cambridge (Russell 1912, chap. 6), when Keynes began his work on induction.

may be significant non-measurable economic phenomena that cannot be directly handled by his testing methods, the effects of these phenomena can be taken into account by using our background knowledge about them (Tinbergen 1939a, 11) and he gave particular examples of how he had incorporated events like major strikes (Tinbergen 1940, 143). Tinbergen thought that, in some contexts, this incorporation would improve the analyses' accuracy. However, Keynes noted that Tinbergen's method for the identification of these exogenous factors was through the size of the residual: the amount of variation that the model did not explain in a correlation analysis. Yet the size of the residual was *also* Tinbergen's criterion for evaluating the analysis's accuracy. So should we (a) discount models that take our background knowledge into account because of their large residuals or (b) take them more seriously due to their greater expected accuracy? Tinbergen's methods become ad hoc and perhaps incoherent at this point (Keynes 1940, 155).

The underlying issue is that Tinbergen lacked a systematic and plausible methodology for evidence synthesis. This is not a problem if there is just one type of evidence. For example, if we can estimate a parameter via confidence intervals *and only confidence intervals*, from disjoint samples, then it is straightforward to combine the evidence into a single estimate.

Unfortunately, when the evidence is heterogenous, then this simple picture falls apart. As Keynes noted, this problem occurs in economics. It also occurs in other sciences. For example, in pharmacology, evidence about the risks and safety of a drug often comes from many sources: randomized controlled trials (RCTs) but also surveys, case series, cohort studies, and so on (De Pretis, Landes, and Osimani 2019). Combining evidence from across these different types is a considerable challenge.

However, recent developments in Bayesian methodology have made progress in evidence synthesis for many sciences, including medicine (Sutton and Abrams 2001; Sweeting et al. 2008; Walach and Loeff 2015; Landes, Osimani, and Poellinger 2018; De Pretis, Landes, and Osimani 2019; De Pretis, Landes, and Peden 2021) psychology (Scheibehenn, Jamil, and Wagenmakers 2016; Heck et al. 2022) and economics (Ades et al. 2006; Baio 2012; Jackson et al. 2019). Using the tools of Bayesian statistics, it seems possible to combine evidence in a systematic way, at least in many contexts. Therefore, it may now become possible for econometricians to respond to Keynes' critique far more systematically than Tinbergen could hope in 1940.

Unfortunately, Bayesian statistics presupposes prior probabilities—probabilities before learning one’s evidence. According to Bayesianism, hypotheses can have probabilities given some evidence. These probabilities are determined by a combination of (a) the probability of the evidence given the hypothesis, (b) the prior probability of the hypothesis, and (c) the prior probability of the evidence. While (a) is often determined mathematically, (b) and (c) are not. So where do they come from? The ‘Problem of the Priors’ has been a persistent and fundamental criticism of Bayesianism (Venn 1876; Nagel 1939; Glymour 1980; Mayo 2018). In brief, worry is that the choice of priors may be subjective and therefore unscientific. In contrast, the methodology used by Tinbergen was frequentism; it is fundamentally based upon the objective error rates of tests, not prior probabilities of hypotheses (see section II).

Both frequentism and Bayesianism have their strengths: the former in avoiding priors, the latter in its tools for evidence synthesis. It is hence unsurprising that philosophers and statisticians have been interested in combining these methodologies. For instance, some contemporary philosophers (Kyburg and Teng 2001; Kyburg 2006; Williamson 2010, 2013) have proposed an intriguing compromise, which I shall call Objective Bayesianism.¹³ According to this approach, Bayesian reasoning takes precedence over frequentist reasoning, but only when the ‘priors’ involved are derivable from background knowledge about the phenomena. These ‘priors’ will be imprecise approximations of the true frequencies, derived from sources like confidence interval estimates, maximum likelihood estimation, or well-supported stochastic theories. Hence, when our background knowledge about phenomena is weak, then frequentist testing will tend to be suitable. We lack evidence-based priors in such situations, but we can fall back on the purely mathematical facts about error rates that frequentists utilise. However, when we can have Bayesian priors based on evidence rather than subjective opinion, then we can use Bayesian reasoning—in the sense of using these priors for techniques like Bayesian evidence synthesis methods.

Thus, it may be possible to systematically address Keynes’ point *if* our background knowledge is sufficiently rich. When interpreted as a constructive challenge, rather than a dogmatic rejection of econometric testing as such, then Keynes’ criticisms offer useful guidance for the further development of econometric methodology. Of course, Objective

¹³ This label is not quite right for Kyburg and Teng, for reasons that are unimportant here.

Bayesianism has its critics (for instance, Howson and Urbach 2006). Additionally, it has not yet been applied to econometrics. As Keynes noted, there are many epistemological and substantive conditions that must be satisfied before probabilistic methods are applicable to a problem. For example, we might know that we are making assumptions about the homogeneity of our subject matter that we know are wrong, such as that the variables involved are not autocorrelated.¹⁴ The persistent disagreements in economics, especially macroeconomics, do not suggest an optimistic view. Nonetheless, it is worth exploring.

V. CONCLUSION

Keynes must have been surprised and confused by Tinbergen: a young economist who claimed to be testing macroeconomic theories in 1939. Even more puzzling must have been the League's apparent endorsement of Tinbergen's research. Keynes would have been justifiably worried that, with World War II beginning, he would be unable to participate for long in the debates about this new method. It would have been tempting to be angry with Tinbergen and attack him superficially. Yet Keynes admired him as an economist and a person (Patinkin 1982, 229-230). He limited himself to only occasional rhetorical flourishes, like describing Tinbergen's work as 'alchemy'. Furthermore, as we have seen, Keynes' criticisms were far from superficial. The Tinbergen-Keynes debates remain both historically and philosophically important. While the macrodynamic programme in econometrics that Tinbergen initiated has made tremendous progress, few are sanguine about it. The philosophical issues raised by Keynes remain controversial. Furthermore, the foundations of econometrics remain obscure. What are the relevant 'probabilities' in econometrics? Degrees of belief, error rates, or something else? How should econometrics be synthesised with other types of evidence in macroeconomics? As I have shown in this commentary, the Tinbergen-Keynes debates remain a thought-provoking means to approach these issues.

Additionally, the reactions to Tinbergen's work by economists other than Keynes have been largely neglected by philosophers of economics. Dekker describes several (chap. 8) and one could add the reactions of Milton Friedman (1940) and Trygve Haavelmo (1943), which are underexplored in the philosophy of economics. Dekker's book is an excellent

¹⁴ A variable like national income or inflation is autocorrelated if its value at t is correlated with its value at $t + n$, for some period n , which could be three months, a year, two years, and so on.

introduction to these controversies, because it gives us a comprehensive impression of Tinbergen's own underlying ambitions and the historical context of his efforts.

Dekker portrays Tinbergen as a passionate man of action, with technocratic tendencies, who sought to use his expert status to achieve ambitious and idealistic goals: achieving a stable and peaceful international order that would end or at least greatly mitigate the suffering caused by business cycle fluctuations, substantially alleviate world poverty, and combine rival economic schools into consensus syntheses. In accordance with this practical orientation, Dekker notes how Tinbergen was highly averse to taking stands on philosophical questions (415). Nonetheless, I have explained how Tinbergen's views in his clash with Keynes do have important philosophical aspects. I have also argued that their agreements are philosophically interesting, as well as their disagreements. We can also use these debates as gateways into the philosophical views of these economists. Furthermore, the methodological issues raised in the Tinbergen-Keynes debates continue to be important for the philosophy of science, especially the philosophy of statistics. Developing methods that can achieve Tinbergen's aims, while respecting the insights of Keynes' critiques, remains an obdurate but exciting challenge for economists and statisticians.

The literature on the philosophical aspects of the Tinbergen-Keynes debates has understandably focused on their differences, since these are what is important for interpreting the cause of Keynes' criticisms and the details of Tinbergen's responses. However, as I have argued in this article, there are also important similarities. For example, it is possible that the similarities were one reason why Tinbergen did not really engage with the epistemological (rather than technical) aspects in Keynes' attack, in addition to Tinbergen's aversion to philosophical controversy (sections II and III above). Additionally, when we see that Keynes attitude to Tinbergian econometrics can be rationally reconstructed as principled methodological caution, rather than as a purely ideological or a fundamental opposition, then we can see prospects for econometrics that combines the insights of Keynes with the aspirations of Tinbergen (section IV). Finally, by supplementing Dekker's discussion with these philosophical points and references to a few key texts in the literature (like Carabelli 1988; Brady 1988; O'Donnell 1989; Morgan 1990) and others, I hope that I have helped readers gain a deeper understanding of the Tinbergen-Keynes debates.

REFERENCES

- Ades, Anthony E., Mark Sculpher, Alex Sutton, Keith Abrams, Nicola Cooper, Nicky Welton, and Guobing Lu. 2006. "Bayesian Methods for Evidence Synthesis in Cost-Effectiveness Analysis." *Pharmacoeconomics* 24 (1): 1-19.
- Andrews, David R. 1999. "Continuity and Change in Keynes's Thought: The Importance of Hume." *European Journal of the History of Economic Thought* 6 (1): 1-21.
- Baio, Gianluca. 2012. *Bayesian Methods in Health Economics*. Boca Raton, FL, London, and New York, NY: CRC Press.
- Bateman, Bradley W. 1987. "Keynes's Changing Conception of Probability." *Economics and Philosophy* 3 (1): 97-119.
- Boettke, Peter J. 1997. "Where Did Economics Go Wrong? Modern Economics as a Flight from Reality." *Critical Review* 11 (1): 11-64.
- Brady, Michael Emmett. 1988. "J. M. Keynes's Position on the General Applicability of Mathematical, Logical and Statistical Methods in Economics and Social Science." *Synthese*, 76 (1): 1-24.
- Brady, Michael Emmett. 2017. "On J M Keynes's Rejection, in General, of Ramsey's Subjective Theory of Probability: The Keynes-Townshend Exchanges of 1937 and 1938." SSRN Working Paper No. 2898643. SSRN, Rochester, NY.
- Carabelli, Anna. 1988. *On Keynes's Method*. London: Palgrave Macmillan.
- Carabelli, Anna. 2021. *Keynes on Uncertainty and Tragic Happiness*. Cham: Springer.
- Davis, John. B. 2019. "Frank P. Ramsey." In *The Elgar Companion to Keynes*, edited by Robert W. Dimand and Harald Hagemann, 87-89. Cheltenham: Edward Elgar.
- De Pretis, Francesco, Jürgen Landes, and Barbara Osimani. 2019. "E-Synthesis: A Bayesian Framework for Causal Assessment in Pharmacovigilance." *Frontiers in Pharmacology* 1-20.
- De Pretis, Francesco, Jürgen Landes, and William Peden. 2021. "Artificial Intelligence Methods for a Bayesian Epistemology-Powered Evidence Evaluation." *Journal of Evaluation in Clinical Practice* 27 (3): 504-512.
- Dekker, Erwin. 2021. *Jan Tinbergen (1903-1994) and the Rise of Economic Expertise*. Cambridge: Cambridge University Press.
- Dow, Sheila C. 2004. "Uncertainty and Monetary Policy." *Oxford Economic Papers* 56 (3): 539-561.
- Dow, Sheila C. 2009. "David Hume and Modern Economics." *Capitalism and Society* 4 (1): 1-29.
- Dow, Sheila C. 2010. "Was There a (Methodological) Keynesian Revolution?" In *Keynes's General Theory After Seventy Years*, edited by R.W. Dimand, R.A. Mundell, and A. Vercelli, 268-286. London: Palgrave Macmillan.
- Fisher, Ronald A. 1974. *The Design of Experiments*. London: Hafner Publishing Company.
- Fitzgibbons, Athol. 1998. "Against Keynes's Recantation." *Cahiers d'économie politique / Papers in Political Economy* 30/31: 539-561.
- Friedman, Milton. 1940. "Review of 'Business Cycles in the United States of America, 1919-1932' by J. Tinbergen." *American Economic Review*, 30 (3): 657-660.
- Gillies, Donald. 2006. "Keynes and Probability." In *The Cambridge Companion to Keynes*, edited by Roger E. Backhouse and Bradley W. Bateman, 199-216. Cambridge: Cambridge University Press.
- Glymour, Clark. 1980. *Theory and Evidence*. Princeton, NJ: Princeton University Press.

- Haavelmo, Trygve. 1943. "Statistical Testing of Business-Cycle Theories." *The Review of Economics and Statistics* 25 (1): 13–18.
- Heck, Daniel W., Udo Boehm, Florian Böing-Messing, Paul-Christian Bürkner, Koen Derks, Zoltan Dienes, Qianrao Fu et al. 2022. "A Review of Applications of the Bayes Factor in Psychological Research." *Psychological Methods* 1–65.
- Howson, Colin, and Peter Urbach. 2006. *Scientific Reasoning: The Bayesian Approach*. Chicago and La Salle, IL: Open Court Publishing.
- Hume, David. (1740) 1975. *A Treatise of Human Nature*. Edited by L. A. Selby-Bigge and P. H. Nidditch. Reprint, Oxford: Oxford University Press.
- Hume, David. (1748) 1975. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Edited by L. A. Selby-Bigge and P. H. Nidditch. Reprint, Oxford: Oxford University Press.
- Jackson, Christopher, Anne Presanis, Stefano Conti, and Daniela De Angelis. 2019. "Value of Information: Sensitivity Analysis and Research Design in Bayesian Evidence Synthesis." *Journal of the American Statistical Association* 114 (528): 1436–1449.
- Keynes, John. M. 1921. *A Treatise on Probability*. London: Macmillan & Co.
- Keynes, John. M. 1936. *The General Theory of Unemployment, Interest and Money*. London: Palgrave Macmillan.
- Keynes, John. M. 1939. "Professor Tinbergen's Method." *The Economic Journal* 49 (195): 558–577.
- Keynes, John. M. 1940. "On a Method of Statistical Business-Cycle Research. A Comment." *The Economic Journal* 50 (197): 154–156.
- Keynes, John. M. 1972. *The Collected Writings of John Maynard Keynes Vol. X, Essays in Biography*. Edited by Elizabeth Johnson and Donald Moggridge. London: Macmillan/Royal Economic Society.
- Keynes, John. M. 1978. *The Collected Writings of John Maynard Keynes Vol. XIV, The General Theory and After: Part II. Defence and Development*. Edited by Elizabeth Johnson and Donald Moggridge. London: Macmillan/Royal Economic Society.
- Kyburg, Henry E. 2006. "Belief, Evidence, and Conditioning." *Philosophy of Science* 73 (1): 42–65.
- Kyburg, Henry E., and Choh Man Teng. 2001. *Uncertain Inference*. Cambridge: Cambridge University Press.
- Landes, Jürgen, Barbara Osimani, and Roland Poellinger. 2018. "Epistemology of Causal Inference in Pharmacology: Towards a Framework for the Assessment of Harms." *European Journal for Philosophy of Science* 8 (1): 3–49.
- Lawson, Tony. 2003. "Keynes's Realist Orientation." In *The Philosophy of Keynes's Economics: Probability, Uncertainty and Convention*, edited by Jochen Runde and Sohei Mizuhara, pp. 154–164. London: Routledge.
- Linsbichler, Alexander. 2017. *Was Ludwig Von Mises a Conventionalist? A New Analysis of the Epistemology of the Austrian School of Economics*. London: Palgrave Macmillan.
- Magnus, Jan R., and Mary S. Morgan. 1987. "The ET interview: Professor J. Tinbergen." *Econometric Theory* 3 (1): 117–142.
- Marchionatti, Roberto. 2010. "J. M. Keynes, Thinker of Economic Complexity." *History of Economic Ideas* 8 (1): 115–146.
- Mayo, Deborah. 2018. *Statistical Evidence as Severe Testing*. Cambridge: Cambridge University Press.

- Mises, Ludwig von. 1949. *Human Action: A Treatise on Economics*. New Haven, CT: Yale University Press.
- Moggridge, Donald E. 1992. *Maynard Keynes: An Economist's Biography*. London: Routledge.
- Morgan, Mary S. 1990. *The History of Econometric Ideas*. Cambridge: Cambridge University Press.
- Nagel, Ernest. 1939. *The Principles of Probability*. Chicago, IL: University of Chicago Press.
- O'Donnell, Rodney M. 1989. *Keynes: Philosophy, Economics and Politics*. New York, NY: Palgrave Macmillan.
- Patinkin, Donald. 1982. *Anticipations of the General Theory? And Other Essays on Keynes*. Chicago, IL: Chicago University Press.
- Popper, Karl. (1980) 2002. *The Logic of Scientific Discovery*. Reprint, London and New York, NY: Routledge.
- Raffaelli, Tiziano. 2006. "Keynes and Philosophers." In *The Cambridge Companion to Keynes*, edited by Roger E. Backhouse and Bradley W. Bateman, 157–176. Cambridge: Cambridge University Press.
- Ramsey, Frank P. 1990. *F.P. Ramsey: Philosophical Papers*. Edited by David H. Mellor. Cambridge: Cambridge University Press.
- Rothbard, Murray N. 2004. *Man, Economy & State with Power and Market*, Auburn, AL: Ludwig von Mises Institute.
- Rowley, Robin, and Omar Hamouda. 1987. "Troublesome Probability and Economics." *Journal of Post Keynesian Economics* 10 (1): 44–64.
- Runde, Jochen. 2003. "On Some Explicit Links between Keynes's A Treatise on Probability and The General Theory. In *The Philosophy of Keynes's Economics: Probability, Uncertainty and Convention*, edited by Jochen Runde and Sohei Mizuhara, 45–53. London, Routledge.
- Russell, Bertrand. 1912. *The Problems of Philosophy*. London: Williams and Norgate.
- Scheibehenne, Benjamin, Tahira Jamil, and Eric-Jan Wagenmakers. 2016. "Bayesian Evidence Synthesis Can Reconcile Seemingly Inconsistent Results: The Case of Hotel Towel Reuse." *Psychological Science* 27 (7): 1043–1046.
- Stove, David C. 1973. *Probability and Hume's Inductive Scepticism*. Oxford: Clarendon.
- Stove, David C. 1982. *Popper and After: Four Modern Irrationalists*. Oxford: Pergamon Press.
- Stroud, Barry. 1977. *Hume*. London and New York, NY: Routledge & Kegan Paul.
- Sutton, Alex J., and Keith. R Abrams. 2001. "Bayesian Methods in Meta-Analysis and Evidence Synthesis." *Statistical Methods in Medical Research* 29 (11): 3424–3454.
- Sweeting, Michael J., Daniela De Angelis, Anthony E. Ades, and Matthew Hickman. 2008. "Estimating Hepatitis C Prevalence in England and Wales by Synthesizing Evidence from Multiple Data Sources: Assessing Data Conflict and Model Fit. *Biostatistics* 9 (4): 715–734.
- Tinbergen, Jan. 1936. *Grondproblemen der Theoretische Statistiek*. Haarlem: F. Bohn N.V.
- Tinbergen Jan. 1939a. *Statistical Testing of Business-Cycle Theories I: A Method and Its Application to Investment Activity*. Geneva: League of Nations Economic Intelligence Service.
- Tinbergen Jan. 1939b. *Statistical Testing of Business-Cycle Theories II: Business Cycles in the United States of America 1919–1932*. Geneva: League of Nations Economic Intelligence Service.

- Tinbergen, Jan. 1940. "On a Method of Statistical Business-Cycle Research. A Reply." *The Economic Journal* 50 (197): 141-145.
- Venn, John. 1876. *The Logic of Chance*. London: Macmillan and Co.
- Walach, Harald, and Martin Loef. 2015. "Using a Matrix-Analytical Approach to Synthesizing Evidence Solved Incompatibility Problem in the Hierarchy of Evidence." *Journal of Clinical Epidemiology* 68 (11): 1251-1260.
- Williamson, Jon. 2010. *In Defence of Objective Bayesianism*. Oxford: Oxford University Press.
- Williamson, Jon. 2013. "Why Frequentists and Bayesians Need Each Other." *Erkenntnis* 78: 293-318.

William Peden is a Research Assistant Professor in the Department of Philosophy at Lingnan University, Hong Kong SAR. He is currently part of the 'Philosophy of Future and Contemporary Science' project, with a focus on peer review and artificial intelligence. He was formerly a postdoctoral researcher at Erasmus University Rotterdam in the 'Jan Tinbergen: the Thinker' project.

Contact e-mail: <williampeden@ln.edu.hk>