The problem of induction in cosmology

ALEX LOMBARD

Abstract

This article problematises a crucial assumption in the methodology of modern cosmology, namely the cosmological principle (CP). Given the physical obstacles and limits peculiar to cosmological observation, cosmological inferences ordinarily take the regularity readily observed in our region of space-time as a basis for the purported self-same regularity of physical processes characteristic of unobservable space-time regions. Just this projection of the locally familiar to the globally unfamiliar is of the essence of the CP. This inductive procedure, however, is beset by serious logical difficulties, an analysis of which is undertaken herein, and the strength of which, it is argued, may constitute grounds for tempering our confidence in cosmological inferences on the nature of the large-scale structure of space-time. The article, all the while, seeks to share in and communicate the excitement at the contributions and prospects of a distinctively philosophical engagement with the conceptual problems of mathematical and theoretical physics, an engagement that is at the very heart of contemporary philosophy of science.

Introduction

Inductive inference has, in various ways, played – and indeed continues to play – a methodologically central role in modern physical cosmology. This type of inference takes past experience as a justification for knowledge of that which has not been experienced. I focus on this aspect of cosmological method, for it is here that the justification of inferences in cosmology faces distinct obstacles over and above the inductive problems that cosmology inherits from the terrestrial (or local) physics upon which it has built.[1] These obstacles chiefly revolve around the legitimacy of the bold extrapolations of experimental observation constitutive of the cosmological principle (CP) and, particularly in the years 1948–1965, of the perfect cosmological principle (PCP), the relevance of which is of more than passing historical interest. This paper argues that in the absence of confidence in the inductive propriety of the CP and

[1] I shall not, therefore, discuss these more classical problems of induction herein.

PCP, the cosmological enterprise is hardly able to get off the ground. The paper proceeds in two stages: first, the CP is defined and two crucial philosophical pitfalls in this definition are identified; second, the CP's problematic temporality is revealed via a discussion of the PCP, and the danger an unsatisfactory treatment of temporality poses to a proper elucidation of the physical dynamics of the universe is emphasised.

The CP and relativistic cosmology

Although only explicitly formulated in 1935 by Milne,[2] the CP is given its first implicit application by Einstein in his 1917 paper applying General Relativity (GR) to the large-scale structure of space-time, the paper which set the foundations for relativistic cosmology.[3] Here Einstein introduces his cosmological constant (λ) which attributes, in the form of an assumption, a structural uniformity – of the distribution of matter and radiation – to the universe necessary for making the latter tractable for GR.[4] Within a decade, Friedmann and Lemaître had independently presented alternative solutions to Einstein's field equations which challenged Einstein's static conception of the universe, providing in effect the mathematical basis for the Big Bang model of the universe.

The subsequent development of Friedmann-Lemaître-Robertson-Walker (FLRW) models in the 1930s explicitly formalised the CP as postulating that the large-scale structure of the universe is (1) spatially homogeneous and (2) spatially isotropic around each space-time point, on *sufficiently large* scales. (1) is defined such that observations obtained from any point in space-time are (approximately) the same as those obtained from any other such point.[5] (2) is defined such that differently directed observations from the same point in space-time are (approximately) the same.[6] Logically, (2) implies (1), but not

---

[2] Jeremy Butterfield, 'On Under-Determination in Cosmology', *Studies in History and Philosophy of Modern Physics*, vol. 46, May 2014, p. 60 and George Gale, 'Cosmology: Methodological Debates in the 1930s and 1940s', *Stanford Encyclopedia of Philosophy*, 21 June, 2017, sec. 3.2, online.

[3] Claus Beisbart, 'Can We Justifiably Assume the Cosmological Principle in Order to Break Model Underdetermination in Cosmology?' *Journal for General Philosophy of Science*, vol. 40, no. 2, 2009, p. 177.

[4] Crucially, Einstein's cosmological constant was also motivated by his need to correct for what he believed to be an otherwise unacceptable catastrophic implosion of the universe as a result of extant gravitational imbalances distributed throughout the cosmos. See Simon Singh's account in his *Big Bang: The Origin of the Universe*, HarperCollins, London, 2005, ch. 2.

[5] See, for example, Claus Beisbart and Tobias Jung, 'Privileged, Typical, or Not Even That? Our Place in the World According to the Copernican and the Cosmological Principles', *Journal for General Philosophy of Science*, vol. 37, no. 2, 2006, p. 234.

[6] Ibid, p. 242.

vice versa.[7] FLRW models therefore describe an expanding and – because homogeneous – unbounded[8] four-dimensional space-time emerging from a unique Big Bang singularity. Unpacking the CP's definition still further, two philosophical issues emerge: a methodological stress upon observation and a logical question concerning the vagueness in the meaning of the adjective phrase 'sufficiently large'. Insofar as the latter is consequent upon the assumption of the former, I examine each successively.

Unlike in typical terrestrial-focused natural sciences, the CP extrapolates from the observed[9] to regions of the universe long believed to be – ignoring for now the prospects of gravitational wave astronomy[10] – *unobservable* in principle. The threshold of the observable universe is taken to depend upon the propagation of light, the surest – because constant – physical signal and transmitter of cosmic information. The propagation of light, however, is taken to have begun with decoupling, about 380,000 years after the Big Bang. Furthermore, because the expansion velocity of the universe outstrips that of light, it follows that much in the way of light signals is precluded from possibly ever reaching us. The state of the early universe, then, and of the universe outside our past light-cone is logically unobservable.[11] Moreover, the ratio of the size of the observable universe to that of the entire universe is continually decreasing given the latter's expansion velocity. 'To put the point very simply,' writes Butterfield, 'in terms of enumerative induction over spacetime regions: the observable universe is such a small fraction of such regions, that it is risky to claim it is a fair sample.'[12] Perhaps the sample can make up for this quantitative (statistical) deficiency by its fine-grained qualitative detail.

Such a resolution of the sample size problem, however, quickly falters in practice, for the selection of appropriate scales for cosmic observations has been governed not by fine-graining, but by coarse-graining.[13] To conserve the similarity of observations demanded by the CP, spatial homogeneity and spatial isotropy have over the century of modern physical cosmology been measured at larger and larger

---

[7] Butterfield, p. 60; H. Bondi, *Cosmology*, Second edition, Cambridge University Press, Cambridge, 1960, p. 14; and Beisbart and Jung, p. 240.

[8] Beisbart and Jung, p. 233, write: 'It follows immediately from the definition of homogeneity that homogeneous systems cannot have boundaries.'

[9] By today's standards, for instance, FLRW models apply GR 'at length scales 14 orders of magnitude larger than those at which it has been tested' (Quoted in Christopher Smeenk and George Ellis, 'Philosophy of Cosmology', *Stanford Encyclopedia of Philosophy*, 21 December, 2017, sec. 1, online).

[10] See, for example, Joseph Silk and Jens Chluba, 'Next Steps for Cosmology', *Science*, vol. 344, 9 May 2014, pp. 586–7 and BF Schutz, 'Gravitational Wave Astronomy: Delivering on the Promises', arXiv, 17 April 2018, esp. sec. 8. The former talks of 'opening up a new window for exploration of the primeval plasma from which all structure originated', p. 587.

[11] Butterfield, pp. 62–3.

[12] Ibid, p. 63.

[13] Ibid, p. 61 and Beisbart and Jung, p. 242.

scales to blur out the obvious inhomogeneities we observe on smaller scales.[14] Most notably, this scale shift has been forced upon the CP by the discovery first of galaxy clusters, and then of galaxy superclusters;[15] the future discovery of further inhomogeneities requiring yet further upwards adjustment in the CP scales is, logically and physically, all too possible.[16] Today, the scale stands at over 300 million light years.[17] The qualification in the CP's definition that measurements of observations be made on 'sufficiently large scales'[18] is troubling for a scientific principle with as much significance as the CP, for it heaps a vague adverb ('sufficiently') upon a vague scaling adjective ('large'). The indeterminacy in the semantics of 'large' endows the CP with the flexibility to ignore the smaller-scale inhomogeneities. Yet this comes at a high cost, as the following section attests.

The PCP and steady-state cosmology

Although not a sufficient condition of the CP, spatial homogeneity is a necessary condition for it to hold. Measurement scales too large to capture physically significant inhomogeneities therefore risk jeopardising the inductive legitimacy of the cosmological enterprise. At issue here is the *temporality* of physics, which in 1948 Bondi and Gold made a central dissatisfaction with the CP and, by extension, of the Big Bang model of the universe.[19] To guarantee the universality of the laws of physics, they argued, the structure of the universe itself – 'which depends upon the physical laws'[20] – needed to be rid of those temporal distinctions by which any atemporal physics seeking to describe it would have to be modified. But the cardinal principle of science, they tell us, is the repeatability of experiments in the formulation of the laws of nature, which implies the *irrelevance* of the place *and* time of experimentation.[21]

The CP's commitment to spatial isotropy, Bondi and Gold believed, rightly does away with spatial contingency but does not do likewise concerning temporal contingency. They therefore proposed an

---

[14] Butterfield, p. 61.

[15] Ibid, p. 63.

[16] This is especially so given the observational constraints we face in studying the cosmos and, relatedly, the limited range of observational techniques at our disposal in cosmology. The progress of the latter shall hopefully reduce the former and give us a better 'view' of the universe. The advent of gravitational wave astronomy, to use a very contemporary example, may allow us to observe more deeply into the cosmos, and bring with it a new picture of the (in)homogeneity of the large-scale structure of space-time.

[17] Butterfield, p. 63 and Beisbart, p. 199.

[18] Quoted in Butterfield, p. 60.

[19] H. Bondi and T. Gold, 'The Steady-State Theory of the Expanding Universe', *Monthly Notices of the Royal Astronomical Society*, vol. 108, no 3, 1948.

[20] Ibid, p. 254.

[21] Bondi and Gold, p. 252 and Bondi, p. 12.

extension of the CP – called the perfect cosmological principle (PCP) – that fulfils this scientific requirement. The PCP defines the large-scale structure of the universe to be (1) spatially homogeneous and (2) spatially isotropic around each space-time point, on sufficiently large scales, *at all times*.[22] The PCP thus founds a new conception of the universe, namely the Steady-State model: like the Big Bang model, it is spatially unbounded, because of the homogeneity condition of the PCP; unlike the Big Bang model, however, it is also temporally unbounded, because of the 'at all times' stipulation of the PCP.

The Steady-State model posits an eternal space-time manifold, not one that emerged from a Big Bang singularity. By 1948, however, breakthroughs in the understanding of nucleosynthesis had been made by Hoyle, soon to become the most notable proponent of the Steady-State theory.[23] Together with Bondi and Gold, Hoyle demonstrated that a Steady-State universe could save the appearances of an expanding universe – the overwhelming observational evidence of which stretches back to Hubble's 1929 discovery of the red-shift in the spectral lines of receding galaxies – without conceding the universe's catastrophic emergence from a Big Bang singularity. This they did by positing a speculative[24] physical process of 'continual creation of matter'[25] via nucleosynthesis, which ensured that within each PCP scale, the same distribution of matter and radiation would be preserved *over time*.[26] Until 1965, the Big Bang and Steady-State models were each capable of fitting the observational evidence, but the latter's theoretical advantage remained its immunity to the logical problem of projecting – via inductive inference – a temporally-contingent physics (the physics developed over the past 400 years) onto a temporally-evolving universe.

With the discovery of the cosmic microwave background radiation in 1965, the Steady-State model – which predicted its *inexistence* – was taken to be empirically refuted, and the Big Bang model – which

---

[22] Bondi and Gold, pp. 254–5.

[23] See ch. 4 of Singh's account.

[24] Speculative, because it openly violated the principle of conservation of energy. But Bondi and Gold defended the physical hypothesis thus:

> In interpreting the universe as stationary we have to assume that such a process of creation is operative; we have to infringe the principle of hydrodynamic continuity. But this principle is not capable of experimental verification to such a precision, and this infringement does not constitute a contradiction with observational evidence. It is true that hydrodynamic continuity has been regarded as an unqualified truth and not as an approximation to physical laws, but this was merely a bold simplifying extrapolation from evidence. Hydrodynamic continuity is no doubt approximately true but this does not compel us to assume that it holds without any deviation whatever. In the conflict with another principle which is much more far-reaching and capable of making many more statements about the nature of the universe and the applicability of physical laws, there is no reason for upholding the principle of continuity to an indefinite accuracy, far beyond experimental evidence.' (Quoted in Bondi and Gold, p. 256.)

[25] Quoted in Bondi, p. 143.

[26] Ibid, p. 143 and Singh, ch. 4. The expansion of the universe, otherwise, would introduce a temporal diffusion of radiation and matter across each PCP scale.

predicted its existence – was considered validated to an unprecedented extent. Indeed, it very soon thereafter came to be referred to as the Standard Model (SM) in cosmology.[27] The SM's vindication, however, resulted not from a resolution of the temporality problem, but rather to its impressive empirical corroboration. Logically, the SM's commitment to the universality of the laws of physics formulated over the past 400 years sits uncomfortably with its description of a temporally inhomogeneous universe. The SM's admission that there is little warrant for applying the supposedly universal laws of physics to the very early universe[28] (and effectively none for the first 10^{-11} s after the Big Bang) is an explicit reminder that the spectre of temporality continues to haunt the SM.[29] And yet, what is at stake here is the very nature of the physical dynamics of the universe. The concern is that matter inhomogeneities – say at the time of decoupling – filtered out of a coarse-grained CP scale might be crucial input data in the elaboration of a relevant micro-dynamics. This comprises the physical mechanisms informing the evolution of these inhomogeneities into the various galaxies, galaxy clusters and galaxy superclusters that we observe today.[30] Insofar as the latter are governed by a macro-dynamics – namely, GR as presented in FLRW models[31] – the risk here is of the micro-dynamics not linking up, commuting, with the macro-dynamics: the 'question is whether this macro-dynamics is induced by a reasonable coarse-graining procedure on the unknown, myriadly complex, micro-dynamics'.[32] A half-century of work on the SM, in sum, has brought it impressive observational corroboration,[33] but little in the way of progress on the (inductive) logical front.

---

[27] Coined by Weinberg in 1972. (Chris Smeenk, 'Philosophy of Cosmology', in Robert Batterman, ed., *The Oxford Handbook of Philosophy of Physics*, Oxford University Press, Oxford, 2013, p. 609).

[28] Namely, the first hundredth of a second after the Big Bang.

[29] Butterfield, p. 57n.

[30] Ibid, p. 65.

[31] Ibid, p. 65.

[32] Ibid, p. 65. A simple counterfactual analogy may illustrate the problem. Suppose that an extra-terrestrial biologist were to land at a nursing home on Earth and encounter for the first time a few hundred adult humans older than 50, having no previous knowledge of human biology. The biologist quickly notices and studies the phenomenon of sarcopenia, or age-related muscle degeneration. The biologist observes all sample subjects undergoing the same types and rates of muscle degeneration, and subsequently infers that sarcopenia is an intrinsic feature of human biology *at all stages of life*, blithely unaware that the mechanism in question only begins beyond the age of 50. Here, the sample size was too coarse-grained to allow relevant inhomogeneities to inform sound inferences about the mechanisms of human biology. The biologist is tripped up by the absence of a *temporally sensitive* biology. The absence is costly: the biologist's inferences fail to capture the fact that *precisely the opposite* of what they conclude occurs at other stages of life.

[33] See, for instance, Singh's account of the decades-spanning COBE mission, in his ch. 5.

Coda

The problem of induction in cosmology cuts across the realism/instrumentalism debate in the philosophy of science. Realism, whereby scientific – cosmological, in this case – theories are, roughly, taken to be representing the physical world, is obviously affected by the problems identified above. Less obvious is the fact that instrumentalism, too, is unable to remedy these problems. For even instrumentalism, whereby scientific theories are taken to be computational devices or instruments for making predictions, presupposes in this case cosmological observations to verify predictions. The interpretation of cosmological observations, however, often employs the very assumptions the observations are expected to corroborate.[34] Most notably, cosmological observation's reliance upon light propagation is problematised by the fact that our physical understanding of the latter is largely derived from *homogeneous* experimental space-time regions,[35] and further by the fact that inhomogeneities are known to distort it.[36] Vicious logical circles thus impinge upon any innocent recourse to instrumentalism to transcend the problem of induction in cosmology.

Bibliography

Beisbart, C, 'Can We Justifiably Assume the Cosmological Principle in Order to Break Model Underdetermination in Cosmology?' *Journal for General Philosophy of Science*, vol. 40, no. 2, 2009, doi.org/10.1007/s10838-009-9098-9

Beisbart, C, and Jung, T, 'Privileged, Typical, or Not Even That? Our Place in the World According to the Copernican and the Cosmological Principles', *Journal for General Philosophy of Science*, vol. 37, no. 2, 2006, doi.org/10.1007/s10838-006-9015-4

Bondi, H, and Gold, T, 'The Steady-State Theory of the Expanding Universe', *Monthly Notices of the Royal Astronomical Society*, vol. 108, no. 3, 1948, doi.org/10.1093/mnras/108.3.252

Bondi, H, *Cosmology*, Second edition, Cambridge University Press, Cambridge, 1960

Butterfield, J, 'On Under-Determination in Cosmology', *Studies in History and Philosophy of Modern Physics*, vol. 46, May 2014

---

[34] Beisbart and Jung, p. 246 and Beisbart, p. 183. Consider for instance Bondi's admission in the preface to Part 2 – on observational evidence – of his influential textbook on cosmology, which reads, in part: 'In the discussion of the observational evidence the cosmological principle is used. This procedure seems to be justified owing to the wide agreement on the validity of the cosmological principle and also because its absence would make any interpretation almost impossible.' (Quoted in Bondi, p. 17).

[35] Beisbart and Jung, p. 246 and Beisbart, p. 184.

[36] Butterfield, p. 65.

Gale, G, ‘Cosmology: Methodological Debates in the 1930s and 1940s’, *Stanford Encyclopedia of Philosophy*, 21 June 2017, plato.stanford.edu/entries/cosmology-30s/

Schutz, BF, ‘Gravitational Wave Astronomy: Delivering on the Promises’, arXiv, 17 April 2018, royalsocietypublishing.org/doi/abs/10.1098/rsta.2017.0279

Silk, J, and Chluba, J, ‘Next Steps for Cosmology’, *Science*, vol. 344, 9 May 2014, doi.org/10.1126/science.1252724

Singh, S, *Big Bang: The Origin of the Universe*, HarperCollins, London, 2005

Smeenk, C, ‘Philosophy of Cosmology’, in Robert Batterman, ed., *The Oxford Handbook of Philosophy of Physics*, Oxford University Press, Oxford, 2013

Smeenk, C, and Ellis, G, ‘Philosophy of Cosmology’, *Stanford Encyclopedia of Philosophy*, 21 December 2017, plato.stanford.edu/entries/cosmology/