Chapter 9
Inductive Inferences on Galactic Redshift, Understood Materially

John D. Norton

This paper is dedicated with gratitude to Professor Roberto Torretti, whose work and personal example exercised a formative influence on me. He set a standard of clarity and precision in philosophical writing that I have long sought to emulate. He also showed me that precision in writing and a wicked sense of humor can cohabit. The exhilarating time we spent together in 1983 as Fellows in the Center for Philosophy of Science at the University of Pittsburgh remains as vivid to me as if it happened yesterday.

Abstract A two-fold challenge faces any account of inductive inference. It must provide means to discern which are the good inductive inferences or which relations capture correctly the strength of inductive support. It must show us that those means are the right ones. Formal theories of inductive inference provide the means through universally applicable formal schema. They have failed, I argue, to meet either part of the challenge. In their place, I urge that background facts in each domain determine which are the good inductive inferences; and we can see that they are good in virtue of the meaning of the pertinent background facts. This material theory of induction is used to assess the competing inductive inferences in the debate in 1972 between John N. Bahcall and Halton Arp over the import of the redshift of light from the galaxies.

I thank Siska De Baerdemaeker for helpful comments on an earlier draft.

J. D. Norton (✉)
Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA
e-mail: jdnorton@pitt.edu

C. Soto (ed.), *Current Debates in Philosophy of Science*, Synthese Library 477,
https://doi.org/10.1007/978-3-031-32375-1_9

9.1 Introduction

Good science is distinguished from other imaginative narratives by the fact that we have good evidence for its extraordinary accounts. If we cannot recover this most basic fact about science, our efforts to develop a cogent philosophy of science have come to nothing. Yet we, as philosophers of science, are faring poorly at the task. Our accounts of inductive inference have so far been a poor fit with the actual practice of science. The failure has stung most for me when I turn to the history of science. For there we see how the scientists found and weighed the evidence advanced in support of our science. In exploring this history, we should be able to use our accounts of inductive inference to identify when the efforts to provide good inductive support for our science have succeeded; and when they have failed.

However, in recounting various episodes in the history of science, I found it hard to use existing accounts of inductive inference to assess the cogency of the inferential claims. In specific cases, I would commonly find some account in the scattered repertoire that would fit. But the fit would be Procrustean. An inference may look initially like a strong inference to the best explanation. But the strength of the inference depended on an assessment of the quality of the explanation. We have no good account of that quality. Take the wonderfully crafted, intricate optical instrument that is the eye. Just why is Darwin’s complicated and incomplete explanation of its evolution better than the simplicity of the creationist’s single supernatural hypothesis?

In other cases, it was not so hard to fit Bayes’ theorem to some judgment of evidential support. The nagging worry was that the scientists themselves were not using Bayes’ theorem or even formulating probabilities. They have their own explicitly non-probabilistic ways of proceeding. Worse, the real inductive work was rarely done by the theorem. It was done in the selection of likelihoods and other probabilities. Once the informal reasoning that led to that selection was exposed and appreciated, actually computing Bayes’ theorem became a superfluous afterthought. Indeed, shrewd selection of the probabilities seemed able to push Bayes’ theorem to give almost any desired result.

Some reflection on the nature of inductive inference was needed. Present accounts of inductive inference are almost exclusively formal and universal. In this regard, they are modeled on accounts of deductive inference. They provide formal schemas or templates for good inductive inferences or for appropriate relations of support; and the schemas are applied by substituting factual content into the slots of the schemas. The schemas are universal in the sense that they can be applied equally in any domain. One merely needs to substitute content from any domain into them.

While each of the accounts in the present repertoire can boast of successes in some specific cases, none succeeded universally. At best the repertoire provides us with a patchwork. One scheme works here; a different one works there. It soon became clear that this patchwork character is actually the essence of the matter. It is central to the material theory of induction, first formulated in Norton (2003). According to it, there are no universally applicable schema for inductive inference.

Rather the licit inductive inferences in each domain are warranted by the true, contingent facts of that domain. This account, elaborated briefly in Sects. 9.2 and 9.3 below, meets the dual requirements for a successful account of inductive inference. First, the account must tell us which items of evidence provide inductive support for which propositions. The background facts of the domain do that. Second, it must provide good reasons for why the particular selections of the account are properly cases of inductive support. Those reasons derive directly from the meaning of those true background facts.

The bulk and the remaining part of the paper, Sects. 9.4–9.12, illustrates the utility the material theory of induction in a particular piece of history of science. It is the debate in 1972 between John N. Bahcall and Halton Arp over the import of the redshift of light from the galaxies. Does this redshift reveal an expansion of the universe, as then standard cosmology asserted? Or does the redshift derive from other processes, as yet not fully understood? We shall see how the material theory of induction allows us to delineate and appraise the inductive inferences of each side of the debate. We shall see that central to the debate was the establishing by each proponent of the background facts needed to support their inductive inferences. They also put considerable effort into impugning the corresponding background facts of their antagonists. We also see how the inductive reach of the two sides depended on whether they could access background facts that would support further inductive inferences.

9.2 The Material Theory of Induction

The central thesis of the material theory of induction is that inductive inferences are not warranted by their conformity with universally applicable schema, but by true background facts.[1] These background facts distinguish those inductive inferences that are well-warranted; and the cogency of the warrant is assured by the meaning of the background facts. Since there is no universally applicable, factual principle of induction, there are no universal warranting facts. Hence each warranting fact obtains only in some limited domain. In such a domain, a warranting fact may be rich enough to induce some formally expressible system or logic of inductive inference. However, that logic will be warranted only within that specific domain and may not apply elsewhere. In this sense, all inductive inference is local.

The simplest argument for the material theory proceeds from the ampliative nature of inductive inference. In such an inference, the conclusion always asserts more factually than the premises. Hence any such inference can fail if attempted in a domain that is inhospitable to the inference. The requirement that the domain be hospitable to that inference is, in its most generic form, the background fact of the domain that warrants the inference.

---

[1] For considerable elaboration on this sketch of the material theory, see Norton (2021, Ch. 1, 2)

9.3 Material Successes

The present literature on logics of induction is enormous and offers us many competing formal systems designed to be universally applicable. Each system works somewhere but turns out to be at best a contrived fit elsewhere. The material theory of induction explains why each such system works where it does: the background facts are hospitable to it. And it explains why they fail when they do: the background facts are inhospitable.

Since there are so many competing accounts of inductive inference, it is practical here only to provide a few illustrations of how these accounts are accommodated by the material theory of induction. Following the categorization of accounts of inductive inference of Norton (2005), we may group accounts of induction into three broad families. Those in the family of inductive generalization proceed under the principle that an instance confirms the generalization. Its simplest form is enumerative induction: from the evidence that some *A*s are *B*, we infer that all *A*s are *B*. The difficulty with this universal schema is that is almost always fails. It only works if the *A*s and *B*s are very carefully chosen. In spite of millennia of efforts, no general rule has been found for specifying which *A*s go with which *B*s.

The material theory of induction entails that no general rule has been found because there is no general rule to be found. The commonality of instances of enumerative induction is superficial. Each or each grouping of them is warranted by background facts peculiar to the pertinent domain. Marie Curie in 1903 found that a mere tenth of a gram of Radium Chloride was crystallographically like Barium Chloride. She had no hesitation in generalizing from that tenth of a gram to all samples of Radium Chloride. This inductive inference was not warranted by a general schema such as enumerative induction. She did not infer to many other conclusions authorized by this schema: that all samples of Radium Chloride must be less than a tenth of a gram, or all must be in Paris, or all must be prepared by Curie. Rather the specific inference she did make was warranted by a hard-won fact of nineteenth century mineralogy known as "Haüy's Principle": all crystalline substances fall into a small set of families, distinguished by the configuration of the axes characteristic of the shape.[2] Other members of the family include Hempel's satisfaction criterion, Glymour's bootstrap, Mill's methods and arguments from analogy.[3] Each can be accommodated within the material theory of induction by comparable analysis.

A second family of accounts of inductive inference, hypothetical induction, is based on the principle that that ability of some hypothesis to entail the evidence is a mark of its truth. This principle, by itself, assigns the mark too indiscriminately to be viable. Many accounts add further conditions in order to restrict this assignment to cases to which it properly belongs. Each account has superficial plausibility,

---

[2] For further details of this example, see Norton (2021, Ch. 1)

[3] For the case of analogy, see Norton (2021, Ch. 4).

but none succeed generally. Rather they succeed where they do because of facts obtaining locally in the pertinent domain.

As an added condition, we might require that the successful hypothesis must also be simple. Notoriously, there is no general account of what makes an hypothesis simple. Rather, simplicity is merely a compactly expressible surrogate for background facts that vary from domain to domain.[4] Alternatively, we might add a severe testing requirement: that had the hypothesis failed to entail the evidence, then it would most likely be false. There is no universal way to implement this severe testing condition. It is realized only by determining factually in each domain what is most likely. Finally, in inference to the best explanation, we require that the hypothesis not just entail the evidence, but that it also explains it. This scheme has proven hardest to explicate in material terms simply because there is no precise formal specification of inference to the best explanation. Notably, there is no general, formal characterization of explanation such that its addition to mere deductive entailment of the evidence would boost the inductive strength associated with that entailment.[5]

In the third family of accounts, relations of inductive support are characterized by some explicit calculus. The dominant example is the probability calculus, as employed in Bayesian confirmation theory. There are many cases in which this sort of probabilistic analysis captures inductive support relations well. However, these are cases in which the particular background facts in turn authorize probabilistic relations of support, such as when we reason over samples drawn randomly from a population of known composition. Might the Bayesian system aspire to cover all relations of inductive support? Might it be that all relations of inductive support are just probabilistic facts; and all generalities about inductive support are theorems of the probability calculus?

These aspirations fail. For example, invariance arguments can show that, as the evidence becomes so weak as to be completely neutral, it enters into relations of inductive support that are inherently non-additive, in contradiction with the additivity axiom of the probability calculus. The Bayesian literature has laudably taken on the burden of proving that all relations of inductive support or, alternatively, all distributions of belief, must conform with the probability calculus. These proofs fail for a simple reason of logic. They seek to prove a contingent matter, that all these relations are probabilistic, by deductive means. Hence, they must begin with assumptions that are logically at least as strong as the conclusion sought. These assumptions must then covertly already presume the very thing we seek to prove. Once one knows to look for it, all the proofs are undone merely by demonstrating their circularity.[6]

---

[4] Or so it is argued in Norton (2021, Ch. 6 and 7).

[5] Or so it is argued in Norton (2021, Ch. 8 and 9), where examples of inferences to the best explanation are accounted for materially.

[6] See Norton (2021, Ch. 10–16) for an extended development of this remark and the other criticisms of Bayesianism's universal aspirations.

Finally, the now dominant subjective Bayesianism has compounded these problems by replacing objective relations of inductive support as the primitive notion with a subjective relation of belief. Objective support relations are now somehow to be wrestled from the subjective relations. This immerses the problem of understanding inductive inference in a larger problem of finding the formal relations that govern both inductive support and belief. However once one entangles mere, ungrounded opinions with objective relations of inductive support, it has proven too difficult to disentangle them. The much-vaunted convergence theorems cannot "wash out the priors" in the simple and common case in which two hypotheses entail the same body of evidence. The ratio of the posterior probabilities always remains equal to the ratio of the prior probabilities. The subjective Bayesian's probabilities are then an inseparable amalgam of arbitrary opinion and inductive support.

9.4 Cosmological Redshifts and the Recession of the Galaxies

The remainder of this paper reviews an historical debate in cosmology over the import of the redshift of light from galaxies. It came to a head in 1972 with a lively confrontation between the astronomers John N. Bahcall and Halton Arp. In it we shall see, as the material theory of induction predicts, that the debate hinges on the background warranting facts required to warrant the inferences on each side, even though they are sometimes only present tacitly. They are the assumption *Typicality*, whose truth is required to warrant Bahcall's inferences; and *Proximity*, whose truth is required by Arp's inferences. Much of debate involves efforts on each side to support their own assumption and to impugn the key assumption needed by the other.

A further consequence of the material view is that inductive inferences are possible only in so far as warranting background facts can be secured. We shall see in the debate below that Bahcall's side had considerable reach in its inductive inferences since it could call upon sufficient background facts to warrant them. Arp's side, however, had limited inductive reach precisely because of the dearth of suitable background facts.

The context of this debate was the single most important astronomical finding underpinning modern cosmological theory. According to the thesis of the recession of the galaxies, on average, the galaxies are receding from us with a velocity that increases in direct linear proportion to their distance from us. The fabled origin of the finding is a paper by Edwin Hubble (1929) with the transparent title: "A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae." Establishing the relation required determinations of both velocities of recession and distances to the galaxies (Hubble's "extra-galactic nebulae"). The determination of the distances proved most troublesome and required some ingenious analysis on Hubble's part. Even so, his estimates of distances were almost an order of magnitude too small.

The determination of the velocities, however, seemed straightforward. Hubble had access to Slipher's measurements of frequency shifts in the spectra of galaxies towards the red. These redshifts were interpreted as Doppler shifts, in accord with an effect widely recognized in more local physics: the frequencies of waves emitted by a receding body are diminished by the recession of the source. The pitch of sound is lowered and the color of light is reddened. The Doppler shift had been established for stars within our galaxies. It reveals the motions of binary stars orbiting about each other, for example.[7] If we extend the effect to the galaxies, the magnitude of the redshift provides a convenient observational proxy for the velocity of recession of the galaxies.

From the outset, the connection between redshifts and velocities of recession seems scarcely to have been challenged. In an early paper reporting some of his first redshift measurements, Slipher (1912) remarked that the size of a velocity inferred "... raises the question whether the velocity-like displacement might not be due to some other cause, but I believe we have at the present no other interpretation for it." Hubble (1929) wrote as if the redshift was an uncontroversial proxy for velocity. More careful reading of his writings, however, reveals a prudent caution. By the time of his semi-popular *Realm of the Nebulae*, his view was asserted as (1936, p. 34): "Although no other plausible explanation of red-shifts has been found, the interpretation as velocity-shifts may be considered as a theory still to be tested by actual observations." In his history of cosmology, Kragh (2007, Ch. 3) recounts Hubble's persistent hesitations throughout his life to accept redshifts as betokening velocities of recession. Kragh also recounts the ideas of a number of dissident cosmologists who explored alternatives to velocities of recession as the origin of the galactic redshifts.

These hesitations had little effect on mainstream astronomy. Weinberg (1972, pp. 417–18) summarized the established view as:

> The announcement by Edwin Hubble[8] in 1929 of a "roughly linear relationship between velocities and distances" established in most astronomer's minds the interpretation of the red shift as a cosmological Döppler effect, and this interpretation has survived through the decades until the present.

A brief survey other cosmology texts written around this time and more recently indicates this view as widespread. There is scant if any mention at all of the possibility of another interpretation of the redshift of the galaxies.

---

[7] Aitken (1918) is a synoptic survey of double stars and gravitationally bound binary star systems. The interpretation of spectral shifts in the light from binary starts as velocity-derived Doppler shifts was then already a standard, heavily exploited method of analysis. Its origins derive from Pickering's August 1889 observation of double lines in a stellar spectrogram (p. 27).

[8] Weinberg's footnote is to Hubble (1929), misdated in an apparent typographical error to 1927.

9.5 The Redshift Controversy

The acceptance of the relationship of redshift and velocity of recession was not untroubled. The trouble was Halton Arp, an observational astronomer who became a strident dissenter. He harbored mounting suspicions that the standard account of the formation of spiral galaxies did not fit well with their observed configurations.[9] These suspicions compounded into a general sense that there were many galaxies whose observed character was similarly "peculiar." The resulting *Altas of Peculiar Galaxies* (1966a) was able to organize these peculiar galaxies into roughly similar groups. The hope expressed in the preface was that the *Atlas* could "not only clarify the working of the galaxies themselves, but reveal physical processes and how they operate in galaxies, and ultimately furnish a better understanding of the workings of the universe as a whole." The hope, we can see, is for science practiced in the Baconian vein. We collect our observations and let the science flow inductively from them.

Arp soon fixated on a quite specific anomaly. As he reported (Arp 1966b) to *Science*, he noted cases of objects that appeared to be physically connected in ways incompatible with the standard interpretation of galactic redshifts as velocities of recession. Most notable were cases in which one object seemed to be ejected from another, so they must be close to each other on cosmic distance scales. Yet these same objects might have greatly differing redshifts. This immediately cast doubt on a simple relationship between distances and redshifts; and thus also on the corresponding relationship of distances and velocities of recession. These "discordant redshifts," as he soon came to call them, formed the basis and the near entirety of Arp's subsequent efforts to impugn the standard view of galactic redshifts.

Arp's concerns could not be ignored. He was well credentialed as an astronomer. He was a Harvard graduate, earned a PhD from Caltech and was a staff member of the Palomar Observatory. His complaints were collecting sympathizers. George B. Field, chairman of the Astronomy Section of the American Association for the Advancement of Science for 1972, noted few opportunities for what he labeled "direct confrontation."[10] He arranged for a debate at the AAAS meeting of December 30, 1972, in Washington, D.C. Halton Arp was to defend his claims of discordant redshifts and John Bahcall, an astronomer at the Institute for Advanced Study in Princeton, would defend the standard view. After the event, papers derived from their presentations, as well as background papers nominated by each debater, were published in the volume, *The Redshift Controversy.*

Field reported (p. 12) that the volume specifically "address[ed] the observational evidence on discordant redshifts." That means that this volume provides us with something tailor-made to the present concerns on evidence and inductive inference:

---

[9] As Arp reported in the Preface to Arp (1966a).

[10] As he reported in his introduction to Arp and Bahcall (1973, p. 4).

it is a self-contained snapshot at the time of the debate of the evidence for and against the standard interpretation of the galactic redshifts. Here are the leading experts on both sides laying out their strongest cases. Hence the analysis below will be confined largely to the cases laid out in the *Redshift Controversy* volume. This volume does not mark the end of the debate. Arp continued to press his case for discordant redshifts. However, it does seem to mark the end of the fleeting interest of mainstream astronomers. Most commonly, Arp's concerns are not even mentioned in cosmology texts after the time of the debate.

9.6 For Redshifts as Distance Indicators

John Bahcall's contribution to the volume was devoted to defending the standard view that the galaxies are receding from us in a unified motion whose magnitude increases linearly with distance. The inference to this standard view has two components: first, there is a roughly linear relationship between galactic distance and redshift; and second, a galactic redshift is linearly related to a velocity of recession. Bahcall's arguments focused just on the first of these two components, for that was the component directly under threat from Arp's arguments. As a result, Bahcall's contribution was explicitly making the case for "redshifts as distance indicators," these words being the title of his contribution. Bahcall's case is in two parts. The first, to be reviewed here, made the positive case. The second, to be reviewed later, sought to overturn Arp's case for discordant redshifts. To be precise, the goal of Bahcall's positive case is to support this claim:

> *Redshift-Distance Relation.* The redshift of light from the galaxies increases linearly with the distance to the galaxy, confounded but not obscured by small deviations due to local "particular" motions of the galaxies.

Bahcall describes (e.g. pp. 65, 77) the positive case as resting on the passing by the standard view of six tests. That is, the view makes predictions that, if unsuccessful, would undermine the view. The tests are passed, he decides.

While this is a popular way of characterizing the evidential import of the evidence, it does not capture the strength of Bahcall's case. For passing a test can fail to provide the extent of support needed. An hypothesis of cosmic contraction predicts a blue shift in light from the galaxies. That prediction provides a test of the contraction hypothesis. Hubble's (1929) paper reported five galaxies whose blueshifts indicated velocities of approach. They are NGC 6822, 598, 221, 224 (Andromeda) and 3031. Had we known only these data, we would have found the test to be passed. However, we might not find the passing of the test a strong support for the contraction hypothesis, since these are mostly nearby galaxies. Their motions might be local, particular motions, not a manifestation of a larger cosmic motion. Worse, I selected these galaxies from Hubble's data precisely because they manifest a blueshift.

An effective regime of testing must somehow probe more fully the hypotheses under test, so that passing the tests is evidentially more potent that merely parrying a threat. That is the case with Bahcall's tests. They are set up and passed in such a way as to give extensive inductive support to the standard view. To see this, we need to move beyond the simple logic of tests. Viewed materially, this stronger relation of inductive support derives from a background assumption that, if true, can warrant the support relations. That assumption is:

> *Typicality.* The relations among redshifts, distances and absolute magnitudes reported for the galaxies surveyed in the evidence are typical of all galaxies in our vicinity.

The assumption assures us that we have not ended up selecting galactic data that will incorrectly bias our inductive inferences. Our Milky Way galaxy precludes easy extragalactic observations in parts of the sky. *Typicality* assures us that galactic behaviors there conform with those found in other, more accessible parts of the sky. The assumption may at first appear benign. However, it will prove to be at the center of the debate. The assumption does not hold for the fictitious case just proposed of a cosmic contraction. The five galaxies in the data set were selected specifically because of their blueshift. They are not typical.

Bahcall's inferences to the redshift-distance relation are authorized by this assumption, in so far as it is true. We shall see that, built into Bahcall's tests, is evidence directly for the assumption.

9.7 Bahcall's Positive Case

Bahcall concluded his positive case (p. 79) with a summary of the six tests passed by the standard view. They are quoted in turn below, with my commentary interleaved.

> *First*, the original relation between redshift and apparent brightness was tested and found valid for the brightest galaxies over a redshift range that is more than one hundred times larger than the range originally available to Hubble.
>
> *Second*, the average apparent brightness of all galaxies irrespective of type decreases[11] like (redshift)$^{-2}$, with a scatter about the mean relation that can be understood in terms of the range in intrinsic luminosities.

These two tests refer to the most direct approach: one checks that the redshift and distance of many galaxies fall under the same linear relationship. If the assumption of *Typicality* is correct, then that assumption authorizes an inductive inference to the linearity of the relation universally.

The complication is that the determinations of distances to galaxies is difficult. The principal technique relies on the fact that the brightness of a galaxy diminishes with the inverse square of distance. However, to use this inverse square diminution

---

[11] Brightness decreases as (distance) $^{-2}$. Thus, a linear relationship between distance and redshift is equivalent to brightness decreasing as (redshift)$^{-2}$.

of brightness to determine distance requires knowing the absolute brightness of the galaxy. Two galaxies of same apparent brightness might be placed at very different distances from us. That could happen if the nearer galaxy has a smaller absolute magnitude and the more distant galaxy a greater absolute magnitude. Hubble's (1929) original work included galaxies close enough to us so that individual stars in them could be resolved. This enabled distance determinations based on the characteristic behaviors of stars like the Cepheid variables. By the 1970s, however, the investigations included galaxies well beyond the distances at which their individual stars could be resolved.

Bahcall reports two strategies to solve the problem. Galaxies commonly collect into clusters of several hundred. If the range of galactic types in each cluster is comparable, then the brightest galaxies in each cluster might have similar absolute magnitudes.[12] Differences in their apparent magnitudes could then serve as a surrogate for differences of distances. Hence, in the first test, Bahcall reports the most recent results of Sandage (1972): the redshift-distance relation holds for redshift and apparent magnitudes of the brightest galaxies in each cluster measured.

The second of Bahcall's test uses a different, less satisfactory way of compensating for differences in the absolute magnitudes of the galaxies. The absolute brightness of galaxies varies considerably. However, if we plot redshift against apparent brightness for many galaxies, we would expect the redshift-distance relationship to manifest as an average amongst scattered data, where the scatter is due to the differences of absolute brightness. Bahcall reports that just such a confounded relationship had been reported in Humason et al. (1956) for many galaxies.

The inductive inferences in these first two tests are warranted by the assumption *Typicality*. This assumption can only supply the warrant if it is true. This second test provides a means for Bahcall to give support to the assumption. He stresses (p. 71, his emphasis) that Humason and et al.'s analysis gives the result "for *all types* of galaxies"; and that:

> The most important inference to be drawn from Fig. V [of Humason and et al.] is that galaxy redshifts seem to be good distance indicators for average galaxies, not just for the brightest ellipticals.

The first test gave more accurate results. However, its data was restricted to the brightest galaxies. Might this selection of special galaxies compromise the typicality of the relation found? That a similar, albeit confounded, relation is found for all types of galaxies is inductive support for typicality.

This last inductive inference requires a background warranting fact. It is the assumption that a failure of typicality would most likely arise as different redshift properties for different types of galaxies; and it is otherwise unlikely. Using this fact we proceed from the agreement of redshift-distance relations among the brightest galaxies and all galaxies to infer that a failure of typicality is unlikely.

---

[12] Sandage (1972, p. 1), a paper cited by Bahcall, concludes just this similarity.

Bahcall's third and fourth tests addresses the confounding of the linear relationship due to local "particular" motions of the galaxies. The confounding must be small, else the linear relation is lost:

> *Third*, the mean redshift of double galaxies was found to be much larger than the difference in redshifts between the two members of each such pair.
>
> *Fourth*, the mean redshift of galaxies in a rich cluster was found to be much larger than the dispersion in redshifts of individual galaxies in the cluster.

These results affirm small confounding in the two cases considered: pairs of close galaxies and galaxies in rich clusters. The assumption of *Typicality* allows us to extend this smallness to all galaxies.

Once again, the inductive inference of the extension depends on *Typicality*. Bahcall notes that (his emphasis, p. 73):

> It is important to note that Page's results [on pairs of close galaxies] refer to double galaxies of *all types*: spirals, ellipticals, and irregulars.

As before, the constancy of the results over all types of galaxies provides inductive support for *Typicality*.

> Fifth, the apparent angular diameter of the brightest galaxies in rich clusters decreases as $(\text{redshift})^{-1}$.

This fifth test relied on a different, geometric method of determining distances. For any fixed, distant object, the distance is inversely proportional to its angular size. Bahcall reported Hubble's demonstration that, for each fixed class of galaxies, angular sizes on average decrease as $(\text{distance})^{-1}$. We also have that the angular diameter of the brightest galaxies decrease as $(\text{redshift})^{-1}$. Combining them we have on average a linear relationship between distance and redshift, for those galaxies mentioned. *Typicality* allows us to extend the relationship to all galaxies.

The final test is:

> Sixth, redshifts that are determined by radio and optical techniques agree to high accuracy.

This agreement is relevant since a velocity-derived Doppler shift in frequency must be the same across the spectrum. Its evidential significance is that it narrows the physical mechanism that might be responsible for the redshift. Through *Typicality,* it precludes mechanisms that produce redshifts by affecting frequencies differentially. However, it does not preclude a gravitational redshift such as general relativity associates with intense gravitational field.

Bahcall concludes, however, that (his emphasis, p. 77): "Note also that the 130 galaxies included in Fig. VIII include galaxies of *all types*." As before, this independence from type provides support for *Typicality*.

9.8 Arp's Discordant Redshifts

Arp's contribution to the *Controversy* volume had an essentially negative goal: to impugn the relationship between redshift and distance of the standard view. The case depended on finding discordant redshifts, that is, galactic bodies at similar distances from us but with significantly differing redshifts.

In the early pages of his contribution, his goal appeared to be a deductive refutation of an exceptionless redshift-distance relation. He prefaced his remarks by recalling how astronomers had become so confident of the redshift-distance relation that they routinely used redshifts to determine distances. He wrote (p. 20):

> … if we can produce just one example of a redshift difference that cannot be explained as a velocity difference, then we have broken the assumption on which the redshift-distance relation is always applied to derive distances.

Arp, however, overstated the import of just one anomalous case. He continued

> In this eventuality it would then become necessary to reexamine each category of the different kinds of galaxies in order to see whether current distance assignments would need to be revised.

A single counterexample would demonstrate only that the relation had an exception. That left the possibility that the counterexamples were so rare that redshift-distance relation was scarcely compromised. It would be reliable but not infallible.

It soon became clear that Arp did have a stronger result in mind. He mentioned in passing on p. 24 that "… evidence to be produced later in this paper [will show] that nonvelocity redshifts are a quite general phenomenon …" Establishing that generality would be sufficient to overturn Bahcall's inductive inferences supporting the standard view. For, if the generality is established broadly enough, it would show that *Typicality* fails. The connections between redshift and distance for the galaxies reported in support of the standard view would simply be true of those galaxies explicitly within Bahcall's data and not a broader generality.

To establish this generality, Arp produced with a catalog of instances. The first concerned quasars (pp. 20–31). They are "*quas*i-stell*ar* objects," that is, objects that look like unresolved stars, but manifest high redshift. The redshifts are so high that quasars must be at very remote distances, if the redshift-distance relation applies to them. It then follows that they must also be of extraordinarily great brightness.[13] Arp urged an alternative reading that escaped the need to posit extraordinary brightness. He urged that quasars are distributed statistically such that they are near galaxies in clusters or groups that are closer to us. Arp (p. 29) also reported what he boasted to be "the 'experimentum crucis'." He had found a line of four quasars that he interpreted as so aligned because they were ejected from a nearby galaxy.

---

[13] This was the standard view at the time of Arp's writing and the one now also accepted. Quasars are now believed to be extremely bright galactic nuclei.

The catalog continued in this vein. Arp turned to galaxies with discordant redshifts. He proceeded thorough various cases. He found that fainter galaxies in clusters tended to have greater redshifts than others in the clusters, for example. He continued with interacting double and multiple galaxies with discordant redshifts; chains of galaxies with discordant redshifts; galaxies connected by filaments with discordant redshifts; and tight groupings of galaxies with discordant redshifts.

Arp's analysis contained an inductive inference that would be targeted for extended criticism by Bahcall. Arp had used proximity in the starfield and other related features as supporting a physical connection between objects so that they are equally distant from us. This assumption was essential to the inference that the discordant redshifts violated the redshift-distance relation. We might express the assumption as:

> *Proximity*. An indicator of equality of distance is proximity of galactic objects in the star field, along with further features such as connecting filaments and alignments compatible with ejection.

The cogency of Arp's case depends upon this assumption. There is an initial plausibility to it. If we see filaments looking as if they are connecting objects in the star field, if we see alignments looking as if one object is ejected by another, perhaps they really are physically connected and thus at the same distance from us. Arp also used his data to support the assumption in a way similar to the way that Bahcall had sought to support *Typicality.* Arp concluded (pp. 55–56):

> It cannot be stressed too strongly, however, that these discordant redshifts are not discovered in just one or two isolated cases that have no relation to each other. But in every case we can test–large clusters, groups, companions to nearby galaxies, companions to middle-distance galaxies, companions linked by luminous filaments, galaxies interacting gravitationally, chains of galaxies–in every conceivable case, we come out with the same answer: the same discordant redshifts for the same general class of younger, fainter galaxies. This evidence, taken together with the same kind of evidence for the quasars–which are a kind of extremely young and, if this evidence is correct, intrinsically faint companion–forms a coherent picture of the kind of galaxies that have excess intrinsic redshifts.

This is evidence for Arp's overall case and, *a fortiori*, his assumption *Proximity*. It is a new inductive inference that is based on a further assumption: Were the indicators of *Proximity* unreliable, we would not recover consistent results such as reported here over many cases; but we likely would recover them if the indicators are reliable.

9.9 Bahcall's Rejoinder

The negative part of Bahcall's contribution to the *Controversy* volume constituted a direct rebuttal to Arp's case. Much of it is a detailed analysis of many cases of discordant redshifts reported by Arp. In each case, Bahcall sought to cast doubt on the discordance. For example, he argued (pp. 83–88) in specific cases that the filaments Arp reported may not be there at all, or may be an artifact of the photographic process, or may just be a chance superposition in our view of bodies widely separated in space. On pp. 107–13, Bahcall summarized the then present case for the standard view of quasars as very bright, very distant objects; and then on pp. 113–15 why Arp's case for discordant redshifts among quasars depended on tendentious statistical analysis.

In relation to general matters of inductive inference, the most interesting of Bahcall's complaints was a general criticism of the method employed by Arp to find discordant redshifts. It is a direct attack on Arp's assumption of *Proximity*. All the indicators of sameness of distance Arp reports could arise purely by chance. Bahcall wrote (p. 82):

> The skies when photographed with large telescopes reveal so many individual objects on any photographic plate that one can find almost any configuration one wants if one just hunts: even stars arranged as four-leaf clovers.

The criticism was raised repeatedly. Bahcall recounted the sustained efforts by Arp to establish a physical connection between the compact luminous galaxy M205 and the normal spiral galaxy NGC 4319, with M205 appearing to lie in an outer spiral of NGC 4319. He matched these efforts with the repeated failure of critics to find the connection. He concluded acerbically (p. 89, his emphasis):

> The moral of the story of the apparent connection between NGC 4319 and M 205 is clear: *Seek and ye shall find, but beware of what you find if you have to work very hard to see something you wanted to find.*

Bahcall then laid out the complaint most systematically as (p. 88, his emphasis):

> REMARKS. The way Arp carries out his observational programs, *searching for peculiarities that are not clearly specified before the observations*, actually prevents one from using the argument that a particular observed configuration is too unlikely to be due just to chance. The reader can see easily that such arguments when applied *a posteriori* may be misleading. Suppose you take a detailed large-scale photograph of Times Square that shows at one time in two dimensions all the people in the Square. The *a priori* probability that all the people have by chance the observed angular separations and apparent configurations seen on the photograph is "negligible"; the number of alternative possibilities is very large. However, the *a posteriori* probability for the observed configuration is unity, just because it is the observed configuration!

The admissibility of *Proximity* then depended on whether Bahcall's complaint could be sustained. The decision lay in a duel of conflicting statistical analyses.

Bahcall proceeded to calculate the number of superposition that would arise accidentally given plausible assumptions about galaxy distributions and found them (p. 92) to match near enough the number observed. Bahcall's complaint was supported in turn by three papers he selected for reproduction in the volume, each finding the arrangements Arp judged significant as consistent with purely chance alignments.

For his part, Arp claimed repeatedly that mere chance could not explain the arrangements. For example, he reported (p. 25) a calculation of a chance of less than one in five hundred of a certain significant alignment of seven quasars with relatively nearby peculiar galaxies. Arp could provide his own repertoire of statistical studies showing that mere chance does not suffice. At least five of the papers he had selected for inclusion in the volume were of such studies. Arp concluded his rejoinder to Bahcall with a plea (p. 129)

> I should like to note that Dr. Bahcall says he has estimated what the probabilities are that the associations I have discussed could be due to chance. Obviously there is always some finite chance that any one association could be accidental. If each association is considered separately, each could be dismissed on these grounds. But what is the chance that two or three or a half dozen could be accidental? Since these cases are independent, their improbabilities multiply, yielding in the end an extraordinarily low figure for the probability of chance occurrence. What value of probability would Dr. Bahcall accept as a demonstration establishing the case? Finally, I would like to ask him, seriously, if discordant redshifts do exist, what he would consider as an observation or a set of observations that he would accept as proof.

Each of these dueling studies proceeds from its own set of background assumptions, most importantly on the chance distribution of various astronomical objects. These assumptions, if true, serve to warrant the inductive inferences underpinning the studies. Evaluating the merits of these dueling studies goes well beyond what can be done here and what needs to be done here. As far as the nature of the inductive inferences are concerned, the duel illustrates the centrality of the warranting assumption *Proximity* in Arp's inferences and that their cogency is to be decided by a determination of this assumption's admissibility.

9.10 Differences in Inductive Reach

According to the material theory of induction, background facts warrant inductive inferences. Thus, the reach of inductive inferences depends upon the richness of the pertinent background facts available. Here Bahcall's standard view and Arp's dissident view differed markedly.

Bahcall's standard view conformed with then current cosmological theorizing. This was a fact to which he drew attention when he commenced his summary of the tests passed by the redshift-distance relation (p. 77):

> Also of great importance is the fact that the specific form of Hubble's law, redshift $\propto$ distance, has a very simple theoretical interpretation. Hubble's law is predicted by all cosmologies that assume the universe is expanding and is (at least locally) homogeneous and isotropic.

Bahcall does not spell out why this is important.[14] Mere conformity with further science is important in itself. However, it has another, narrower importance. The conformity enables an inference to a subset of possible cosmological models. Background cosmological models by themselves could not pick among universes

---

[14] On pp. 81–82, he reports that large redshifts are evidence for the expansion of the universe; and that the overall results conform with the laws of physics found terrestrially.

that were contracting, expanding or even static (if the cosmological constant were appropriately tuned). Among expanding universes, a linear relationship is recoverable only for nearby galaxies. The observed redshift-distance relation then is the evidence that enables an inference to those cosmological models that are expanding and with a region of linearity corresponding to the distances of galaxies observable. There were indications in the most recent studies reported by Bahcall that this was the limit of the region of linearity. Sandage (1972, p. 1) was already able to report a non-zero deceleration parameter, which is a parameter that gauges deviations from linearity.

The inferences to this set of cosmological models were warranted by the general cosmological theory. Its truth seemed fairly secure. The theory required only that general relativity, the theory of gravity then accepted locally, also apply cosmically; and that the universe is on the largest scale roughly homogeneous and isotropic.

In comparison, Arp's analysis had limited inductive reach, contrary to his Baconian expectations. At best it could only establish a negative: the failure of the redshift-distance relation. Arp could call upon nothing further in existing theory to assist him in inferring more from his results. Indeed, he had to report the awkward fact that his conclusions did not conform with then present science. He wrote (p. 18):

> The explanation for any such noncosmological redshifts is not readily available from current physics. Because the discordant redshifts are overwhelmingly redshifts and not blue shifts, peculiar Doppler velocities cannot be invoked. This is because in the case of high random velocities as many approach (blue shift) anomalies as recession anomalies should be observed. Gravitational redshifts require too much mass and too abrupt local gradients of field strength to be reconciled with observations of diffuse galaxies, even if complicated models could be made to work for the more compact quasars.

Bahcall's report on this difficulty was more forthright (p. 82): "If discordant redshifts truly exist, then the known laws of physics do not apply to some galaxies."

It is presumably for this reason that Arp sought out new physics that would be compatible with his discordant redshifts. The papers he selected for inclusion in the volume included those by Fred Hoyle that proposed non-standard physics, such as a joint paper by Fred Hoyle and Jayant Narlikar that non-velocity redshifts might derive from matter with electrons of low mass.

9.11 Who Won?

The *Redshift Controversy* volume has no one serving as an umpire to declare who won the debate. The two sides made their best cases and the decision was left to readers in the community. In a simple, sociological sense, it is clear that Bahcall and the standard view won. For Arp's work and his notion of discordant redshifts has all but completely disappeared from the literature after the debate. In this sense, the astronomical and cosmological community has decided.

However, it is also clear that the weight of evidence strongly favors this standard view. There was no communal irrationality in the decision. For, as we saw in the last section, the redshift-distance relation conformed with the standard view in cosmology. Its place became even more secure as new results in cosmology were built around it. In contrast, Arp's dissident view could not be developed without in turn developing alternative cosmology and astrophysics.

Arp recognized this difficulty. His later publications continued to make fitful references to such physics. In his 1987 development of his claims on discordant redshifts, speculation on their explanation in non-standard physics, such as "tired light," is isolated in a few concluding pages (pp. 178–84) at the end of the volume. Arp's (1998) publication was entitled with autobiographical candor, *Seeing Red…* It is a boisterous mixture of new observational results favoring discordant redshifts, anecdotes largely of Arp's perceived mistreatment by mainstream astronomy and fragments of alternative science. The last included Hoyle and Narlikar's Machian-based theory of gravity that leads to lower mass electrons (p. 108); Hoyle, Burbidge and Narlikar's 1993 "quasi-steady state" cosmology (p. 238); and Arp's own non-expanding cosmology (pp. 251–52).

While Arp continued to publish his evidence for discordant redshifts, he was unable to secure its acceptance. Reports continued to contradict his claim. Iovino and Hickson (1997), for example, concluded in their abstract: "Our results confirm that projection effects alone can account for the high incidence of discordant redshifts in compact groups."

Arp also required novel astrophysics. His view required quasars not to be distant, extraordinarily bright objects, but dimmer, closer objects, possibly ejected from nearby galaxies. The standard view of quasars grew from strength to strength. They came to be widely accepted as the brilliant nuclei of active galaxies. The nuclei contain, most likely, a supermassive black hole and massive amounts of radiation are emitted as matter falls into the hole. A volume published in 2012 celebrated 50 years of quasar research. The idea of quasars' great brightness and distance had become so well established that Peterson, writing in the preface, needed to remind readers with astonishment that this idea was ever doubted (D'Onofrio et al. 2012, p. vii):

> But over at least the first two decades of quasar research, the question that dominated discussion was whether quasars were indeed at the cosmological distances implied by their redshifts!

Meanwhile, the sort of alternative accounts of quasars required for Arp's view fared poorly. Tang and Zang (2005) found no evidence in astronomical survey data for certain models of how supposedly nearby quasars could have significantly greater redshifts than their parent galaxies.

Chapter 2 of the fifty-year retrospective volume (D'Onofrio et al. 2012) included reminiscences prompted by interview questions from astronomers involved in the history of quasars. Halton Arp and Jayant Narlikar were included in the section

"2.11 Challenging the Standard Paradigm." A narrator first needed to remind readers that once there had been other views (p. 61):

> We now move to an entirely different view of quasars, and in many ways to a different universe. Not everyone back in the 1960s accepted at once that quasars were "fast and far." Several astronomers suggested that quasars were not at the distances implied by their redshift . . .

Arp (pp. 61–72) reviewed at length the basis of his past and continuing disagreement with the standard view. The weakness of Arp's position became clear when the interview question was "Do you think that some alternative fundamental physics could help explain the quasar phenomenon?" The response was brief and hesitant, mentioning Hoyle and Narlikar's views and suggesting exploration of Bose-Einstein quantum states. The interview concluded with a polite dismissal:

> Thank you, Halton for your considerations. Since the above discussion involves unconventional theories, interested readers may look at the more extended presentation in Halton Arp's book (Arp, 1998).

His views had become so remote from the mainstream that no rebuttal was deemed necessary.

In sum, the standard view thrived, moving from evidential success to evidential success. Arp's view languished. Its basic results remained under challenge. If it was to match the strength of evidential support of standard view, it needed to come up with the alternative science that could make sense of its non-expanding cosmology and its novel astrophysical processes. This was a massive evidential debt that remained undischarged. The longer it remained so, the worse it was for Arp's view.

9.12 Conclusion

The recounting above of the inductive relations connection evidence and theory in the Arp-Bahcall debate may well seem to labor the obvious, at times to the point of tedium. Indeed, I hope that the reader has developed this sense. For the narrative displays the inductive relations of support, while drawing nowhere on general theories of confirmation. There are no enumerative inductions, analogies, inferences to the best explanations or displays of Bayes' theorem. They are not needed to display and assess the cogency of the inductive inferences in the debate. All inductive relations are traced back to their warrant in background facts. We see that a critical goal of each side was to provide further support for their own warranting fact, while at the same time impugning that of the other side. The example shows that, when we begin to explore the details of evidential cases laid out in science in at least this one case, the material approach fits closely with the practice of the science while also giving the means to assess the validity the inferences.

References

Aitken, R. G. (1918). *The binary stars.* (No publisher listed).

Arp, H. (1966a). *Atlas of peculiar galaxies*. California Institute of Technology.

Arp, H. (1966b). Peculiar galaxies and radio sources. *Science, 151*(3715), 1214–1216.

Arp, H. (1987). *Quasars, redshifts, and controversies*. Interstellar Media.

Arp, H. (1998). *Seeing red: Redshifts, cosmology and academic science*. Apeiron.

Arp, H., & Bahcall, J. N. (1973). *The redshift controversy*. W. A. Benjamin, Inc.

D'Onofrio, M., Marziani, P., & Sulentic, J. W. (Eds.). (2012). *Fifty years of Quasars: From early observations and ideas to future research*. Springer.

Hubble, E. (1929). A Relation between distance and radial velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Sciences, 15*, 168–173.

Hubble, E. (1936). *The realm of the nebulae*. Oxford University Press/London:Humphrey Milford.

Humason, M. L., Mayall, N. U., & Sandage, A. R. (1956). Redshifts and magnitudes of extragalactice nebulae. *Astrophysical Journal., 61*, 97–162.

Iovino, A., & Hickson, P. (1997). Discordant redshifts in compact groups. *Monthly Notices of the Royal Astronomical Society, 287*, 21–25.

Kragh, H. S. (2007). *Conceptions of cosmos: From myths to the accelerating universe: A history of cosmology*. Oxford University Press.

Norton, J. D. (2003). A material theory of induction. *Philosophy of Science, 70*, 647–670.

Norton, J. D. (2005). A little survey of induction. In P. Achinstein (Ed.), *Scientific evidence: Philosophical theories and applications* (pp. 9–34). Johns Hopkins University Press.

Norton, J. D. (2021). *The material theory of induction*. BSPSOpen/University of Calgary Press.

Sandage, A. (1972). The redshift-distance relation. II. The hubble diagram and its scatter for first-ranked cluster galaxies. *Astrophysical Journal, 178*, 1–24.

Slipher, V. M. (1912). The radial velocity of the andromeda nebula. *Lowell Observatory Bulletin, No. 58, II*(8), 56–57.

Tang, S. M., & Zhang, S. N. (2005). Critical Examinations of QSO Redshift Periodicities and Associations with Galaxies in Sloan Digital Sky Survey Data. *The Astrophysical Journal, 633*, 41–51.

Weinberg, S. (1972). *Gravitation and cosmology: Principles and applications of the general theory of relativity*. Wiley.

Cristián Soto
Editor

Current Debates in Philosophy of Science

In Honor of Roberto Torretti

Springer

Editor
Cristián Soto
Departamento de Filosofía
Universidad de Chile
Ñuñoa, Chile

Newton International Fellow
British Academy/CPNSS, LSE
London, UK

ISSN 0166-6991 ISSN 2542-8292 (electronic)
Synthese Library
ISBN 978-3-031-32374-4 ISBN 978-3-031-32375-1 (eBook)
https://doi.org/10.1007/978-3-031-32375-1

© Springer Nature Switzerland AG 2023
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland