The Evidence for the Accelerating Universe: Endorsement and Robust Consistency

G. Guralp

San Diego State University

`gguralp@sdsu.edu`

Abstract

The 2011 Nobel Prize in Physics was awarded to researchers from the Supernova Cosmology Project and the High-z Supernova Search Team for the discovery of the accelerating expansion of the universe. In this paper, I provide a historical analysis of the supernova cosmology evidence put forward by these teams for the accelerating universe, in terms of an iterative model of scientific progress developed by Hasok Chang in the context of his study of the development of measurement standards. I argue, using the key concept of *epistemic iteration,* that the iterative model adequately accounts for evidence production in experimental science as well. In order to apply Chang's model to the experimental evidence for the accelerating universe, I introduce the concept of *endorsement* as a particular mode of progress, and argue that supernova scientists produced an endorsed measurement system to claim evidence for their discovery. Furthermore, I show that the credibility of the evidence was not based on a particular measurement, rather, what proved to be decisive was the "robust consistency" of many individual results.

1   Introduction

According to contemporary cosmology, $\sim 68\%$ of our universe is composed of *dark energy*, which is responsible for the late-time acceleration of our universe (Planck Collaboration et al., 2018). Today, there are several lines of evidence that point towards the existence of dark energy, such as the baryon acoustic oscillation measurements, or weak gravitational lensing (Chuang et al., 2013; Abbott et al., 2018). However, the initial evidence that confirmed the existence of dark energy was provided by *supernova cosmology*, which established the accelerated expansion of the universe by making use of comparative distance measurements of both nearby and high-redshift supernovae. These measurements were conducted in late 1990s by two competing research teams. One of these was the Supernova Cosmology Project (SCP), based in Lawrence Berkeley National Laboratory in California, Berkeley. The other team was known as the High-z Supernova Search Team (High-z), which was an international team of astronomers that formed a collaboration to compete with the SCP. In 2011, researchers

from both these teams were awarded a Nobel Prize in Physics for their work on the discovery of the accelerating universe.[1]

In this paper, I analyze the supernova cosmology evidence for the accelerating universe in terms of an iterative model of knowledge production. This model was first proposed by Hasok Chang, in the context of his study of the development of measurement standards in the history of thermometry (Chang, 2004). The central concept in Chang's analysis is *epistemic iteration.* Chang characterizes epistemic iteration as a "process in which successive stages of knowledge, each building on the preceding one, are created in order to enhance the achievement of certain epistemic goals" (Chang, 2004, 45). This process begins by "adopting an existing system of knowledge, with some respect for it but without any firm assurance that it is correct ..." (Chang, 2004, 6). One then conducts scientific research within this adopted system "that result[s] in the refinement and even correction of the original system" (Chang, 2004, 6).

I aim to show, on the basis of a historical analysis of discovery of the accelerating universe, that epistemic iteration not only enables one to understand formation of measurement standards[2] (as Chang originally intended), but can also shed light on the question of how evidence is generated in experimental science. Chang identifies two forms of iteration that drive scientific progress, namely, *enrichment* and *self-correction.* Whereas in enrichment, the initially adopted system is improved (say, by increasing its precision,) in self-correction, the system is partially modified, or changed completely. Even though these two modes of progress of the iterative model elucidate several key aspects of the development of supernova measurements for the accelerating universe, the model still needs to be amended to account for experiments that produce scientific evidence. By studying the argumentative strategies of the supernova scientists, we may generate additional tools that bring about these required amendments. To this end, I will introduce a new mode of iterative progress that I refer to as *endorsement.* Endorsement denotes a form of iterative development through which an experimental system gains *evidential authority.* I will argue that, in the case of the discovery of the accelerating universe, supernova scientists formed an *endorsed* measurement system to produce credible evidence.[3]

During the discovery of the accelerating universe, after an endorsed measurement system was formed, several measurements were made within the endorsed system, and scientists examined whether these measurements gave consistent results. Only after obtaining this consistency, the High-z and the SCP teams claimed to have obtained evidence for the accelerating universe. Thus, it was not a single measurement that established the accelerating universe result, but a consistent conglomerate of many.

While discussing the consistency of their measurement results, supernova scientists made explicit use of the concept of *robustness,* referring in numerous occasions to their results as

---

[1]These scientists were Adam Riess and Brian Schmidt from the High-z, and Saul Perlmutter from the SCP.

[2]For two recent studies that apply the iterative model to standardization in time and sensory measurements, see Tal (2016), and Barwich and Chang (2015), respectively.

[3]I intend the endorsement framework as providing a "descriptive account" (borrowing a term from (Sullivan, 2008)) of how the accelerating universe result was justified by the supernova cosmologists.

being "robust." As is well known, this concept has been studied extensively in the philosophy of science literature. Beginning with the pioneering work of William Wimsatt, considerable effort has been put into clarifying what exactly the justificatory rationale behind robustness is.[4] For Wimsatt, the significance of robustness stemmed from "its implications for increasing *reliability*," which could pertain to a "phenomenon, object, or result [...]" (Wimsatt, 2007, 44, emphasis added). Under the rubric of *robustness analysis*, he offered a recipe of analytical procedures that all relied on introducing variations into either the measurement processes, or theoretical determinations, that sought to identify entities or quantities that remained stable (i.e., reliable) under these variations.

As Wimsatt himself noted, his concept of robustness was "expressed in very general terms," and several authors subsequently introduced finer distinctions within this concept (Wimsatt, 2007, 44). Given that robustness is an actor's category in the discovery of the accelerating universe, it is important to consider these distinctions to decipher how this concept is understood by the practicing scientists themselves, in order to bring about a clarification of its justificatory role in this episode. A very useful starting point is provided by Brett Calcott, who distinguishes three forms of robustness in science, namely, *robust theorems*, *robust phenomena*, and *robust detection* (Calcott, 2011). Robust theorems are formal results that can be arrived at via multiple models. Robust phenomena, on the other hand, are phenomena that withstand perturbations, and thus can be reliably taken to be a "property of the world" (Calcott, 2011, 284). Finally, robust detection characterizes an empirical claim, when "there are multiple, independent ways it can be detected or verified" (Calcott, 2011, 284). This latter concept of robust detection has been most salient in the discussions of experimental practice, and many authors identified it as a prominent feature of experimental science. In recent literature, robust detection is usually characterized in terms of *consistency*, i.e., as multiple independent tests results converging to the same outcome.

As I mentioned above, robust detection, in the sense of several measurement outcomes being consistent, was employed by the supernova scientists. However, as I aim to show, what gave robust detection a justificatory value for the supernova scientists was not the mere consistency of measurement outcomes. Rather, it was the *procedural robustness* of the experimental system that substantiated the robustness claim. This notion of procedural robustness was recently introduced by Koray Karaca, on the basis of a case study on a high-energy physics experiment, as a distinct form of robustness. According to Karaca, robustness of an experimental procedure must be characterized "in terms of the invariance of the procedure's maintenance of its intended function under possible variations in its inputs" (Karaca, 2018, 9). In other words, a robust experimental procedure fulfills its function despite possible variations in the conditions under which it operates. Following this characterization, I will argue that what made the robust detections of the supernova teams trustworthy was that they were the outcomes of robust procedures. These robust procedures, in turn, were constructed through a process of iterative endorsement. It was on this basis of having constructed a *robust procedure*, the teams were able to claim their individual results were

---

[4]For two recent articles discussing this issue, see (Stegenga and Menon, 2017; Schupbach, 2018). A general collection of articles on robustness can be found in (Soler et al., 2012).

robust.

In several of his publications, Robert Hudson objected to the view that robust detection arguments play a significant role in the confirmation of empirical claims (Hudson, 1999, 2009). More recently, in his book *Seeing Things: The Philosophy of Reliable Observation*, he specifically argued that the two teams responsible for the discovery of the accelerating universe did not justify their claims via a robustness argument (Hudson, 2013). In setting up his argument, Hudson relies on Calcott's concept of robust detection. He argues that the mere convergence of "two different physical observational processes" does not provide any "support for the reliability of the processes and for the representational accuracy of the observed results..." by itself (Hudson, 2013, 6). I agree that the SCP and High-z scientists did not provide a robustness argument in the way this concept is characterized by Hudson. However, once the robust detection versus robust procedure distinction is recognized, we see that both a form of robust detection argument was at play, and also that scientists had good reasons for making such an argument.

The plan of the paper is as follows. In Section 2, I begin with laying out the conceptual framework. After introducing Chang's perspective, I describe the idea of endorsement as an extension of the iterative model. I then explicate the claim that confirmation of the accelerating universe result on the basis of experimental evidence was built on the convergence of experimental outcomes in an *endorsed* system of measurement. The remainder of the paper aims to demonstrate this claim through a historical analysis of the confirmation by the High-z and the SCP teams. In Section 3, I first examine the initiation of the supernova cosmology program by two Danish scientists, and then study how the High-z and the SCP teams, following the Danish team, prepared their supernova cosmology measurement system, and presented their initial evidence. In the next section, I consider Hudson's objection that robustness arguments did not play a role in the confirmation of the accelerating universe. In Section 5, I study the publications of the two teams after the initial announcement, and show that the aim of the "post-discovery" papers was to conduct more robust measurements that rule out possible systematic effects that could mimic an accelerating universe. That these robust measurements (understood as robust detections resulting from a robust procedure) pointed to a positive cosmological constant played the central role in validating the accelerating universe result for each individual team. Furthermore, the fact that these robust results obtained by both the High-z and the SCP were consistent with each other was an extra step in the argument supporting the accelerating universe claim. In Section 6, in order to distinguish the robustness claim of a single team, from the convergence of the results of several experimental groups, I introduce the notion of a *robust consistency claim*. As I understand it, a robust consistency claim ensues when several distinct measurement programs, each individually robust, give a consistent result. Finally, in the concluding Section 7, I examine the general applicability of this account to other cases.

2 The Idea of Endorsement and Experimental Evidence

The iterative model sees knowledge growth to occur through stages. In each stage, there exists an affirmed *system of knowledge*, within which scientists can make various measurements. Through the modes of *self-correction* and *enrichment*, scientific knowledge progresses from one stage to the next. Whereas in self-correction, the affirmed system is "altered in its content as a result of inquiry based on itself," in enrichment, "the initially affirmed system is not negated but refined, resulting in the *enhancement* of some of its *epistemic virtues*" (Chang, 2004, 228, emphasis added). In this conception of enrichment, Chang understands epistemic virtues within a *theory–dominated* perspective, and he does not consider how iteration might operate for experimental knowledge. To see this, consider how Chang introduces the notion of an *epistemic virtue.* In his discussion of how epistemic iteration leads to progress, Chang writes that Kuhn "lists accuracy, consistency, scope, simplicity and fruitfulness as the 'values' or 'standard criteria for evaluating the adequacy of a *theory'*" (Chang, 2004, 227, emphasis added). Next, he considers a similar list given by Bas van Fraassen, who "mentions elegance, simplicity, completeness, *unifying power and explanatory power*" (Chang, 2004, 227, emphasis added). Finally, he refers to William Lycan's work that "gives the following as examples of epistemic or *'theoretical' virtues:* simplicity, testability, fertility, neatness, conservativeness, and generality (or explanatory power)" (Chang, 2004, 227, emphasis added). He concludes the discussion by saying that he "will refer to all of these various criteria of judgment as 'epistemic values' or 'epistemic virtues.' " (Chang, 2004, 227). Hence, from his discussion of epistemic virtues, and the specific examples he gives for them following Kuhn, van Fraassen and Lycan, it is clear that Chang has in mind *theoretical* virtues, in constructing his concept of enrichment. However, some of these virtues in the list, such as *unifying power, explanatory power,* or *testability,* do not apply to experiments. For others, it is not clear that they should be considered as "virtues" for an experiment at all, such as *neatness,* or *conservativeness.* Thus, in examining how an experimental system[5] is considered to be capable of making evidential claims, enrichment does not seem to be helpful, and we need a different conception to account for iterative progress that results not in a theoretical refinement but in experimental evidence.

Furthermore, for a given system there are many possible ways of refining it, and enrichment by itself is not specific enough to designate a form of transition in which the end result is the production of *evidence.* In other words, there are situations in which an initially affirmed system is enriched, even though its evidential viability remains the same. For example, extending the scope of the temperature measurements beyond the freezing point of mercury by constructing air and alcohol thermometers is a form of enrichment, but this simply increases the domain of applicability of the instrument (i.e., the thermometer), and does not increase its evidential value. As both air and alcohol thermometers reproduce the results of the mercury thermometers in their overlapping domains of application by their very design, this extension does not augment our confidence in these instruments to a level

---

[5]My usage of the term "experimental system" follows Hans-Jörg Rheinberger's understanding of it. For Rheinberger, experimental systems are "the smallest integral working units of research," which scientists manipulate to generate new knowledge (Rheinberger, 1997, 28).

beyond the mercury thermometers. Rather, in case the extension is successful, we have "reasonable assurance" that our initial conception of temperature can be carried beyond its initial domain (see (Chang, 2004, Chapter 3).)

Consequently, we need a concept of iteration that captures a situation in which new measurements do not simply extend the scope of a concept, but enable an experimental system to gain evidential authority. For measurements conducted in an experimental system that does not represent the target phenomena in an evidentially adequate way, scientists have good reasons not to trust that their measurements are capable on adjudicating on hypotheses in a reliable way. What makes a measurement system evidentially authoritative is not an enhancement of "some of its epistemic [i.e., theoretical] virtues" but a transition in which all its components succeed in fulfilling their experimental task in such a way that the end result is evidentially valuable. Thus, in order to understand how a measurement system becomes evidential, the criterion of the iterative transition should be given from a strictly experimental perspective.[6] In order to build such a criterion, I first need to introduce a distinction within the concept of *system of knowledge.*

In general, the term "system of knowledge" refers to any form of knowledge that is assumed to be valid throughout the course of an inquiry. However, from the perspective of experiment, a system of knowledge may comprise two different types of knowledge, namely, *the knowledge of the measurement system,* and *the knowledge of target phenomena.* The knowledge of the measurement system is, as the name suggests, the knowledge the experimentalists have of the measurement system, which they aim to use in their experiment. I use the term "measurement system" in the most general sense to include all the components that are made use of in the experiment. Thus, in the case of supernova cosmology, not only the instruments such as the CCD cameras, spectrometers, or telescopes are included in the system of measurement, but also the particular experimental strategy, and the natural objects such as supernovae, photons, and galaxy dust as well, as all these items are made use of during the execution of the measurement. On the other hand, the *target phenomena* are the phenomena that the scientists aim to investigate using the measurement system. The knowledge of the target phenomena for supernova cosmologists consisted of their estimations of the total mass in the universe, or, equivalently, the presumed deceleration of the universe, and the family of cosmological models they based their calculations on (to the extent that the model contributed to the scientific characterization of the phenomena).[7] For the ex-

---

[6]To be clear, I do not claim that enrichment or self-correction do not play a role in experimental science. They certainly do. My claim is rather that they are insufficient to account for the transition of a measurement system from a pre-evidential stage to the evidential one.

[7]As one reviewer pointed out, the distinction between "knowledge of the target system" and the "knowledge of the measurement system" may not always be clear-cut. One possible example for this might be the modeling of quantum mechanical measurement processes, in which the measurement apparatus and the target system are treated as a single entangled system. But even in quantum mechanics, I would submit, one starts out with a clear distinction between the apparatus and the system to be measured, and the knowledge that pertain to them. An example for this would be relying on the Schrödinger equation as the correct description of the time development of unobserved particles, while testing for the violations of Bell inequalities. Rheinberger also introduces a similar distinction, between the objects of inquiry that he refers to as "epistemic things," and the "experimental conditions," or "technical objects" within which they are studied.

perimentalist, the relationship between the target phenomena and the measurement system needs to be clearly understood in order for the experiment to be successfully conducted. Furthermore, that this relation actually holds must be empirically established.[8] Hence, in a supernova cosmology measurement, the experimentalist should ensure that the conditions such as the right kind of a CCD camera is used for obtaining sufficient number of galaxies for the search, the correct form of K-corrections[9] are applied to the raw data, the supernova standardization method actually provides the precision needed for the measurement, the search strategy is well optimized to deliver sufficient objects in reasonable time etc. I will use the term "epistemic access" to refer to this type of situation: If the set of conditions of the experiment that are delineated by the experimentalist are satisfied, then the target phenomena can be said to be epistemically accessible by the measurement system.[10] On the basis of this idea, I define endorsement as follows:

> **Endorsement**: A system of measurement is endorsed when the target phenomena is judged to be epistemically accessible within the existing (system of knowledge of the) measurement system.[11]

In order to judge that the target phenomena lies within the epistemic purview of the measurement system, scientists need to provide *good empirical reasons.* These reasons are provided on the basis of various epistemic tools including data models, Monte Carlo simulations, error analysis, and previously conducted measurements. Once all the checks are completed, *one endorses the stage of knowledge that one is in as evidentially authoritative.*[12] As I aim to show in the next section, the discovery of the accelerating universe shows the need for

---

Even though he recognizes that this distinction is contextual, in that a "sufficiently stabilized" epistemic object can function as a technical object within an experimental system, he still submits that "their distinctness is clearly perceived in scientific practice" (Rheinberger, 1997, 29,30). In line with Rheinberger, my claim is not that this distinction is structurally fixed, but rather in plenty of cases of experimental research, we see that the practitioners do employ a form of it.

[8]Hence, this is a two step procedure: The right conditions for the experiment cannot be known *a priori,* but require many iterative corrections to be found out. Once the experimental conditions are established, actually implementing them requires further iterative trials

[9]The K-correction is a relativistic term that needs to be calculated in order to "compare the magnitudes and light curves of objects of different redshifts" (Hamuy et al., 1993b, 787). Even though the K-correction is fundamentally a technical issue with a straightforward interpretation, it is an essential part of the formalism of supernova cosmology.

[10]Note that, this set of conditions need not be unique. Thus, two experimental teams working on the same measurement may delineate different sets of accessibility conditions depending on the particular experimental system that they work with. On the other hand, a single team may also choose to vary the specific conditions in order to make their access optimal. For example, if the standardization method is not trustworthy, one can opt for a different search strategy to increase the number of detected supernova, or, if one cannot sufficiently reduce the biases due to the K-correction, one can chose to use additional filters for color.

[11]To repeat, what is at stake in endorsement is not an enhancement of a theoretical virtue such as scope or precision, but the establishment of an experimental system that satisfies the accessibility requirements of the measurement for the target phenomena. Although the construction of the measurement system requires many iterative enrichments, endorsement need not involve any enrichment, as it essentially involves coordinating the different components of the system in such a way that the target becomes accessible.

[12]I use the term "authority" due its normative connotation, following Robert Brandom. In his *Making*

introducing the concept of endorsement. When experimenters decide to make a measurement, they cannot, and do not, *just go and do it.* Before the experiment can be done, there are many preliminary steps that involve planning and design. All this is done in an affirmed system of knowledge. What is distinctive (and instructive) of supernova cosmology is that, when the project of using supernovae measurements for cosmological purposes was first conceived, all scientists agreed that the existing system of knowledge was not yet capable of realizing this goal. In other words, the target phenomena was beyond the reach of the existing knowledge of measurement system. Using the terminology introduced above, one can also say that the target phenomena laid beyond the *epistemic access* of supernovae cosmology. Not enough was known about the nature, and the standard candle properties of supernovae, and the technological knowledge for making better instruments did not exist as well. Similarly, the strategy to systematically detect supernovae was not in place either. Before all these steps were taken, it was not possible to obtain any experimental results that would count as evidential.

This decision of endorsement is an empirical one, taking into consideration all the aspects of the measurement system, including the data collection and analysis procedures, the precision of instruments, the ability of the scientists and so on. Scientists need to give empirical arguments to the effect that what they aim to measure is within the epistemic access of their measurement system. As I show below, during the years 1997-1998, the supernova system of measurement was endorsed. In other words, both the SCP and the High-z scientists were convinced that their investigations brought their respective measurement systems to a level at which the system's pronouncements concerning the target phenomena would *constitute* experimental evidence. I will refer to an endorsed experimental system as a *measurement protocol.* By the end of 1998, the two teams succeeded in constructing measurement protocols that produced evidence for the accelerating universe.

The framework of endorsement differs from the alternative accounts of experimental evidence in that I situate the issue within the iterative formation of the measurement protocol, as opposed to focusing the analysis on the results obtained with an already existing one. Among the various prominent accounts of evidence in experimental practice, Allan Franklin argues that a collection of *strategies* that scientists employ, which he refers to as the "epistemology of experiment," provides "reasonable belief in experimental results" (Franklin, 2002, 2). Deborah Mayo urges a statistical framework, according to which a body of data constitutes evidence for a hypothesis H "(just) to the extent that test T severely passes H with [the data]" (Mayo, 2005, 100). A further approach is provided by James Woodward, who argues that for data to constitute evidence for an experimental claim, experimentalists

---

it Explicit, Brandom proposes to explain authority of norms over a community as due to the fact that rules that members of a (rational) community endorse become binding for them. As he writes, "our own *acknowledgement* or *endorsement* of a rule is the source of its authority over us. . . " (Brandom, 1998, 51). In the context of experimental knowledge, what I referred to as the necessary conditions for the execution of the measurement can be seen as *norms* that are delineated by the scientists. Once the experimental system is judged to obey these norms, i.e., the requirements that the experimentalists delineate in the course of their inquiry are verified to be met by the experimental system, the resulting system is endorsed as evidential, and thereby gains authority for the scientific community.

must establish a "systematic pattern of counterfactual dependence" between the data and the claim (Woodward, 2000, S166). Even though all these authors acknowledge the importance of a competently formed experimental system, their analyses of evidential relationships assume the existence of an experimental system that produces trustworthy data as given. The framework of endorsement aims to be complementary to these accounts, as it purports to demonstrate the evidential relevance of the very construction of a reliable experimental system.[13]

Once a measurement protocol is constructed, the iterative progress of the *knowledge of the measurement system* is paused, at least momentarily. For the focus of scientists is now turned towards executing the measurement, instead of improving the measurement system. In this case, a shift in the iterative process occurs. As now the primary aim is learning about the target phenomena, the affirmed knowledge that is subject to iteration is the existing *knowledge of the target phenomena*, in our case, the existing cosmological model.

In the case of the accelerating universe, scientists conducted many experiments before validating the experimental result, and these experiments followed an iterative pattern as well. In this sense, epistemic iteration not only plays a central role in building measurement standards, but is also essential in conducting "high-level" experiments for confirmatory purposes. As the measurement protocol is established, scientists check to see whether experiments conducted within the protocol produce *robust* results. In other words, for each team, the results should persist under different data sets, instruments, or methods of analysis. In addition, that two separate measurement protocols, each with individually robust results, agreed on the accelerating universe, played a significant role in its validation for the community of astronomers.[14] Thus, the confirmation of the accelerating universe was established through a consistency of the two teams' individually robust measurements. In the next section, I will illustrate these ideas through an analysis of the discovery of the accelerating universe by supernova cosmology.

---

[13]In a recent paper proposing an iterative account of data processing, Sabina Leonelli reaches a similar conclusion concerning how data constitutes evidence. Using the concept of *evidential space*, understood as the "range of phenomena" for which a given body of "data could plausibly serve as evidence," she writes: "The very process of defining what constitutes acceptable evidence helps researchers to narrow their investigative focus to specific phenomena of interest within that target system ...Procedures of data production and processing thus define the evidential space ..." (Leonelli, 2019, 22). The concept of endorsement aims to capture the normative mechanics of this very process of "defining the evidential space" within empirical practice.

[14]Royal Swedish Academy of Sciences attested to this fact in the text that accompanied the Nobel Prize announcement, providing the scientific background to the discovery of the accelerating universe: "The fact that both groups independently presented similar - albeit extraordinary - results *was a crucial aspect for their acceptance within the physics and astronomy community*" (NobelPrize.org, 10, emphasis added).

3   Constructing the Measurement Protocol

3.1   The Danish–Durham Team

The first research program that aimed at using *high redshift* supernovae for cosmological purposes was a collaboration of Danish and British astronomers in the mid 1980s. The impetus of their research program came from a talk that Gustav Tammann gave at a conference dedicated to the astronomical uses of the *Space Telescope*[15] (Tammann, 1979). In this talk, Tammann focused on the following four possible uses of the telescope: the determinations of the Hubble parameter (H_0), and the deceleration parameter (q_0)[16], as well as using the telescope for a detailed mapping of the local expansion field, and understanding the nature of the galactic redshifts. With respect to q_0, Tammann argued that Type I supernovae are the best objects to use for the determination of this parameter, due to their *standard candle* property[17]:

> Present data suggest that SNe of Type I in E (and S0) galaxies are at maximum light *very good standard candles.* The observed magnitude dispersion amounts to $\Delta m_{pg} = 0^{m}\!.4$,[18] but the true dispersion is certainly smaller and possibly vanishingly small ... SNe I in other types of galaxies suffer from internal absorption, as do SNe II, and they exhibit therefore a much wider luminosity scatter at maximum... (Tammann, 1979, 335, emphasis added).

Hence, Tammann's claim that supernova are standard candles involved *Type I supernovae found only in a restricted class of galaxies.* He compared supernovae with other astronomical objects which could be used as standard candles at high redshifts, such as galaxy clusters, from various perspectives, including the existence of evolutionary effects,[19] difficulties in photometry, or the Malmquist bias.[20] For Tammann, the main advantage of supernovae was that they were expected to be free of these problems. For example, concerning evolution, he wrote that "the explosion of a Type I SN, whatever its origin may be, seems to be a physically well defined event, which is not expected to vary with time" (Tammann, 1979, 335). As supernovae are less bright than galaxy clusters, and hence more difficult to detect,

---

[15]Until its re-naming in 1983, the Hubble Space Telescope was referred to as the "Space Telescope." For an early history of its construction, written by one of the pioneers of the project, see (Spitzer, 1979).

[16]Before the discovery of dark energy, the majority of cosmologists believed that the expansion of the universe was slowing down, due to the attractive nature of gravity. The deceleration parameter (q_0) quantified the rate of this slowing down of the universe.

[17]Initially, supernovae were thought to be two types, and it was the Type I supernovae that were the considered to be the best standard candle candidates. In around 1985, Type I supernovae were further classified into Type Ia and Ib, and it was understood that the Type Ia's were the homogeneous objects (relative to Ib's) to serve as potential standard candles.

[18]m_{pg} stands for the apparent photographic magnitude.

[19]The term "evolution" denotes the possible difference between supernova in the early universe (high-redshift), and more recent (low redshift) ones.

[20]Malmquist bias is a selection effect that stems from the fact that astronomical observations tend to detect intrinsically bright objects compared to faint ones. It was first described by Karl Gunnar Malmquist in 1922 (Malmquist, 1922).

they were the ideal objects to study with the Space Telescope. In addition, Tammann believed that the faintness of supernovae meant that the space telescope would not simply observe these objects, but also would be used to *search* them. He therefore delved into questions concerning where the objects could be found with the highest rate, and what the search strategy should be. He calculated that to find 25 objects, one would need "4 days of telescope time," adding that this "is indeed a modest price for the goals…" (Tammann, 1979, 337).

Referring to Allan Sandage's classic 1961 article "The Ability of the 200-INCH Telescope to Discriminate Between Selected World Models" (Sandage, 1961), Tammann then estimated that only 6 supernovae, discovered at the redshift of $z = 0.5$, would be enough to discriminate between the flat universe model ($q_0 = 0$), and the open universe model ($q_0 = \frac{1}{2}$). Finally, he added that, for the determination of q_0, a local set of supernovae "at small and intermediate redshifts with uniform photometry" were required, in addition to the high redshift ones that would be discovered and studied by the Space Telescope (Tammann, 1979, 338). In anticipation of the low redshift search that Calán/Tololo survey realized in 1990s, he explained:

> The search for these additional SNe I in E/S0 galaxies[21] shall turn out to be quite time-consuming, regardless if performed from the ground or with ST. *The details of an optimum observing programme are still to be devised* (Tammann, 1979, 338, emphasis added).

Tammann concluded his talk with an optimistic outlook, writing: "The conclusion is that ST shall be helpful for the determination of H_0, that it shall be very important for the mapping of the expansion field, and that it offers a unique chance to determine q_0 from SNe and to verify the Doppler nature of cosmological redshifts" (Tammann, 1979, 341).

At least two people in the audience shared his optimism. While listening to Tammann, Leif Hansen and Hans Ulrik Nørgaard-Nielsen, two early career researchers from Denmark, conceived of a possible project to embark on: determination of the deceleration parameter using high redshift supernovae. The Danish scientists, due to their affiliation with the Copenhagen University Observatory, had access to the Danish 1.5 m. telescope at La Silla, Chile. This telescope was built in 1978, and became operational one year later. The Danes understood that, contrary to Tammann, finding the supernovae with the Space Telescope was too "expensive." In their first paper describing their program, they explained that a ground-based telescope can be used to find supernovae. As they wrote:

> With the launch of the Hubble Space Telescope (HST) it will become possible to do photometry on distant SNe to magnitudes fainter than 25, and the cosmological goal is then within reach. *The first and difficult problem is to find the SNe [Type] I.* G. A. Tammann…estimates that a Coma-like cluster at z = 0.5 will show a rate of 0.5 SN I per year within the field of the Wide Field Camera of the

---

[21]In Hubble's classification of galaxies, the letter "E" denotes elliptical, and the letter "S" denotes spiral galaxies. "S0" denotes an intermediate type between the E and S classes.

HST. However, observing time on the HST is very expensive, and fortunately the job can be performed from the ground. The Danish 1.5-m telescope at La Silla is ideal for the task (Hansen et al., 1987, 46, emphasis added).

The finding strategy of the group was as follows: First, in order to increase the chance of finding a supernova, they observed "rich clusters of galaxies with redshifts from 0.2 to 0.6 if possible each month from September through April with the Danish 1.5-m telescope" (Hansen et al., 1987, 46). Then this exposure is compared with a previously obtained "standard" exposure of the same field.[22] As a result of this comparison, one obtains a *difference image* by subtracting the standard image from the new one. One then analyzes the difference image to identify a supernova explosion. The group used a "special colour scale" to emphasize the differences in brightness, such that "[p]ixel-values deviating more than about 2.5 σ will appear either black or red" (Hansen et al., 1987, 47). As a result of this, a supernova explosion "will stand out very prominently in the difference as black or red" (Hansen et al., 1987, 47). Once the identification of the object is made, it needs to be confirmed by ruling out cosmic rays, or instrumental effects. The group seems to have planned to confirm the object through spectroscopy within the very same run, as they warn that "a delay of one day" in identifying an object "may have sad consequences if a candidate appears on the last night of observing" (Hansen et al., 1987, 48).

The Danish scientists expected to find, on the basis of the supernovae rate studies by Tammann, 5 objects per year (assuming one search run every month). Yet, after their first 5 runs, which should have produced 2 objects, they found none. They provided several reasons for the existence of the discrepancy:

Part of the answer may be that although many of the clusters are really impressive some are less rich than Coma.[23] One should also notice that the local supernova-rate is uncertain with a factor of 2 according to Tammann (private communication). Further, there is at present no evidence of the rate at earlier epochs of the universe (Hansen et al., 1987, 49).

The last sentence is important: there was no evidence at the time concerning whether supernovae *existed* in the early universe. This indicates the *explorative* nature of the experiment the Danish scientists conducted: they were not testing a hypothesis with a controlled experimental procedure, but were exploring an uncharted territory in the universe.[24] This

---

[22]For the comparison images of clusters, the Danish scientists contacted two astronomers at Durham University, namely Richard Ellis and his post-doc, Warrick Couch. These scientists had obtained a big catalogue of clusters for a red-shift survey they were conducting. After Ellis offered to collaborate on the supernova cosmology project, the scientists agreed to "take it in turns to go to Chile," and image the clusters every month (Richard Ellis, oral history interview with the author).

[23]Coma is a large cluster of galaxies. The idea here is that as the scientists used Tammann's calculation of supernova occurrence rate, which was based on the assumption that the clusters they observe resemble Coma, the failure of this assumption may be one reason why the expected number of supernovae was not observed.

[24]The notion of "exploratory experimentation" was independently introduced by Richard Burian and

exploratory aspect of supernovae cosmology proved to be a continuing theme of the research in supernova cosmology, showing that this research could not apply an existing system of knowledge to the measurement, but it had to construct the system from scratch.

The group found their one—and only—Type Ia object on 8/9 August, 1988, at the end of a 2 year search. The object discovered was at redshift $z = 0.31$, and it was the highest redshift supernova found up until that date. A spectroscopic identification was made on August 19 (Norgaard-Nielsen et al., 1989, 524). By observing the spectrum of the supernova AC118[25], scientists inferred that it was discovered 17 days after reaching its maximum brightness. This affected the precision in the determination of the value of maximum brightness, as several assumptions had to be made to obtain this number. Using this data, the group published a very tentative result on the q_0.

> Although the AC118 supernova has behaved in all respects like a typical nearby supernova, we cannot yet judge whether it is a representative standard candle for our purposes, and hence a formal result for q_0, however uncertain, is premature. The implied bounds, $-0.6 \leq q_0 \leq +2.5$ ... are, however, sensible, considering the fact that maximum light was not observed directly (Norgaard-Nielsen et al., 1989, 524).

This measurement was the first determination of q_0 using supernovae in history. But the main success of the paper was not the particular number they measured. As the authors wrote in their concluding paragraph, it was that they "demonstrated that it is possible to extend the study of supernovae to cosmologically interesting distances ($z \leq 0.5$)" (Norgaard-Nielsen et al., 1989, 525). Moreover, the group was aware that the work that had to be done was more than simply finding more objects:

> By extending our survey to a cluster sample with $\bar{z} \approx 0.5$ we will need to detect ~ 15 supernovae to determine q_0 with an accuracy of 0.1, in the best possible case (that is, *if the type Ia supernovae are completely well behaved objects with a cosmic dispersion of ~ 0.2 mag.*) (Norgaard-Nielsen et al., 1989, 525, emphasis added).

The question whether supernovae are *well behaved objects* was to prove decisive for the developments of supernova cosmology, as it was this question that carried supernova cosmology into its next stage.[26]

Friedrich Steinle as a type of experimentation distinct from a theory-driven one. (Steinle, 1997; Burian, 1997). For further examination of this concept in the context of cases taken from biology, and particle physics, see (Franklin, 2005), and (Karaca, 2013), respectively.

[25]Being found in the galaxy cluster AC118, the supernova was referred to by the same name.

[26]Even though Ellis found their result encouraging for continuing the project, to his disappointment, the Danish scientists disagreed. Nørgaard-Nielsen and his colleagues thought that the project was too demanding to continue, and showing that supernova cosmology was in principle a feasible endeavor was satisfactory enough for them. As it was "their telescope," and "they got the observing time," it was not possible for Ellis and Couch to continue without the Danish scientists (Richard Ellis, oral history interview with the author). The Danish-Durham supernova project ended in 1989.

In 1979, when the Danish team first conceived of their project, and began conducting their research, there were four significant assumptions that they made, which constituted their *affirmed system of knowledge.* First, the scientists assumed that supernovae were standard candles. Second, they assumed that there were two types of supernovae (Type 1 and Type 2) without any sub-divisions. Third, they assumed Tammann's supernova rate calculation for E/S0 type galaxies was correct. Finally, they assumed that the deceleration parameter (q_0) was positive. These assumptions reflect the general convictions of the supernova community of the era, and we now know that all of these assumptions are incorrect. In fact, when the team published their discovery of AC118 in 1988, the Type Ia-b distinction had been established. Still, on the basis of these three assumptions, the achievement of this earliest stage of supernova cosmology was to show that high-redshift supernovae do exist, and that these objects could be found with the existing technology, provided a suitable search strategy was devised. This not only taught scientists something new about the universe, namely, the behavior of supernovae explosions in the early universe, but it also showed the experimental capabilities of CCD astronomy, which was to prove decisive for the discovery later on. The work of the Danish–Durham team showed that the existing supernova measurement system was not yet capable of delivering the evidence it had sought. Still, even though it proved to be extremely difficult—only one object in two years as opposed to a projected 10—a research program that began with three false assumptions led to progress in science, as specified in the iterative model.

3.2 The First Measurement Paper by SCP and an Erroneous Result

Nine years after the work of the Danish team, SCP published the first paper in supernova cosmology that contained a measurement of the cosmological parameters Ω_M and Ω_Λ.[27] This paper, which in fact reported an incorrect result, is best characterized as a transition paper towards the construction of the full supernova cosmology measurement protocol. Here, I would like to focus on the overall structure of this first evidence paper, and highlight its transitory function. How did the SCP construct the argument for the results they reported?

The paper, entitled "Measurements of the Cosmological Parameters Ω and Λ from the First Seven Supernovae at $z \geq 0.35$," begins with a brief history of the standard candle approach to cosmological measurements, and discusses the advantages and disadvantages of using supernovae as cosmological tools. This discussion is important because it shows how the SCP distinguished their experimental situation from the previous attempts in using supernova as cosmological tools. They explain a major difficulty that hampered earlier work as follows:

> The disadvantages of using supernovae are . . . obvious: they are rare, transient events that occur at unpredictable times and are therefore unlikely candidates for

---

[27]Here Ω_M denotes the mass density of the universe, and Ω_Λ denotes the density parameter of the cosmological constant.

the scheduled observations necessary on the largest telescopes. The single previously identified high-redshift (z = 0.31) SN Ia ... was found *(at an unpredictable time) several weeks after it had already passed its peak luminosity* ... (Perlmutter et al., 1997, 565, emphasis added).

After explicating the difficulty of obtaining quality data in this way, the paper then explains how this problem is overcome with a new strategy of observation that enabled the SCP to discover many objects in a comparably shorter time span. As they write:

> To make high-redshift supernovae a more practical "cosmological tool," the Supernova Cosmology Project has developed a technique over the past several years that allows the discovery of high-redshift SN Ia's in groups of 10 or more at one time (Perlmutter et al. 1997a). These "batch" discoveries are scheduled for a particular night, or nights, which thus also allows follow-up spectroscopy and photometry on the large aperture telescopes to be scheduled. Moreover, the supernova discoveries are generally selected to be on the rising part of the light curves and can be chosen to occur just before new moon for optimal observing conditions at maximum light (Perlmutter et al., 1997, 565-566).

The ability to find supernovae in numbers that are sufficient for cosmological purposes was a *major advance* in supernova cosmology, and thus it was crucial for the group to announce that the measurement protocol they aimed to construct was capable of delivering more objects in principle. The fact that most of these objects can be found before they reached maximum brightness increased the confidence in the determination of the peak brightness values. These points are concretely made as follows:

> Since our demonstration of this technique with the discovery of SN 1992bi at z=0.458 (Perlmutter et al. 1994, 1995a), we have now discovered more than 28 supernovae, most in two batches of ~ 10 (Perlmutter et al. 1995b, 1997b). Almost all are SN Ia's detected before maximum light in the redshift range z= 0.35-0.65. We have followed all of these supernovae with photometry and almost all with spectroscopy, usually at the Keck 10m. telescope (Perlmutter et al., 1997, 566).

Thus, the group provided evidence for both the sufficiency and the quality of the data delivered by their measurement protocol design that aimed to establish supernovae as feasible "cosmological tools."

After discovering and classifying the supernovae, the aim was to construct the light curve of each object so as to measure the maximum brightness it reaches during the explosion. The first step in obtaining the data points that are used toward construction of the light curve is known as the *photometry reduction* in which, the scientists translate the image(s) obtained on a given night to a *number* that represents the flux, or brightness of the object. This translation procedure requires "cleaning up" the raw data image by subtracting the host galaxy light contribution. Here, another important point concerning the construction of the

measurement protocol emerges: an image is rejected if an error of 20% or more is present at the end of host galaxy subtraction.[28] The paper then adds a parenthetical remark:

> (Planned HST[29] light-curve measurements will make these steps and checks practically unnecessary, since the HST point-spread function is quite stable over time and small enough that the host galaxy will not contribute substantial light to the supernova measurement) (Perlmutter et al., 1997, 570).

In the formation of a measurement protocol, instrumental upgrading is a key element for controlling error: the future research that SCP planned involved the use of the Hubble Space Telescope that the scientists indicated would largely increase the evidential strength of the results by reducing the error introduced through photometric corrections due to the host galaxy light.

After obtaining the photometry data points (including the uncertainties), the next step is to fit them to obtain a light-curve. The light curve is fit to standardize the objects via a template.[30] The fitting process involves two further correction terms. These are the $K-$correction, and the correction required due to the absorption of supernova light by our own galaxy, known as *extinction*. Once these corrections are added, the final peak brightness can be constructed. The peak brightness thus obtained for each supernova is the modeled data, and these are compared with the theoretical predictions of various cosmological models to find the best fit model. This comparison is made by fitting the data to various cosmological models by varying cosmological parameters.[31]

In making the main claim of their paper, the SCP essentially relied on a contour plot, which showed the confidence regions for the cosmological parameters depicting the space of cosmological models (Fig. 1). Applying a flat universe prior, SCP discussed the cosmological constant issue as follows:

> The two special cases represented by the solid lines of [Fig. 1] *yield significant measurements:* We ... fit ... the high-redshift supernova magnitudes ... with the fit constrained first to a $\Lambda = 0$ cosmology (the horizontal line of [Fig. 1]) and then to a flat universe (the diagonal line of [Fig. 1], with $\Omega_{\text{total}} \equiv \Omega_M + \Omega_\Lambda = 1$). In the case of a $\Lambda = 0$ cosmology, we find the mass density of the universe to be $\Omega_M = 0.88^{+0.69}_{-0.60}$. For a flat universe (the diagonal line of [Fig. 1], with $\Omega_{\text{total}} \equiv \Omega_M + \Omega_\Lambda = 1$), *we find the cosmological constant to be* $\Omega_\Lambda = 0.06^{+0.28}_{-0.34}$,

---

[28]The error introduced by the subtraction process can be measured by applying it galaxies which contain no supernova. In this case, the subtraction of the galaxy light should give a flat light-curve. Deviations from this ideal case show the scientists how much error is introduced.

[29]Hubble Space Telescope.

[30]The raw supernova data for each object must be "standardized" since not all supernova have exactly the same brightness. This is a technically important point but further explication is beyond the scope of this paper. The method SCP used for supernova standardization is known as the "stretch" method, as this involved stretching the width of a template light curve to fit individual objects.

[31]In other words, in supernova cosmology, each value of the parameters $(\Omega_M, \Omega_\Lambda)$ correspond to a different model of the universe. Thus, in (Fig. 1) each point in the graph corresponds to a model of the universe.

Figure 1: SCP Contour Plot for the Cosmological Parameters Ω_M and Ω_Λ. Source: (Perlmutter et al., 1997, 577). Reproduced by permission of the AAS.
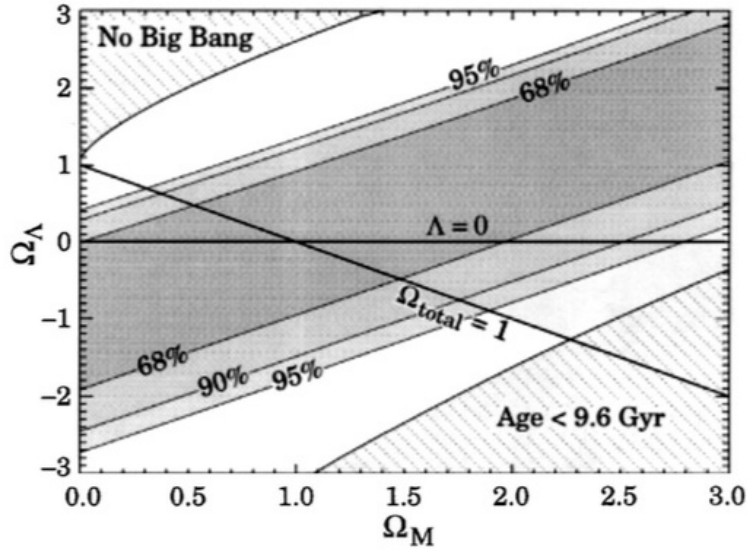
consistent with no cosmological constant (Perlmutter et al., 1997, 575, emphasis added).

Thus, the result of the paper was that there is no cosmological constant, and for a flat universe, $\Omega_M \approx 1$. This result was obtained for the 5 objects with the above mentioned stretch standardization applied, yet the group also analyzed the same 5 objects without applying this correction and showed that this did not change the result in a meaningful way. Finally, the full set (7) objects are analyzed without the correction, with the result $\Omega_\Lambda = 0.15^{+0.24}_{-0.28}$, for a flat universe, still consistent with a ($\Lambda = 0$) universe. This approach of using both the corrected and "measured" data is justified in terms of *robustness*. Thus, concerning the correction, the authors say:

> To make our results *robust* with respect to this correction, we have analyzed the data (1) as measured, i.e., with no correction for the width-luminosity relation[32] ... and (2) with the correction ... (Perlmutter et al., 1997, 567, emphasis added).

The concluding discussion of the paper also invokes a similar robustness argument by mentioning that "supernova data sets provide enough detailed information ... that we can perform tests for many possible sources of systematic error by comparing results for supernova subsets" (Perlmutter et al., 1997, 579). As a result of this comparison, one can examine whether a result is sustained, and to what extent, when different subsets of the entire data set is used in the analysis. In both these cases, the term "robustness" is used in Wimsatt's

---

[32]i.e., the stretching of the template.

sense of introducing variations into the data analysis to check whether the result remains the same.

Furthermore, the continuing aspect of the research is emphasized. In other words, the current result is not presented as the "conclusive" measurement, but a tentative one: "the analysis of our next set of high-redshift supernovae will test and refine these results" (Perlmutter et al., 1997, 580). Hence, the expectation of the SCP scientists was that the next data set was not going to bring in a major revision in their measurement, but would make the result more precise. Thus, in their discussion of the Ω_Λ measurement, they write:

> The uncertainty on the Ω_Λ measurement from the high-redshift supernovae . . . is likely to narrow by $3^{1/2}$ as we reduce the data from the next ~ 18 high-redshift supernovae that we are now observing (Perlmutter et al., 1997, 580).

However, as their next paper showed, further data brought an unexpected result: the possibility of a non-zero Λ.

## 3.3 The *Nature* Paper and the Iterative Correction of Error

In January 1998, SCP published a new result in *Nature* that revised their early claim concerning Ω_M being approximately 1. The argument of the paper is based on a single supernova, namely, SN1997ap, which was the "most distant spectroscopically confirmed" object at the time (Perlmutter et al., 1998, 51). The central claim of the paper was put very carefully. There was no announcement of a cosmological constant, but simply that Ω_M could be less than 1. More specifically, the abstract of the paper stated the conclusion as follows:

> When combined with previous measurements of nearer supernovae, these new measurements suggest that *we may live in a low-mass-density* universe (Perlmutter et al., 1998, 51).

The "new measurements" that are referred to here are the photometry and spectrum observations of SN1997ap. Before presenting the main conclusion of the paper, several points are made that are worth considering. First, photometry observations of SN1997ap were made by the Hubble Space Telescope (HST), in addition to ground based telescopes, as was promised in the earlier paper. Second, the group became more confident in the supernova standardization: "The width of the light curve of a type Ia supernova has been shown to be an excellent indicator of its intrinsic luminosity, both at low redshift and at high redshift..." (Perlmutter et al., 1998, 51). Third, the group recognized the problem of host galaxy extinction. Yet, on the basis of a comparison of the light curve of SN1997ap with a " 'normal' low-redshift type Ia supernovae," they argued that no host galaxy extinction seems to be present. Note that each of these points constitute iterations of earlier analyses with better data, or instrumentation.

The cosmological analysis of the paper followed that of (Perlmutter et al., 1997). The object SN1997ap is added to the previous five objects for the fit, all of them corrected for luminosity. This fit of the combined set

Figure 2: SCP Contour Plot with SN 1997ap. The confidence regions for supernovae at redshift 0.4 and 0.83 are shown in blue and yellow, respectively. Source: (Perlmutter et al., 1998, 53). Reprinted by permission from Springer: *Nature*, © (1998).
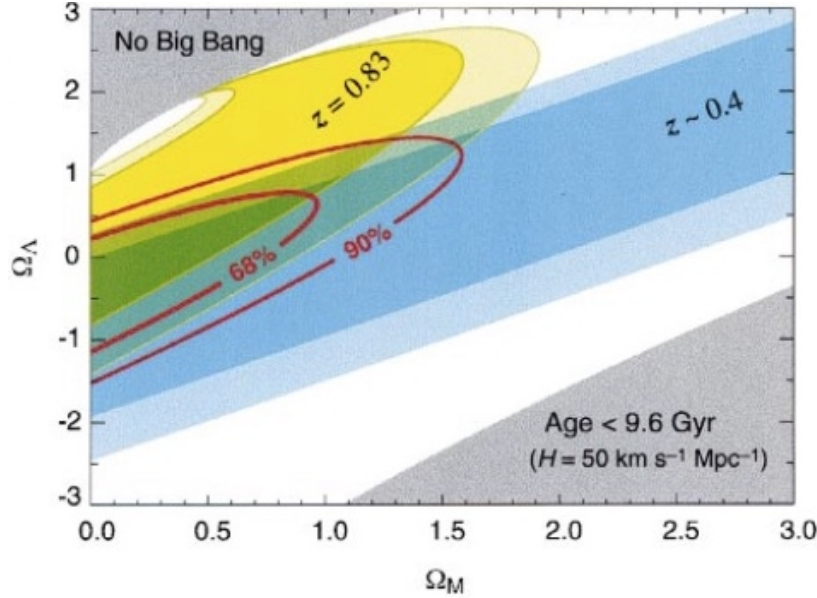
corresponds to a value of $\Omega_M = 0.6\pm0.2$ if we constrain the result to a flat universe ($\Omega_M + \Omega_\Lambda = 1$), or $\Omega_M = 0.2 \pm 0.4$ if we constrain the result to a $\Lambda = 0$ universe. *These results are preliminary evidence for a relatively low-mass-density universe.* The addition of SN1997ap to the previous sample of lower-redshift supernovae decreases the best-fit M by approximately 1 standard deviation compared to the earlier results (Perlmutter et al., 1998, 53, emphasis added).

Thus, with one very high redshift object, SCP revised their previous conclusion and claimed that the early result of $\Omega_M \approx 1$ may be false.

This iterative revision[33] of the earlier claim concerning the mass density of the universe is followed by a passage that is of key importance for the thesis of this paper. After declaring that there is now *preliminary evidence* that the mass density of the universe may be lower than the previous estimate, the paper observes why this measurement is important as follows:

Our data for SN1997ap demonstrate: (1) that type Ia supernovae at $z > 0,8$ exist; (2) that they can be compared spectroscopically with nearby supernovae to determine supernova ages and luminosities and check for indications of supernova evolution; and (3) that calibrated peak magnitudes with precision better than the intrinsic dispersion of type Ia supernovae can be obtained at these high redshifts. *The width of the confidence regions in Fig. 4 [Fig. 2] and the size of*

---

[33] As the leader of the SCP, Saul Perlmutter, put to me in an oral history interview: "What it felt like was constant iteration and revision, so there was learning from error all the time."

the corresponding projected measurement uncertainties show that with additional type Ia supernovae having data of quality comparable to that of SN1997ap, a simultaneous measurement of Ω_Λ and Ω_M is now possible (Perlmutter et al., 1998, 51, emphasis added).

This passage can be read as the precise moment of the *endorsement* of the empirical program of supernovae cosmology, in the SCP context: Ω_Λ and Ω_M were now epistemically accessible to supernova cosmology. Three main components of the measurement protocol are identified, and an empirical argument is given as to how these components satisfy the epistemic requirements for a successful measurement of cosmological parameters: First, the SN1997ap proves that supernovae at redshifts as high as $z = 0.83$ do exist, and can be found. The significance of this fact is clearly articulated in the paper. For, in the redshift of SN1997ap,

a flat $\Omega_M = 1$ universe is separated from a flat $\Omega_M = 0.1$ universe by almost one magnitude, as opposed to half a magnitude at $z \approx 0.4$ . . . Thus, at such redshifts even individual *type Ia supernovae become powerful tools* for discriminating amongst various world models, provided observations are obtained, such as those presented here, where the photometric errors are below the intrinsic dispersion of type Ia supernova (Perlmutter et al., 1998, 53, emphasis added).

Secondly, it is shown that, through spectroscopic analysis, one can control perhaps the most significant source of error: supernova evolution. If supernovae in the early universe are different from the nearby (i.e. later) objects, treating them as identical would introduce error. Hence, it is crucial for the supernova measurement protocol to be able to empirically ascertain this for high redshift objects. Thirdly, the central aim of the photometry, i.e., the maximum brightness, can be obtained with confidence.

As a final point, note also that the paper offered a *robustness* argument for the measurement of peak magnitudes:

These peak magnitudes are *relatively insensitive to the details of the fit:* if the date of maximum is left unconstrained or set to the date indicated by the best-match spectrum, or if the ground- and space-based data are fitted alone, *the peak magnitudes still agree well within errors* (Perlmutter et al., 1998, 52, emphasis added).

The next paper the group published fully applied this endorsed experimental system to conduct the measurement with a much larger data set. It is this paper that is usually taken to be the *primary* evidence paper of the SCP, and that is referred to in connection with the Nobel Prize and the discovery of dark energy. Before examining the evidential strategy presented there, I will now examine two transitory papers by the High-z that represents that group's stage of *endorsement.*

3.4   Two Further Papers by the High-z

The High-z published three cosmology papers in the year 1998. The first of these was a short article that came out in the February issue of the *Astrophysical Journal Letters*, Peter

Garnavich being the lead author (Garnavich et al., 1998). The second paper was written by the lead author Brian Schmidt, in which the team laid out the general methodology of the High-z (Schmidt et al., 1998). The third was led by Adam Riess, which contained the announcement of the evidence for the accelerating universe (Riess et al., 1998).[34] In contrast with Riess's paper, neither of the two other papers made any evidential claims about acceleration. These two papers represent the transition to the endorsement stage of High-z's work. In order to understand how they relate to each other, and to Riess's evidential paper, we need to analyze how the arguments each of them offered were constructed. On this basis of the comparative analysis, we can see how endorsement works, as a distinct form of iterative progress. Historically, Schmidt's paper was planned to be published before the others, as it lays out the High-z teams's methodological approach to the problem. On the other hand, (Garnavich et al., 1998), can be compared to SCP's *Nature* paper, where a preliminary result is presented, and the need for further data is expressed.

Schmidt's programmatic paper, entitled, "The High-Z Supernova Search: Measuring Cosmic Deceleration and Global Curvature of the Universe Using Type Ia Supernovae," has two main parts (Schmidt et al., 1998). In the first part, it lays out the general components of the measurement protocol, including the cosmology formalism, the search strategy and software, K-corrections, and the possible sources of error such as host galaxy extinction, or selection biases. In the second part, the paper reports a cosmological measurement based on a single high-redshift object, SN 1995K. Both these sections include important elements from the perspective of the endorsement of an experimental system of knowledge. Thus, in discussing the optimum search strategy for finding supernovae for cosmological purposes, (Schmidt et al., 1998) argues that even though it is technically possible to find $z \approx 1$ objects

> it is efficient to measure cosmological parameters by observing objects in the range $0.35 < z < 0.55$. When systematic effects are small, the leverage gained with high-redshift objects is offset by the difficulty in obtaining accurate measurements ... objects observed at $0.55 < z < 0.9$ ... are currently less powerful tools for measuring cosmological parameters than their lower redshift siblings. The [Hubble Telescope] could acquire accurate rest frame B and V measurements for SNe Ia at $z \approx 1$. These objects hold the promise of establishing powerful constraints on cosmology within this more distant observational window. *From the ground, however, the band* $0.35 < z < 0.55$ *gives the best combination of measurements and systematics to investigate cosmology* (Schmidt et al., 1998, 49, emphasis added).

Hence, the search strategy does not simply aim for finding very high redshift objects (which would be preferred in an ideal epistemic setting), but finding objects that can be adequately analyzed within the *current technical and computational limitations of the measurement system.* The planning of the experiment and the data collection process not only relate to

---

[34]Although submitted a few months before Riess's, Schmidt's paper appeared last in print. Schmidt drafted the paper in 1997, yet its completion and submission was delayed due to many responsibilities he had within the group (Schmidt, 2007).

the cosmological model that is to be "tested," but they are also intimately connected with the affirmed (knowledge of) system of measurement.

The programmatic section of the paper ends with the discussion of the various possible sources of systematic error, such as host galaxy extinction, supernova evolution and selection biases. After summarizing their contributions to the uncertainty in the measurement, Schmidt concludes:

> The data presented by Garnavich et al. (1998) and Riess et al. (1998) indicate that we can measure the distances to high-z supernovae with a statistical uncertainty of $\sigma = 0.2$ mag (10%) per object. *With only 10 objects, a comparison of $z \approx 0$ and $z \approx 0.5$ can be made to a precision of better than 5%–leaving systematic uncertainties as a major contributor to the total error budget.* A summary of the contributions to high-redshift supernova distance uncertainties ...*shows that our program to measure cosmology will most likely be limited by the possibility of the evolution of SN Ia explosions* ...Future work to address this possible problem will be as important as obtaining large numbers of objects at high redshift (Schmidt et al., 1998, 55, emphasis added).

Hence, before the actual measurement was done, a clear delineation of the measurement system was presented in terms of the required number of objects with their redshift range, and the possible sources of statistical and systematic error. The scientists argued that their measurement system was capable of delivering a meaningful result due to its design in data collection and analysis, in spite of the uncertainties involved that still needed to be dealt with.

The section of the paper which reports the measurement of Ω_Λ and Ω_M on the basis of SN 1995k contains a detailed description of the observation and the analysis of this object. In particular, the photometric calibrations, the application of K-corrections, and the light curve fitting techniques are discussed. After these detailed discussion, a cosmological conclusion is presented as follows:

> *With a single object, it is difficult to make serious conclusions about cosmological parameters, regardless of the distance precision it offers, because there is no way to judge systematic errors in an empirical way.* However, taken at face value, if $\Omega_\Lambda = 0$, SN 1995K yields an estimate of $\Omega_M = -0.2^{1.0}_{-0.8}$ for the matter density of the universe ...For a spatially flat universe composed of normal matter and a cosmological constant, we find $\Omega_M = 0.4^{+0.5}_{-0.4}, \Omega_\Lambda = 0.6^{+0.4}_{-0.5}$. An $\Omega = \Omega_M = 1$ universe is excluded with greater than 80% confidence from this single distance estimate (Schmidt et al., 1998, 58-59, emphasis added).

Hence, even though Schmidt reported a non-zero cosmological constant, he did not have sufficient data to confidently make this claim (as can be seen from the error estimates, which are very large in comparison to the reported value): phenomena were beyond the epistemic access of his protocol.

Figure 3: High-z Precision Analysis. Source: (Schmidt et al., 1998, 60). Reproduced by permission of the AAS.
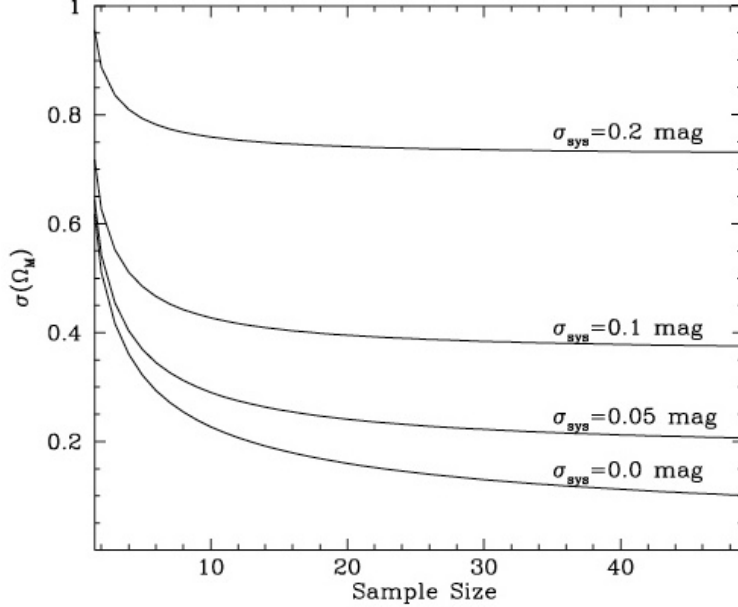
From the perspective of the construction of the measurement protocol, the most significant part of the paper is the last section, entitled "The Future" (Schmidt et al., 1998, 59). Here, Schmidt discussed the required number of objects (sample size) and the necessary level of precision for an evidentially viable measurement:

> SN 1995K is the first of more than 30 confirmed SNe Ia discovered by the High-Z SN Search (Schmidt et al. 1995, 1997a, 1997b; Kirshner et al. 1995; Suntzeff et al. 1996; Garnavich et al. 1996a, 1996b, 1997). The observations presented here indicate that these objects, and those of Perlmutter et al. (1997), should provide an accurate method of tracing out luminosity distances to high redshifts. Figure 12 [Fig. 3] shows *the level of precision likely to be achieved by a sample of N [number of] SNe Ia with observations of comparable precision to those of SN 1995K; a measurement of Ω_M to ± 0.2 should be achieved from objects at $z < 0.5$ given our expected systematic uncertainties* (Schmidt et al., 1998, 59, emphasis added).

Thus, before the actual measurement is conducted, the High-z scientists studied how the uncertainty in their results would change under different systematic errors and sample sizes. This analysis was a necessary step for the construction of the measurement protocol, for only on its basis, one could claim evidential value for the observations.

Schmidt also provided a comparative discussion of the SCP results with the High-z team's work. After cautioning his readers that a comparison with (Perlmutter et al., 1997) could be problematic as not all of their objects were spectrally confirmed as Type Ia's, and the SCP group applied no correction for the host galaxy extinction, he wrote:

However, taken on equal terms, SN 1995K suggests that Ω_M may be lower than the central value of $\Omega_M = 0.88^{+0.60}_{-0.69}$ they find. Recent HST observations by the two groups (Garnavich et al. 1998; Perlmutter et al. 1998) seem to confirm this view... *More data will elevate this discussion* (Schmidt et al., 1998, 59, emphasis added)

Note the careful language Schmidt uses here, which is reminiscent of SCP's *Nature* paper, which announced that their "new measurements *suggest* that we *may* live in a low-mass-density universe" (Perlmutter et al., 1998, 51, emphasis added). For both papers, the main claim is not a new cosmological discovery, or a definitive measurement of a parameter, but explicating the potency of a measurement protocol in the making, and the conditions of its full realization. In this sense, both are transitory papers to the measurement protocol.

The (Garnavich et al., 1998) paper, in addition to being a transition to the more comprehensive (Riess et al., 1998), is also the first paper of the High-z, where the group actually made a measurement claim, albeit on limited data. The paper builds its main argument on data collected from observations made with the HST. It contains three supernovae that were photometrically followed by the HST, and only two of these were spectroscopically confirmed to be Type Ia's. Similar to earlier cases that we have seen, the paper computes the cosmological parameters both with this object, and without it, as a precautionary measure to ensure the robustness of the results. The analysis includes these three objects, in combination with the SN 1995K, that is also analyzed by (Schmidt et al., 1998).

In spite of being a "short notice,"[35] Garnavich paper contains a substantial discussion of the details of the observations, including the specifics of data reduction, filter information, calibration of the HST observations, and the combination of the ground-based and HST data. The sources of possible error are discussed, and for the details the reader is referred to the main methodological paper of the group, (Schmidt et al., 1998). It can be said that Schmidt's paper functions as the main background reference for the argument of (Garnavich et al., 1998), as, in addition to the error budget, the data reduction techniques and K-corrections also follow this paper. I emphasize this fact, as it shows how intimately connected the arguments of these works are, which points to the unity of the measurement protocol that supports them. None of these papers' arguments can be assessed in isolation, but can only be understood as part of the whole measurement process.[36]

The cosmological results of the paper are presented as follows. First the three spectroscopically confirmed objects are used for the fits. For the flat universe prior ($\Omega_M + \Omega_\Lambda = 1$), the paper obtains $\Omega_M = 0.4^{+0.3}_{-0.3}$ and $\Omega_M = 0.3^{+0.3}_{-0.3}$, using two different light-curve fitting methods, respectively. For a zero cosmological constant prior $\Omega_\Lambda = 0$, the analysis gives: $\Omega_M = -0.1 \pm 0.5$. As a result, Garnavich concludes:

From either method we find that $\Omega_M < 1$ with 95% confidence. Including SN 1997ck in the analysis tightens the constraints on Ω_M and Ω_Λ but does not alter

---

[35]The paper was published in the *The Astrophysical Journal Letters*. According to its website, this journal "is a peer-reviewed express scientific journal that allows astrophysicists to rapidly publish short notices of significant original research" (APJ).

[36]A very similar relationship exists for the SCP papers as well.

the best-fit values. Our data suggest that the matter density is low. Either the universe is open or flat; if flat, then *a cosmological constant makes a considerable contribution...* (Garnavich et al., 1998, L56, emphasis added).

This is the first non-zero lambda claim in the history of supernovae cosmology. Note that, due to the limitations of the data, the claim is made conditional on the flat universe. Hence, if the universe is open, the data still allowed a zero cosmological constant. On this basis, the ultimate conclusion of the paper is that, similar to the previous cases, further research is necessary. The concluding paragraph of the paper emphasizes that the main contribution of the paper is the endorsement of the measurement program:

> *We have shown that SNe with redshifts as large as $z \approx 1$ can be discovered and successfully studied with a combination of ground-based telescopes and HST.* Refinements will be made to this data set once template images are acquired with HST, and additional ground-based photometric data are obtained. Our initial sample of four SNe is inconsistent with a high matter density $\Omega_M \approx 1$, although the strength of these conclusions should be tempered by the less than perfect data set for SN 1997ck and the small size of our present sample. *Additional objects will allow us to increase the precision of our measurement and test for sources of systematic error* (Garnavich et al., 1998, L56, emphasis added).

It was the (Riess et al., 1998) paper that fulfilled the requirements stated in the last sentence of this paragraph, and executed the measurement within the protocol that (Schmidt et al., 1998) and (Garnavich et al., 1998) endorsed.

3.5   Analysis of the High-z Team Evidence

The primary publication of the High-z group that put forward the acceleration claim explicitly is (Riess et al., 1998). The evidence claim in this paper is built on the endorsement of the measurement protocol that the group established in their previous publications, and is defended by various *robustness* arguments. The paper defends its claim by arguing on four main fronts: the acquisition of data, its analysis, its cosmological interpretation, and the ruling out of error.

Data acquisition in supernova cosmology has three main components: discovery, classification, and follow-up photometry. The acquisition protocol of (Riess et al., 1998) is built on the methodological program of (Schmidt et al., 1998), targeting the $0.3 < z < 0.6$ redshift range. The paper analyses 10 new objects which are all discovered through the (Schmidt et al., 1998) prescription. For all 10 objects introduced, the paper provides their identification spectra, which gives evidence that the objects are Type Ia. Finally, for the follow up photometry, the group designed specific filters to match the observation of the high redshift objects to standard rest frame filters. These filters were standardized using local standard stars. The construction and the standardization scheme of the filters are described in both (Schmidt et al., 1998) and (Riess et al., 1998) in detail.

As we have seen, the data analysis consists of two components: K-corrections and light curve fitting.[37] In order to calculate the K-corrections, (Riess et al., 1998) applied the prescription outlined in an earlier paper by the SCP group (Nugent et al., 2002). However, this calculation involves a circularity, and as one can expect, this is overcome through an *iterative* strategy:

> This prescription requires the age and observed color for each observation to be known before its K-correction can be calculated. The age is best determined from fitting the light curve's time of maximum. Yet we must use the K-correction to determine the time of maximum and the true color of each epoch. *This conundrum can be solved by iteratively converging to a solution by repeated cycles of K-correcting and empirical fitting of the light curves* (Riess et al., 1998, 1019, emphasis added).

The iteration involved in K-corrections employs a computational analysis, but it still enacts an epistemic function, as it begins with an affirmed form of the $K-$correction, and improves this form based on the computations conducted using the initial assumption. This aspect of the data analysis shows that these two forms of iteration (i.e., computational and epistemic) co-exist in the production of evidence claim, and mutually support each other.[38]

Of the 10 new objects analyzed in the paper, only 6 had sufficient photometry to be amenable to the MLCS or template-fitting methods.[39] The remaining 4 objects (augmented by 2 previously published ones, which similarly had a few photometry data points) were analyzed with a method that Riess developed earlier. In this "snapshot" method, a single photometric observation is combined with spectrum information to estimate distance. Hence, of the total 16 objects analyzed, 10 objects are fitted by the MLCS and the template fitting, and 6 objects by the "snapshot" method.

3.5.1 Data Interpretation: Evidence for Acceleration

(Riess et al., 1998) contains a detailed discussion of the interpretation of the data, compared to the early papers of the team that I examined. This interpretation has two aspects.

---

[37]The High-z used two prescriptions for light curve fitting, known as the Multicolor Light-Curve Shapes (MLCS) and the template-fitting methods. The former was formulated by Adam Riess in his doctoral thesis, and the latter was utilized by the members of the Calán/Tololo Supernova Survey, some of whom later joined the High-z. Calán/Tololo was a supernova search conducted at the Cerro Tololo Inter-American Observatory during the years 1990-1996 (Hamuy et al., 1993a).

[38]According to Chang, epistemic iteration differs from computational iteration in two crucial ways. First, "the latter is used to approach the correct answer that is known, or at least in principle knowable, by other means" (Chang, 2004, 45). And second, as opposed to epistemic iteration, "mathematical iteration relies on a single algorithm to produce all successive approximations..." (Chang, 2004, 46). As the computation of the K-correction is not "in principle" knowable by other means, and many different algorithms must be innovated depending on the particular filter system that one uses, an epistemic iteration underlies the computational one in this case.

[39]In addition to these 6 objects, (Riess et al., 1998) also fitted the SN 1995K, and the three HST objects studied in the (Garnavich et al., 1998) paper with the two prescriptions.

On the one hand, the paper determines the best fit values of the cosmological parameters, and on the other, the probability density function of the parameters is constructed. The probability density function is then used to generate joint confidence intervals that provide a broader understanding of the behavior of the parameters. These analyses are built on a formalism which combines elements from frequentist statistics in determining the best fit parameters, with an overall Bayesian approach. This Bayesian machinery undergirds a *robustness argument*, as I explain below.

The main function that the paper uses to calculate the cosmological parameters is a frequentist χ^2 statistic:

$$\chi^2(H_0, \Omega_M, \Omega_\Lambda) = \sum_i \frac{[\mu_{p,i}(z_i; H_0, \Omega_M, \Omega_\Lambda) - \mu_{0,i}]^2}{\sigma_{\mu_{0,i}}^2 + \sigma_\nu^2} \tag{1}$$

(Riess et al., 1998, 1021).

Here $\mu_{p,i}(z_i; H_0, \Omega_M, \Omega_\Lambda)$ stands for the theoretical distances predicted by the cosmological model, and the $\mu_{0,i}$ are the measured distances. The $\sigma_{\mu_{0,i}}^2$ denotes the error associated with each measurement. Entering the computation in the denominator, these functions weigh the contribution each data point makes to the result in an inverse way: the larger the error, the less the associated $\mu_{0,i}$ contributes to the measurement. Finally, σ_ν denotes the dispersion in the galaxy redshift, which is assigned a constant value of 200 km s^{-1}. Using this statistic one determines the cosmological parameters $(\Omega_M, \Omega_\Lambda)$ by minimizing the difference between the predicted values by various choice of parameters and the actual data obtained. In order to determine how the data constrains the two dimensional parameter space $(\Omega_M, \Omega_\Lambda)$, one constructs the two-dimensional *confidence regions* using the probability density function (PDF). To obtain this PDF, Bayes theorem is indispensable:

> The χ^2 statistic of equation (1) is well suited for determining the most likely values for the cosmological parameters $H_0, \Omega_M, \Omega_\Lambda$ as well as the confidence intervals surrounding them. For constraining regions of parameter space not bounded by contours of uniform confidence (i.e., constant χ^2), we need to define the probability density function (PDF) for the cosmological parameters.[40] The PDF (p) of these parameters given our distance moduli is derived from the PDF of the distance moduli given our data from Bayes's theorem,
>
> $$p(H_0, \Omega_M, \Omega_\Lambda | \mu_0) = \frac{p(\mu_0 | H_0, \Omega_M, \Omega_\Lambda) p(H_0, \Omega_M, \Omega_\Lambda)}{p(\mu_0)} \tag{2}$$
>
> where μ_0 is our set of distance moduli ... (Riess et al., 1998).

---

[40]The idea expressed here can be understood as follows: On the two-dimensional parameter space, there are regions which contain parameter values that correspond to different types of universes, such as decelerating universes, re-collapsing universes etc. In order to quantify these regions probabilistically, one needs a function that can assign a probability value to each point on the parameter space. This is achieved with the use of the Bayes theorem.

Figure 4: High-z MLCS Contour Plot. Source: (Riess et al., 1998, 1023). Reproduced by permission of the AAS.
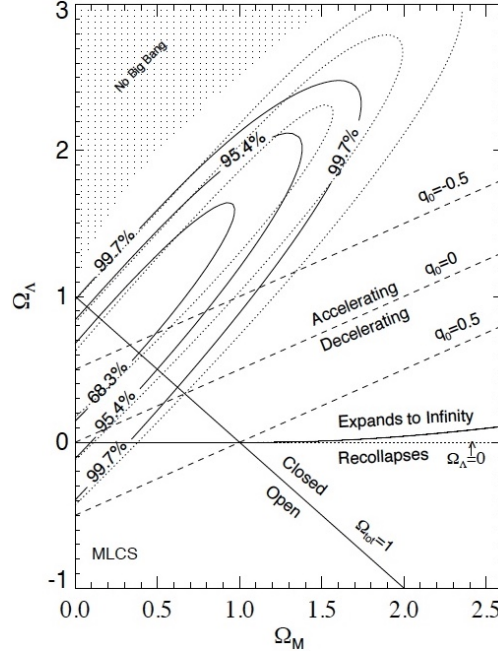
As the determinations of the cosmological parameters depend on the *relative difference* of the brightnesses of the nearby versus distant supernovae, the measurement is independent of the absolute value of the Hubble constant. Hence H_0 can be integrated out from Equation (2) to obtain the final form of the two dimensional PDF of the parameters, given the data. This function, denoted as $p(\Omega_M, \Omega_\Lambda | \mu_0)$, is the main analytical tool for all the cosmological computations in the paper. Thus, using this function one obtains the probability that the cosmological constant

is greater than zero ... by summing the likelihood for this region of parameter space:

$$p(\Omega_\Lambda > 0 | \mu_0) = \int_0^\infty d\Omega_\Lambda \int_0^\infty p(\Omega_M, \Omega_\Lambda | \mu_0) d\Omega_M \ . \tag{3}$$

(Riess et al., 1998, 1024).

The probability that the cosmological constant is non-zero is obtained by evaluating this integral "numerically over a wide and finely spaced grid of cosmological parameters..." (Riess et al., 1998, 1024). In other words, one sums up the area of the probability space where Ω_Λ is greater than zero (See Figures 4 and 5.)

On the basis of this machinery, a *robustness* argument for the evidence of accelerating universe is built as follows. First, using the nine objects which have well observed light-curves, and which are spectroscopically identified as Type Ia's, the paper infers that "a non-negligible positive cosmological constant is strongly preferred at the $99.6\%(2.9\,\sigma)$ and greater

Figure 5: High-z Template-Fitting Contour Plot. Source: (Riess et al., 1998, 1023). Reproduced by permission of the AAS.
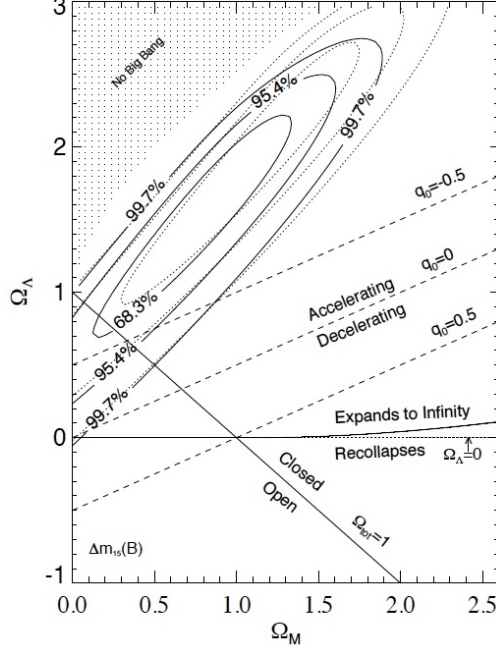
than $99.9\%(3.9\ \sigma)$ confidence levels for the *MLCS* and *template-fitting methods*, respectively" (Riess et al., 1998, 1024). Thus, the two fitting methods give similar probabilities for the positive lambda. Two constraints are introduced to more specifically determine the value of the parameters. First, if one imposes the universe to be spatially flat, i.e., that $\Omega_M + \Omega_\Lambda \equiv \Omega_{tot} \equiv 1$, then the MLCS gives $\Omega_\Lambda = 0.68 \pm 0.10$ for the cosmological constant, and $\Omega_M = 0.32 \pm 0.10$ for the matter density. Similarly, template fitting method gives $\Omega_\Lambda = 0.84 \pm 0.09$ and $\Omega_M = 0.16 \pm 0.09$.

The above results are computed with the set of 9 well observed objects, using two different light curve fitting methods. In addition to these results, other computations with different sets of objects are provided as well, always using both of the light curve fitting methods. In total, the paper fits four different data sets (with both light curve fitting methods) to the model to obtain the parameters, which all converge towards the same results within error limits (Table 1).

3.5.2 Systematics and the Final Claim

After this conclusion is provided, a final discussion is devoted to the possible systematic effects that could be influencing the results. The following seven sources of systematic error are discussed: (i) Supernova Evolution, (ii.) Dust Extinction, (iii.) Selection Bias, (iv.) Effect of a Local Void, (v.) Weak Gravitational Lensing, (vi.) Light Curve Fitting, (vii) Method Sample Contamination. For each case, the paper argued against the possibility that the effect in question might be responsible for the dim appearance of supernova, as opposed

| Light Curve Fit / Data Set | $p(\Omega_\Lambda \geq 0)$ | Constraint | | Number of Objects |
		$\Omega_{tot} = 1$	$\Omega_\Lambda = 0$	
MLCS	99.6%(2.9 σ)	$\Omega_M = 0.28 \pm 10$	$\Omega_M = -0.38 \pm 0.22$	9
ΔM_{15}	> 99.9%(3.9 σ)	$\Omega_M = 0.16 \pm 0.09$	$\Omega_M = -0.52 \pm 0.20$	9
MLCS + Snapshot	99.7%(3.0 σ)	$\Omega_M = 0.28 \pm 10$	$\Omega_M = -0.34 \pm 0.21$	15
ΔM_{15} + Snapshot (15)	> 99.9%(4.0 σ)	$\Omega_M = 0.17 \pm 0.09$	$\Omega_M = -0.48 \pm 0.19$	15
MLCS + Snap. + 97ck	99.5%(2.8 σ)	$\Omega_M = 0.24 \pm 0.10$	$\Omega_M = -0.35 \pm 0.18$	16
ΔM_{15} + Snap. + 97ck	> 99.9%(3.9 σ)	$\Omega_M = 0.21 \pm 0.09$	$\Omega_M = -0.41 \pm 0.17$	16
MLCS + 97ck	99.5%(2.8 σ)	$\Omega_M = 0.24 \pm 0.10$	$\Omega_M = -0.38 \pm 0.19$	10
ΔM_{15} + 97ck	> 99.9%(3.8 σ)	$\Omega_M = 0.20 \pm 0.09$	$\Omega_M = -0.44 \pm 0.18$	10

Table 1: Cosmological Results of the High-z Primary Evidence Paper. Source: (Riess et al., 1998, 1025).

to a cosmological constant. As mentioned above, one possible cause of supernovae appearing dim could be *evolution*, i.e., supernovae in the early universe (high redshift) might differ from the more recent ones (nearby.) The group used comparisons of the spectra and light-curve shapes of the more distant and nearby objects in their sample to check this effect. Even though they found that "evolution provides an inadequate explanation" (Riess et al., 1998, 1028) for their observations, they did not fully rule it out either, ending with a cautious note: "Although our current observations reveal no indication of evolution of SNe Ia ... evolution remains a serious concern that can only be eased and perhaps understood by future studies" (Riess et al., 1998, 1029).

After similar considerations on the remaining list of the systematic effects, the paper concluded that these effects "currently do not provide a convincing substitute for a positive cosmological constant" (Riess et al., 1998, 1034). Still, this did not imply that the result was decisively established, as the authors cautioned in their final sentence that " [f]urther studies are needed to determine the possible influence of any remaining systematic uncertainties" (Riess et al., 1998, 1029).

3.6   Analysis of the SCP Evidence

The primary evidence paper of the SCP that established the existence of a positive cosmological constant is entitled "Measurements of Ω and Λ from 42 High-Redshift Supernovae" (Perlmutter et al., 1999). As we have seen, before this publication, SCP first argued that Λ equaled zero with their (Perlmutter et al., 1997), but then modified this claim a year later, writing that their new results "are preliminary evidence for a relatively low-mass-density

universe" (Perlmutter et al., 1998, 53). (Perlmutter et al., 1999) refined this result further, and made an explicit argument for the existence of a cosmological constant.

The introductory section of (Perlmutter et al., 1999) explains the main improvements of this paper over (Perlmutter et al., 1997) (following the SCP practice, I will refer to this paper as the P97). Firstly, the classification of the "[a]lmost all of the new supernovae" were made through spectroscopy (Perlmutter et al., 1999, 566). This is a major improvement over P97, which had only two objects that were spectroscopically identified. Secondly, the calculation of K-corrections are updated: "We have now recalculated these K-corrections ... using improved template spectra, based on an extensive database of low-redshift SN Ia spectra recently made available from the Calán/Tololo survey ... Where available, IUE [International Ultraviolet Explorer] and HST spectra ... were also added to the SN Ia spectra, including those published previously" (Perlmutter et al., 1999, 566). These recalculations are done iteratively, as the K-corrections depend on using a "template" object, which in turn requires a K-correction to be applied for its very construction (as explained above.)

After the introductory section which reports on the improvements on data analysis, the results are presented. The main computational procedure is the same as followed in P97: the data are fitted to a cosmological model based on the Friedmann metric, in order to obtain the *best fit* cosmology parameters. In total, the paper presents 12 different fits, formed by different data subsets, which are labeled alphabetically from A to L (See Table 2). The main aim of these different fits is ensuring the *robustness* of the result, which is emphasized very strongly in the paper. Thus, in the section where the results are announced, we read:

> From Table 3 [Table 2] ... it is clear that the results of fits A, B, and C are quite close to each other, so we can conclude that our measurement is *robust* with respect to the choice of these supernova subsets. The inclusive fits A and B are the fits with the least subjective selection of the data. They already indicate the main cosmological results from this data set. However, to make our results *robust* with respect to host-galaxy reddening, we use fit C as our primary fit in this paper (Perlmutter et al., 1999, 579, emphasis added).

This strategy of checking the robustness of the results by employing different fits is repeatedly used in the paper. This is a feasible strategy for SCP for they have a large number of high-redshift objects to work with. The total number of high-redshift supernovae that the SCP uses in the paper is 42. Combined with the low redshift sample published by the Calán/Tololo survey, the most inclusive fit (fit A) of the paper includes 60 objects. Another fit (fit B) is obtained by removing 4 high-redshift objects from fit A set. Two of these objects are removed due to being "the most significant outliers from the average light curve width...", and two others for being "the largest residuals" from fit A (Perlmutter et al., 1999, 571). The removal of the former two objects is justified via a robustness argument: as the remaining set of supernovae are all very close to the average object, the particular form of the *standardization* relation applied to the remaining set would not have a large influence on the results. In other words, the removal of two objects results in a robust set with respect to the standardization relation. On the other hand, the two latter objects are removed from the full set *iteratively.* Using fit A one obtains the best fit results for the cosmological

			Constraint		
Fit	Fit Description and Data Set	$p(\Omega_\Lambda > 0)$	$\Omega_{tot} = 1$	Best Fit $(\Omega_M, \Omega_\Lambda)$	Number of Objects
A	All Supernova	0.9984	$\Omega_M = 0.29^{+0.09}_{-0.08}$	$0.83, 1.42$	60
B	Fit A, Excluding Residual and Stretch Outliers	0.9992	$\Omega_M = 0.26^{+0.09}_{-0.08}$	$0.85, 1.54$	56
C	Fit B, Excluding two likely reddened	0.9979	$\Omega_M = 0.28^{+0.09}_{-0.08}$	$0.73, 1.32$	54
D	No stretch correction	0.9972	$\Omega_M = 0.25^{+0.10}_{-0.09}$	$0.76, 1.48$	54
E	Bayesian	0.9894	$\Omega_M = 0.29^{+0.12}_{-0.10}$	$0.35, 0.76$	53
F	Fit B, Supernovae with colors measured	0.9991	$\Omega_M = 0.26^{+0.09}_{-0.08}$	$0.85, 1.54$	51
G	Fit C, Supernovae with colors measured	0.9974	$\Omega_M = 0.28^{+0.09}_{-0.08}$	$0.73, 1.32$	49
H	Fit G, excluding seven reddest and two faintest high red-shift supernovae	0.9857	$\Omega_M = 0.31^{+0.11}_{-0.09}$	$0.16, 0.50$	40
I	Fit C with $+0.03$ mag systematic offset	0.9994	$\Omega_M = 0.24^{+0.09}_{-0.08}$	$0.80, 1.52$	54
J	Fit C with -0.04 mag systematic offset	0.9912	$\Omega_M = 0.33^{+0.10}_{-0.09}$	$0.72, 1.20$	54
K	Empty beam metric with Bayesian Prior	0.9984	$\Omega_M = 0.35^{+0.12}_{-0.10}$	$2.90, 2.64$	54
L	Partially filled beam metric	0.9974	$\Omega_M = 0.34^{+0.10}_{-0.09}$	$0.94, 1.46$	54

Table 2: Cosmological Results of the SCP Primary Evidence Paper. Source: (Perlmutter et al., 1999, 573).

Figure 6: SCP Contour Plots. Each plot represents a data sub-set, different types of corrections, or fitting methods, as well as combinations thereof. Source: (Perlmutter et al., 1999, 574). Reproduced by permission of the AAS.
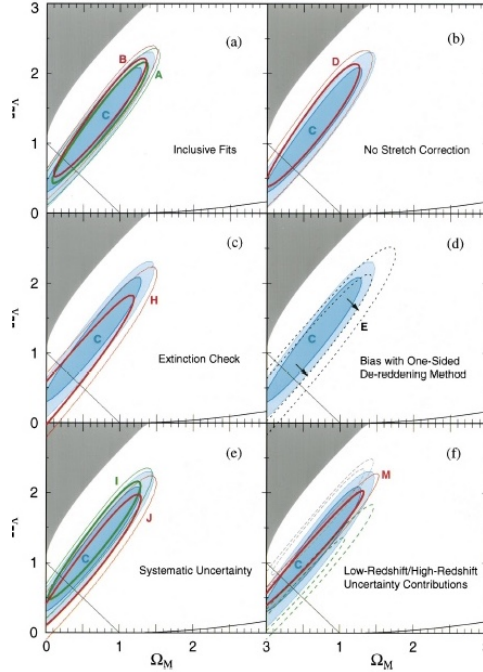
parameters. On the basis of these parameters a prediction for supernovae brightnesses can be obtained for any given redshift. After obtaining this prediction, two objects (which were originally included in the set of objects for obtaining this prediction) are now removed as outliers, one being "fainter than the best-fit prediction and one is brighter" (Perlmutter et al., 1999, 571). From the resulting set of objects, the group removed "two likely reddened objects" further, in order to obtain the primary fit of the paper (fit C in Table 2) (Perlmutter et al., 1999, 573). In a contour plot, the group compared each of the fits to the "result of fit C (shaded regions) compared with fits to different subsets of supernovae, or variant analyses for the same subset of supernovae, to test the robustness of the fit C result" (Perlmutter et al., 1999, 574). (Fig. 6.)

Using the primary fit C, the group stated their conclusion concerning the measurement of the cosmological parameters as follows:

> For fit C, we find $\Omega_M = 0.28^{+0.09}_{-0.08}$ in a flat universe. Cosmologies with $\Omega_\Lambda = 0$ are a poor fit to the data at the 99.8% confidence level (Perlmutter et al., 1999, 573).

Similar to the High-z evidence paper, (Perlmutter et al., 1999) defended this conclusion on the basis of a detailed study of possible systematic effects. After a quick discussion of the error budget, i.e., the "relative statistical and systematic uncertainty contributions" to the result, the paper stated its positive cosmological constant conclusion as follows:

> *All of the alternative fits listed* in Table 3 [Table 2] indicate a positive cosmo-
> logical constant with confidence levels of order 99%, even with the systematic
> uncertainty included in the fit ... (Perlmutter et al., 1999, 581, emphasis added).

Hence, the convergence of the different fits, i.e., the *robustness* of the result, played a central role in the SCP evidence claim.

4   Hudson's Objection and Robust Procedures

Robert Hudson criticized the view that robustness arguments play a significant role in scientific knowledge production. In *Seeing Things*, he specifically argued, concerning the discovery of the accelerating universe, that "robustness reasoning (to the extent that it occurs at all) is not fundamental to the thought process of discoverers" (Hudson, 2013, 166). Furthermore, after mentioning the "striking convergence" of the results of the High-z and the SCP teams, he claimed:

> In the key research articles to the discovery, one doesn't find this *(or any other)* robustness reasoning introduced in any authorized fashion: The convergence of results is stated as more of an after thought, introduced after the 'real' work of adequately justifying one's observational methods is accomplished (Hudson, 2013, 166, emphasis added).

As this quote makes it clear, according to Hudson, no form of robustness reasoning played any significant role in the work of the High-z and the SCP.[41]

More specifically, Hudson singles out two possible robustness arguments that could be generated in the context of the accelerating universe discovery (Hudson, 2013, 155-157). One of these is the concordance between the supernova data, galaxy cluster observations, and the cosmic microwave background (CMB) measurements of the COBE and WMAP satellites. The CMB data implied that the universe was flat (i.e., $\Omega_M + \Omega_\Lambda = 1$), and the galaxy cluster measurements implied that $\Omega_M = 0.3$. This information strongly supported the SCP and the High-z claim that cosmological constant was positive.[42] The other argument was the

---

[41]According to Hudson, the supernova cosmologists employed a form of reasoning that he calls "reliable process reasoning," which involves "targeted testing." Whereas robustness deduces the reliability of a process, or a result, from the convergence of the test outcomes, reliable process reasoning "assumes the reliability of a process and then, on this basis, infers the truth of an observed result" (Hudson, 2013, 68). Targeted testing is a form of reliable process reasoning in which scientists decide between two competing hypotheses by *targeting* a specific aspect of the empirical situation that is sufficient to break the underdetermination. The assessment of Hudson's arguments against robustness in general, as well as the issue of whether "targeted testing" does capture the reasoning of the two teams' is beyond the scope of this paper. Here, my modest aim is to show that a form of robustness argument played a significant role in the argument of both of the teams—and justifiably so.

[42]Strictly speaking, this "convergence" cannot be characterized as robust detection, as these results are not the same. The supernova data constrains $\Omega_M - \Omega_\Lambda$, whereas the CMB measurements quantify $\Omega_M + \Omega_\Lambda$. Finally, the galaxy cluster measurements were made with the assumption that $\Omega_\Lambda = 0$. Due to this reason, the SCP referred to these results as "complementary" and the High-z characterized them as "orthogonal" to their results, and did not invoke a robustness claim.

convergence of the results of the two teams themselves. Following this, Hudson argues that neither of these arguments were instrumental in convincing "the astrophysical community that it should embrace the reality of dark energy" (Hudson, 2013, 157). In fact, none of these arguments were put forward by the teams as supporting a positive lambda result. In this sense, Hudson's claim that a robust detection argument did not play a role in the confirmation of the accelerating universe result has merit.

However, these two arguments do not exhaust the possible set of robustness arguments that were available to the supernova scientists: the teams relied on their own measurement procedures to achieve robustness. Hence, by examining the scientists' practice, we can discern a different form of robustness reasoning that was at play. As we have seen above, the SCP provided 12 different fits to their data (indicated with capital letters A-L). These fits employed variations in the data set, or the analysis methods. For example, the most inclusive Fit A employed the complete supernova data set. Fit B excluded 4 objects from this set, 2 of them due to being largest residuals to the Fit A, and the remaining two for being stretch correction outliers. In Fit C, two more objects are removed due to being dimmed by extra-galactic (i.e., outside Milky Way) dust. The group submitted the Fit C as the primary fit of their paper on the basis of a robustness claim as follows. In Fit D, Fit C objects are used but without applying them the stretch correction. Similarly, in Fit H, a subset of Fit C objects with colors measured (for extinction information) is analyzed, excluding from this set 9 objects which were either too faint or reddened. The result of this analysis showed that "the cosmological parameters found for Fit H differ by less than half of a standard deviation from those for Fit C" (Perlmutter et al., 1999, 575). In other words, a variation in the data set did not change the result in any significant way. In fact, the group explicitly stated that the aim of these variations was to check robustness, writing in their explanation of (Fig. 6): "Each panel shows the result of Fit C ... compared with fits to different subsets of supernovae, or variant analyses for the same subset of supernovae, to test the *robustness of the fit C result.* "(Perlmutter et al., 1999, 574, emphasis added). The 12 different fits are employed in 4 types of robustness analyses. The first type, referred to as "Comparison Analysis Techniques," includes fits that vary the analysis method by introducing a Bayesian correction, or analyzing the data without applying the stretch correction. The second type, "Effect of Reddest Supernovae," incorporates fits that vary the data set by using only the objects with measured colors, or excludes various objects that are faint, or reddened. Third type, "Systematic Uncertainty Limits," introduces magnitude offsets to each of the high redshift objects used in Fit C to mimic systematic effects. The final type, "Clumped Matter Metrics," studies the effects of gravitational lensing on the supernova light by using a different metric expression in the distance–luminosity formula that represents clumping of matter in the form of compact objects such as MACHOs[43]. None of these variations introduced any significant change in the result, in other words, the primary fit of the paper was found to be robust.

A similar strategy is used by High-z as well. As we have seen above, the High-z used two

---

[43]This acronym stands for "massive (astrophysical) compact halo object," a hypothetical candidate for dark matter.

different light-curve fitting methods, namely the MLCS and the ΔM_{15}, and compared their results. Furthermore, they applied these methods to different subsets of their objects. First and foremost, the High-z primary evidence paper studied 9 high-redshift objects that were fitted with both fitting methods. Next, they added 6 more objects (whose distances were measured with a different method called "snapshot") to the MLCS and ΔM_{15} data sets, and calculated the cosmological parameters with a total of 15 objects. Third, they added another supernova (SN 1997ck), which could not be spectroscopically confirmed to be a Type Ia (but was otherwise consistent with Type Ia objects), and repeated the analysis with a 16 object data set. Finally, the group also analyzed MLCS and the ΔM_{15} objects combined only with the SN 1997ck. All these results were summarized in a table in the paper (Table 1 above). Concerning the comparison of the snapshot and light-curve fitted distances, the authors remarked that it "yields conclusions that are less precise but fully consistent with the *statistically independent* results from the well-sampled SN Ia light curves" (Riess et al., 1998, 1024, original emphasis). A consistency achieved as a result of statistical independence indicates that the authors had a robustness claim in mind.

In an oral history interview that I conducted with Adam Riess, I specifically asked about the two fitting methods that the High-z group used. This is how he explained its methodological import, and the role robustness played in this context:

> RIESS: Well, yeah that...one would worry about a systematic error that the method that you are using is mis-estimating. So, that's why its very powerful to have two different methods. Now, in a way, ours was fortuitous in that we had this heritage of two different groups within our group: the Calán/Tololo survey and the Harvard group. It was natural the Harvard group, you know, I was working on it at the time, to say "Well I wanna use MLCS because that's what I developed and I know how to use it." Mark Phillips and the Calán/Tololo survey was like "Well, we'd like to use our ΔM_{15} method." I think right away we all realized "hey, why don't we use both of them because *if we don't get the same answer both ways, thats not very reliable*." I think the end result was between having two methods to cross-check each other and having a way to account for dust. When we saw the results from both methods which came basically at the same time. We brought both results in, there was nothing really left to question, if that makes sense. Could there be something wrong with the MLCS? Sure, but ΔM_{15} gets us the same thing ...

> AUTHOR: Yeah because if you look at their [SCP] paper I think they have one method—Saul's stretch method—but what they do is, they try with different sub-samples.

> RIESS: I think we didn't have as much data as them and so sub-samples wouldn't be as...we couldn't do as much of that as we could look at different ways of analyzing the data.

> AUTHOR: It seems this is an important methodological difference.

RIESS: But also, similar in this vein is, now look at it from the community standpoint. They hear this crazy result but they see two different teams which they knew were not collaborative, they were competitive. Yet they were getting this same result. So just like we said with the MLCS and ΔM_{15}, "well I guess this must be the right answer because both methods get it," the community does the same thing. They say "well I guess this must be the right answer because both teams get it." Scientists love the cross-check. *Something is really not true until you see it more than one way.* That was instantaneous essentially with those two groups pursuing it.

AUTHOR: This just reminded me of a term. In philosophy, people talk about "robustness" when different methods agree. And I think in statistics this term is also used.

RIESS: Sure.

AUTHOR: I remember in your papers you sometimes refer to a result being "robust."

RIESS: Correct.

AUTHOR: Do you use this in this sense?

RIESS: Yes, that's what we mean. Usually what we mean is: impervious to or insensitive to a certain variable, to a certain flavor of the analysis.

AUTHOR: If the effect is already there, it should still be there even if you. . .

RIESS: That's right and likewise, if you see the effect with one way of analyzing the data and you have another way that you think is just as valid, starts with maybe different assumptions but is just as valid. *But then if you don't see the result, then the result isn't. . . you can't be confident that its real.* It may be an artifact, it could be an artifact in one method or it could be that the other method has an artifact or bias that is not seeing it, but you can't discriminate between those two possibilities.

Hence, we see that for Riess, the two teams getting the same result *did* play a role in the acceptance of the accelerating universe result by the astronomical community. Furthermore, for Riess, using two methods of light-curve fitting increased the confidence in the result by making it more robust, understood as being insensitive to variations in the analysis.

Consequently, both the SCP and the High-z made a robustness argument to support their claim. To which extent the robustness claim of these groups were *justified?* As I argued above, the main robustness claim of the groups did not stem from the convergence of their results with other team's work, but rested solely on their own data acquisition and analysis procedures. In a recent paper, Karaca distinguishes robustness of an experimental procedure from a robust detection. For Karaca, a data acquisition procedure is robust when it has the "capacity . . . to maintain its intended function invariant despite possible variations in its inputs" (Karaca, 2018, 9). The general aim of the supernova cosmologists in setting up their measurement protocol was to measure the deceleration parameter using the comparative

measurements of supernova brightnesses. The intended function of the data acquisition procedure was to obtain high quality supernova data that could adjudicate between open and closed cosmological models. Constructing a robust data acquisition procedure was a non-trivial issue for the High-z and the SCP. As we have seen, not having such a procedure was the major reason the Danish team to abort their research program. To further document the issue, I will quote from an oral history interview that I conducted with Saul Perlmutter, the leader of the SCP. The following excerpts from the interview document the very early stages of the SCP, during which the team was struggling to construct a robust data acquisition procedure.

In our interview, I asked Perlmutter what types of scientific "hurdles" they had to overcome before being able to obtain reliable data. He explained that one of the first problems they ran into was the appearance of "ghost" objects:

> PERLMUTTER: The first one was the trial run that we did with the 60 inch [Palomar Observatory]. We had impossible time distinguishing possible supernova from other things with that data because there were all these ... what we called "ghosts." The optics would bounce light back to itself and back again so bright stars would appear over and over again in different places of the image, and they wouldn't be in the same place when the sky rotated and the night had changed. So it was much harder to pull out just the supernova because you were getting these other things changing.

As I explained above, identifying supernova requires subtracting a sky image from a reference image obtained earlier. The appearance of these "ghosts" showed that the optics of the SCP data acquisition system was not robust enough: the variations in the input would not consistently lead to identifying supernova, but would also introduce spurious objects. Before they could claim to have a reliable data acquisition system, the group had to improve the subtraction software.

But even a bigger issue was not having a reliable data collection procedure in the first place. When the SCP first began, the search strategy was imaging galaxies one by one, and subtracting an earlier image of a galaxy from a later one to find supernova. Later, the strategy was changed to subtracting images of hundreds of galaxies in a wide field at the same time, and to this end the group decided to develop their own wide field instrument:

> PERLMUTTER: We did this complicated deal with Australia, because there was very little interest in most of the observatories in developing a wide field camera at that time. (Five years later, there was a huge amount of interest in doing that!) But at that time, all anybody wanted to do was spectrographs, so we couldn't get anybody to work on a wide field imager, except the Australia group was interested. So they agreed that they would trade us contributing the chips and the optics, (and they would help work on the electronics)—for twelve nights, twelve and a half nights, of telescope time, for us to actually try the project, using the English-Australian telescope. And so we built the instrument, and that was actually very successful instrument, it actually ended up becoming the mainstay for that telescope for years to come, everybody was

using it. *But for our purposes, we ran into the problem that . . . not surprisingly you can get hit by bad weather over and over and over again. In our case we got . . . out of twelve nights, we got two and a half nights of good weather. And so that wasn't enough to really nail the project.*

The group did not have any additional resources to deal with bad weather, and this significantly crippled the data acquisition process. In addition, following up the candidate objects was also a major problem. During the early days, the data was shipped through the "fledgling internet" from Australia to LBNL, and was analyzed there to find candidate objects:

PERLMUTTER: We were printing out the images with little arrows showing where the computer thought we had a supernova, potentially. And then we would look by eye to see whether or not . . . how we would rank them . . . what we thought the likelihood was. And then . . . some of them we were able to . . . you know look at them more closely . . . some happened that we were able to get some follow up observations and we saw it . . . looked like a supernova was actually rising. *But we didn't get one early enough or at a time when we could convince somebody to take an observation of it with a spectrograph on a big telescope to confirm it was a supernova.*

Given the group size was still relatively small, they did not have access to spectroscopy in the crucial times that was essential for confirming supernovae. So the team would call up telescopes around the world, and try to convince fellow astronomers to obtain spectroscopy for them. As Perlmutter explained:

PERLMUTTER: And so the failing you know at that point was this combination of that we hadn't budgeted for quite enough bad weather, but also that the way we confirmed supernova at the lower redshift end was . . . the thing is you never knew when the supernova was going to happen. So you would have to call all the observatories around the world, and I used to have all the schedules, you know, on my table. And I would look for names I knew and people who were using instruments that would . . . you know to get the spectrum. And I would call up the control rooms, I had the control room phone numbers . . . and I would say "I'm really sorry to bother you, I know you are in the middle of important science: by any chance, is there any way you could be interested in getting a spectrum of one of these supernova?"

Although this method was relatively successful for finding low redshift objects, it was quite unreliable for a high redshift search:

PERLMUTTER: But the problem is, these [high-redshift] ones, you know, are much fainter, and much more distant, and you only can go to the very biggest telescopes, to ask them, you can't do it at the smaller ones. These telescopes are, you know, much more . . . they have even more expensive time from the point of view of the astronomers because they've managed to just get their one or two nights of this big telescope, and

there are a fewer of them, so the odds of having one that has an instrument on, at the right time of the month, is slimmer. And people were genuinely not able to get the spectra that we wanted, to confirm the one or two that we had found.

Thus, even though in principle the group had been able to demonstrate that they could find and identify supernova, the data acquisition system was still far from being optimum for the problem at hand. This was a big set back for the group:

PERLMUTTER: Since we were doing this somewhat by the vagaries of when the weather was gonna be good, you didn't know whether the supernova was gonna be visible at the right time of the night, for different parts of the country ... where you go for. You didn't know whether it would be near the new moon or the full moon, when you found them. And near full moon you couldn't do these things ... objects, anyway, nobody would put the spectrograph on the telescope in full moon anyway. By that point they would put other instruments on. So the odds you could get it were slim enough that given how few nights we really managed to search, and how many times we asked people, it was only one or two candidates, and we didn't get them.

It was clear to Perlmutter that this method of "calling people up to ask for spectra" was not viable in the long run. He reasoned that he only way to overcome this problem was developing a data acquisition procedure in which one could schedule spectroscopic observations beforehand:

PERLMUTTER: It still felt to me like there is this randomness of when you were going to need the spectroscopy method: you were still calling people up and asking them for spectra. And that didn't feel to me as a viable plan for the long run. And nobody had ever yet been able to apply for spectroscopic time and say: "this night is the night that we will definitely have delivered you a supernova. So that was the point at which we started doing this new technique where I ... in retrospect it seemed obvious but at the time it wasn't. Which was this idea of: time it exactly, so you do all of your reference images with no supernovae, you know, with supernovae not yet there, right after new moon. Wait a two and a half weeks till just before the next new moon, and then get all the same images. And now, by fixing that time gap just right, and only collecting the ones that are brightening, you know that they are just before maximum, and now you can schedule the follow up to be on maximum light. And you can actually schedule spectroscopy. And since we now had a demonstration of a proven discovery, we were able actually to apply for spectroscopic follow up, and that was the first time we tried doing that, and by doing enough images with a big enough telescope (wide enough image) that we can guarantee half a dozen. Not just one. And by guaranteeing half a dozen, now its worth having a night scheduled for spectroscopy.

Once this new data acquisition system was in place (referred to as the "batch" method[44]), the group began applying to observation sites with the best weather–such as the Cerro Tololo observatory in Chile—in a regular fashion.

---

[44]I discussed how the group presented this improvement, and its consequences for their measurement program, in Section 3.2 above.

All these improvements that are mentioned above, such as making the subtraction software more precise, or re-organizing the data collection procedure, aimed at making the data acquisition process robust, in the sense of being insensitive to variations of input, including variations in fields that are imaged, availability of spectroscopy, or the weather conditions. The discovery of the accelerating universe presents us with a case in which a robust outcome is produced not through a mere convergence of measurements, but as a result of a robust procedure.[45] The robust detection claim is made on the basis of data obtained from a robust procedure that underlies the variations introduced into the data analysis. A robust data acquisition procedure in supernova cosmology provided high quality supernova data in sufficient numbers in a timely manner, which was essential for the program to succeed.[46] It was on the basis of this data acquired with a robust procedure, were the supernova cosmologists justified in claiming that the consistency resulting from their variations of data-subsets, or analysis methods, rendered their detection robust.

We thus see that the data acquisition procedures that are described in the key evidential papers of both teams had to be developed through stages to become more and more reliable. Development of a robust data acquisition procedure made the scientists confident in their data, and justified their trust in the robustness claims they made on the basis of them.

5 The Post "Discovery" Papers, and the Confirmation of the Result

Even though the statistical significance achieved in the measurements played a salient role in the primary evidence papers, neither of the teams made the claim that their results were conclusive. As we have seen above, after dealing with the statistical errors, (Riess

---

[45](Basso, 2017) analyses measurements of migraine prevalence and material poverty by social and behavioral scientists, and argues that scientists can avoid basing their robustness claim on a "simple convergence of the measurement results," by examining the "fit between the expected and actual convergence" of their results (Basso, 2017, 58). This is compatible with my account, as the construction of the robust data acquisition procedure requires an endorsed measurement system, which, in turn, depends on assessing the "fit" between the system and the target phenomena. In each case, a robustness claim requires the assessment of whether a measurement system behaves *as it should.*

[46]In a paper examining the issue of interpreting and validating raw-data produced by a new technique, Gandenberger identifies an argumentative strategy that can be seen as aimed at constructing a robust data acquisition procedure. This strategy, which he refers to as "direct causal inference," was used by the scientists Erlanger and Gasser to demonstrate that their cathode-ray oscillograph "could follow rapidly changing voltages with fidelity" (Gandenberger, 2010, 388). In order to test this, the scientists varied the voltages applied to the apparatus, using AC currents, constant voltages, or brief voltage spikes (Gandenberger, 2010, 388-389). Similarly, in another test of the reliability of their apparatus, scientists varied the pressure applied to nerves and recorded the resulting voltage changes in the nerve (Gandenberger, 2010, 390). In each of these cases, that the outcome of the tests were "as expected constituted a successful experimental check of their apparatus for rapidly changing voltages" (Gandenberger, 2010, 388). Even though Gandenberger situates direct causal inference within the general context of experimental strategies that Franklin refers to as "the epistemology of experiment," introducing variations in inputs to check whether the apparatus functions as intended indicates that a robust procedure strategy was in play as well (Franklin, 2002, 2-6).

et al., 1998) further asked how reliable their conclusion is. Examining the possible sources of systematic error, the paper concluded that "[f]urther studies are needed to determine the possible influence of any remaining systematic uncertainties" (Riess et al., 1998, 1034). Similarly, the SCP group admitted that a "conspiracy" of systematic effects being responsible for the result was "unlikely but not impossible" (Perlmutter et al., 1999, 582).

In line with this approach, both teams continued to probe the result they obtained. These studies began immediately after the publication of the acceleration result. They had two primary aims: to re-analyze the existing data set (as well as adding new objects to it to further increase the confidence in the result,) and to search for ways of testing whether possible systematic errors that could not have been ruled out in the earlier experiment influenced the result. To this end, the SCP group published a paper in 2003 aiming to provide "an independent set of high-redshift SNe that *confirms previous SN evidence for an accelerating universe*" (Knop et al., 2003, 102, emphasis added). This paper also aimed to check further the possible systematic error of dust extinction with higher precision, and stated its conclusion as follows:

> The precision of the measurements is such that it is possible to perform a host galaxy extinction correction directly for individual SNe without any assumptions or priors on the parent $E(B - V)$ distribution. Our cosmological fits using full extinction corrections confirm that dark energy is required with $P(\Omega_\Lambda > 0) >$ 0.99, a result *consistent* with previous and current SN analyses that rely on the identification of a low-extinction subset or prior assumptions concerning the intrinsic extinction distribution (Knop et al., 2003, 102, emphasis added).

Hence, three years after their original announcement of the acceleration, the SCP still found it necessary to provide further confirmation for the earlier result. Note also that the paper stated that the new results were consistent with extant work, which relied on different assumptions, or used less than ideal data.

Similarly, the High-z group also continued to work on the question of confirming the earlier result. They published three papers (co-authored by Adam Riess and several other members of the collaboration, as well as other scientists who were not originally part of the High-z) that aimed to test the systematic effects that might be influencing the result. The first of these papers, entitled "Tests of the Accelerating Universe With Near-infrared Observations of a High-redshift Type Ia Supernova," was published in 2000 (Riess et al., 2000). This paper is devoted to testing a particular model of dust extinction, namely the *intergalactic gray dust scenario*, that could dim the supernovae light in a specific way that resembles a cosmological constant (Riess et al., 2000, 66). As a result of the near-infrared observations of the single supernova SN 1990Q, the paper argued that the gray dust scenario was disfavored, but the authors were still cautious concerning the acceleration result being fully confirmed, ending the paper with the following remark:

> Continuing studies of high-z SNe Ia should ultimately provide the extraordinary evidence required to accept (or refute) the accelerating universe (Riess et al., 2000, 67).

Riess continued to seek this extraordinary evidence. A year later, he co-authored another paper devoted to testing the acceleration result. One way to distinguish the accelerating universe from supernovae appearing dim due to dust is to observe supernovae at even larger redshifts than the high redshift data sets both group's used in their primary evidence papers. As the very early universe was matter dominated, supernovae at very high redshift should appear brighter indicating an early decelerating phase. On the other hand, if dust extinction causes supernovae to appear dim, the very high redshift supernovae should not exhibit any relative brightness. In his 2001 paper, Riess examined a supernova at the redshift of $z = 1.7$ and on the basis of this object, argued that the results were

> inconsistent with simple evolution or gray dust, the two most favored astrophysical effects which could mimic previous evidence for an accelerating universe from SNe Ia at $z \approx 0.5$ (Riess et al., 2001, 67).

Still, observation of a single supernova was not suffcient to confidently assert that the universe was accelerating. Riess and his co-authors kept their cautious stance, writing that "observations of more SNe Ia at $z < 1$ are needed to test more complex challenges to the accelerating universe hypothesis..." (Riess et al., 2001, 67). Consequently, Riess "led a proposal in 2001 to ... find and measure half-a-dozen Type Ia supernovae at z's greater than 1..." (Riess, 2011). The outcome of this research, which discovered 16 objects, 6 of which had redshifts higher than 1.25, was published in 2004 (Riess et al., 2004). In the conclusion of this paper, we finally find a confident assertion: "The data reject *at high confidence* simple, monotonic models of astrophysical dimming that are tuned to mimic the evidence for acceleration at $z \approx 0.5$" (Riess et al., 2004, 686). With this new data, the paper provided an updated contour plot, which showed more stringent limits on the parameters, compared to the earlier "evidence paper" data set (Fig. 7.) It was only after many checks with new data sets that the scientists were ready to affirm the acceleration result.

6   Robustness and Robust Consistency

In both of the teams' primary evidence papers, we have seen the argument for acceleration is based on the *robustness* of the results. This means that both groups demanded the results they obtained to be invariant over the particular data set, fitting procedure, or statistical method used in the measurement. Recall that, both groups made more than one measurement of the cosmological parameters. For example, the High-z used two light-curve fitting methods and compared the results. The SCP, on the other hand, had only one light-curve fitting method, but they applied various different analyses to their data. Furthermore, the groups required that this invariance of the outcome of different measurements did not disappear when combined variations on the data sets or analysis methods were introduced. In other words, the scientists created several combinations of their sub-samples and analysis methods.

Thus, even though SCP used a single method for their light curve fits, they employed twelve different data sets to fit to the model. And they applied both Bayesian and frequentist

Figure 7: High-z Post-Discovery Contour Plot. Note the iterative improvement of the contours over those of Fig. 4 and Fig. 5. (Riess et al., 2004, 678). Reproduced by permission of the AAS.
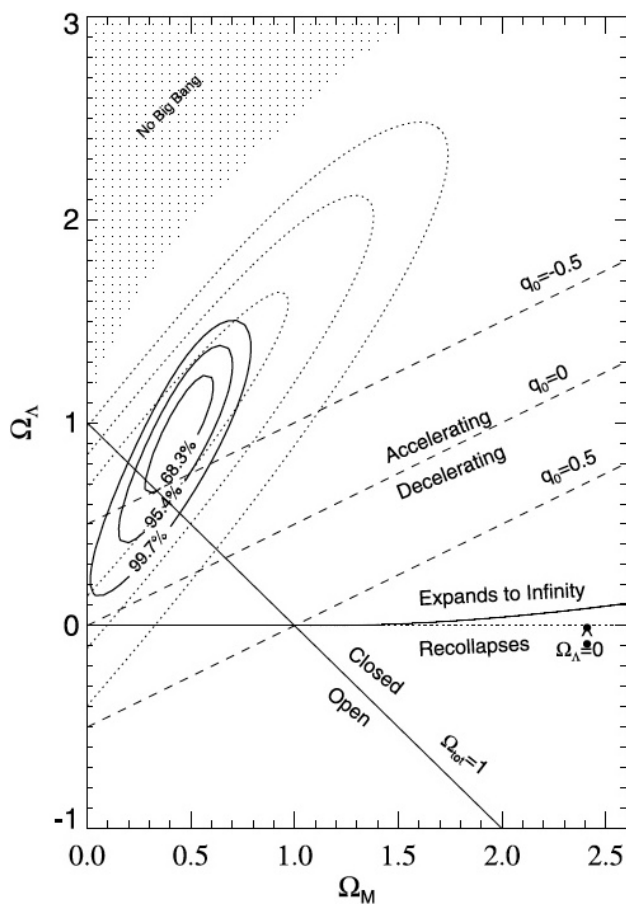
Figure 8: SCP Post-Discovery Robust Consistency Fits. This plot compares its results (indicted as "this paper") with both teams' earlier results. (Knop et al., 2003, 120). Reproduced by permission of the AAS.
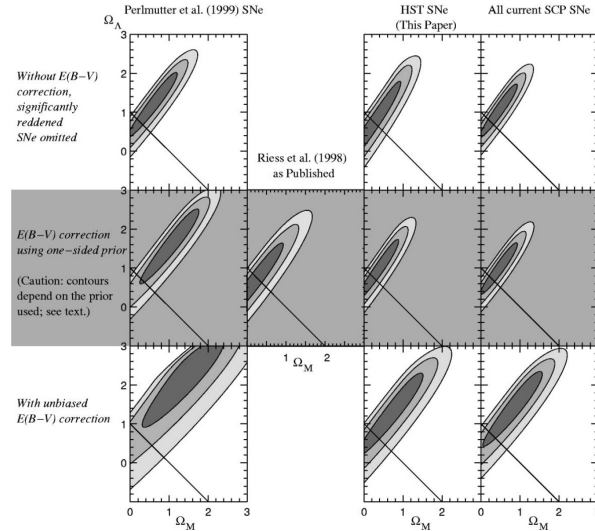

FIG. 9.—Plot of 68.3%, 95.4%, and 99.7% confidence regions for Ω_M and Ω_Λ using different data subsets and methods for treating host galaxy extinction corrections. The top row represents our fits to the low-extinction primary subset, where significantly reddened SNe have been omitted and host galaxy extinction corrections are not applied. The second row shows fits where extinction corrections have been applied using a one-sided extinction prior. These fits

analysis methods to these different data sub-samples. In total, 19 of all these 20 different fits gave a $\Omega_\Lambda > 0$ result with the confidence $p(\Omega_\Lambda > 0) = \%99$ and the remaining one with $\%89$. Thus, the results were robust under different combinations of analysis methods, and data sets. On the other hand, while (Riess et al., 1998) built the robustness claim on the agreement of two methods of light curve fitting, it also fitted a total of 4 different sets of supernova with these two methods, which all led to at least $p(\Omega_\Lambda > 0) = \%99.5$ confidence.

Finally, scientists continued to apply this methodology even after publishing their primary evidence papers. In their post-discovery papers, both the SCP and the High-z members remeasured the cosmological parameters (respectively (Knop et al., 2003) and (Riess et al., 2004)) and explicitly referred to the *robustness* of their results. For example, in the SCP post-discovery paper, one reads:

> This new set of SNe observed with the HST [Hubble Space Telescope] confirm and strengthen previous SN evidence for an accelerating universe and show that those results are robust even when host galaxy extinction is fully accounted for (Knop et al., 2003, 128).

The group produced new cosmology fits with "different data subsets and methods for treating host galaxy extinction corrections," all converging to the same result of a positive lambda (Fig. 8) (Knop et al., 2003, 120).

Note that there are two kinds of robustness claims to be distinguished here. On the one hand, there is the individual robustness claim made by each team separately. Building on a robust data acquisition procedure, each team claimed their measurements to be robust

Figure 9: SCP Plot for the Cosmological Parameters combining the SCP and High-z (HZT) Data. Source: (Knop et al., 2003, 121). Reproduced by permission of the AAS.
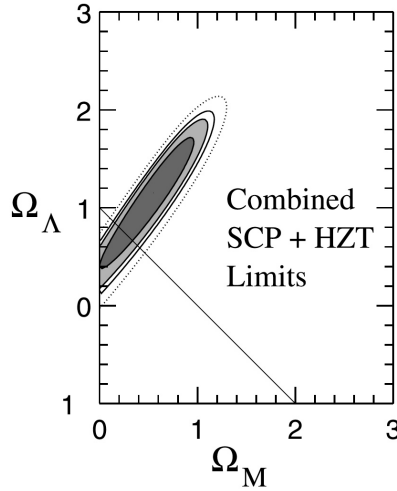

FIG. 10.—Plot of 68%, 90%, 95%, and 99% confidence regions for Ω_M and Ω_Λ, combining the high-redshift data of the SCP (this paper and P99) and Riess et al. (1998). The fit includes subset 2 SNe from the SCP plus the nine well-observed confirmed SNe Ia from Riess et al. (1998).

on the basis of the convergence of the results they obtained by varying the data sets, or analysis methods. On the other hand, there is the robustness claim stemming from the fact that the results of the two teams agreed.[47] Even though this latter convergence was not put forward as supporting the acceleration result by either of the teams in their evidence papers, the (Riess et al., 1998) data set was used in (Knop et al., 2003) in a combined analysis of the two teams' data (Fig. 9). The combined analysis also implied a positive cosmological constant, showing the full consistency of the two teams' results.

To emphasize this distinction between these two forms of robustness, I will refer to the latter case as the *robust consistency requirement.* Robust consistency ensues when several experiments, each with robust data acquisition procedures, produce consistent results.[48] Thus, what made the two groups' result *believable* was not one specific measurement by itself, but the coherence of many individually robust measurements taken together. Accordingly, both teams' papers produced several measurements of the cosmological parameters, and demonstrated to the astronomical community that these results were the outcome of a robust data

---

[47]Recall that Adam Riess also alluded to this distinction in his interview when he compared the two light-curve fitting methods of the High-z giving consistent results, with both teams obtaining the same result.

[48]An example to illustrate this distinction between *robustness* and *robust consistency* is the contemporary measurements of the rate of expansion of the universe, known as the *Hubble constant.* Here several distinct and robust measurement protocols, including supernova, cosmic microwave background, and gravitational lensing measurements, give inconsistent results, at the 4.4σ level (Riess et al., 2019). Instead of choosing the "most reliable" of these measurements, and declaring the Hubble constant problem to be solved, scientists talk about a "Hubble tension," and suspect that new physics might be lurking behind the discrepancy (ibid.) This indicates that for cosmologists consistency between individually robust measurements is required before a measurement result can be deemed conclusive.

acquisition procedure, and they agreed not only with their own previous measurements, but also with the other group's results. The accelerating universe result was not established by a single *most reliable* measurement, but through the *consistency* of many robust measurements. Finally, note also that these iterative measurements were all taken in endorsed measurement systems, which, in turn, were constructed iteratively.

How does robust consistency differ from robust detection? As mentioned above, in Calcott's characterization, robust detection of an empirical claim occurs when there exists multiple, independent ways of detecting or verifying it. The examination of the accelerating universe case demonstrates that a robust result obtained from variations within a *single* measurement protocol, and a robust result obtained from the consistency of results obtained from *multiple* measurement protocols have differing epistemic import. Whereas the former is actively used by the experimentalists to put forward an evidence claim, the latter fulfills a social epistemological function of augmenting the trust of the general scientific community in the validity of the result. Yet the concept of robust detection as defined by Calcott is too coarse-grained to capture this difference, as it does not characterize how the multiplicity of detections is achieved. In the context of the present case study, I propose to reserve the concept of robust detection for the former case (i.e., the individual robust results by each team,) in which the multiplicity of detections is achieved within a given measurement protocol. In contrast, robust consistency describes situations in which several measurement protocols agree on a result.[49]

Even though an in-depth analysis of the rationale underlying the the various robustness claims scientists employed in this case is beyond the scope of this paper, a few preliminary points can be made. Among the several accounts available in the current literature on robustness, Kuorikoski and Marchionni argue that the epistemic rationale of a robustness argument lies in demonstrating that different methods or procedures that scientists employ "do not share the same kind of error or bias" (Kuorikoski and Marchionni, 2016, 241). On the other hand, Basso sees the value of robustness reasoning in its ability to enhance reliability by evaluating the "fit between actual and expected convergence" of result obtained from different procedures (Basso, 2017, 58). Basso emphasizes that it is the complementarity of the sources of uncertainty (as opposed to their independence) that is critical in this evaluation. Here, complementarity means that no procedure has an uncertainty dominating the others'. Another author who analyses robustness in terms error and uncertainty is Staley, who sees robustness analysis as "a strategy for assessing the security of an evidence claim by investigating classes of potential error scenarios..." (Staley, 2018, 13). According to Staley, one uses robustness analysis to rule out error scenarios that "are not compatible with a given evidential conclusion" (Staley, 2018, 13). Finally, another influential view is provided by Schupbach, who argues that robustness analysis enables scientists to rule out "competing

---

[49]Another example that illustrates the distinction is the 1995 discovery of the top quark. There were two experimental teams involved in the search for the top quark, namely, the Collider Detector at Fermilab (CDF), and the DZero (DØ), collaborations. As (Staley, 2004a) demonstrates, the CDF group resorted to robustness arguments as part of their evidence claim for the top quark (robust detection.) This is distinct from the fact that both teams reported evidence for the top quark on the basis of their separate data, and methods of analysis (robust consistency.)

potential explanations" of the result they aim to validate (Schupbach, 2018, 287).

In the primary evidence papers of the SCP and the High-z teams, we do find variations introduced by the scientists to rule out competing explanations, as Schupbach describes. For example, introducing variations into the metric expression in the data analysis can be seen as an attempt to rule out the alternative explanation that the dimming of the supernovae is caused by gravitational lensing. However, for the majority of the variations introduced, the rationale of the supernova scientists seems to be best captured by Staley's account. The variations incorporated by the SCP, which include stretch correction, extinction checks, addition or subtraction of magnitude offsets to mimic unknown systematic effects, or analyzing data with and without a Bayesian prior etc., are all introduced to identify and eliminate possible error scenarios and biases. For example, in checking (via a robustness argument) for the systematic error that can be introduced by supernova evolution, the group wrote:

> *If the same cosmological results are found for each measurement* based on a subset of low- and high-redshift supernovae sharing a given host-galaxy classification, *we can rule out many evolutionary scenarios* (Perlmutter et al., 1999, 579, emphasis added).

The group treated these evolutionary scenarios as possible sources of systematic uncertainty, and not as competing explanations on a par with the accelerating universe hypothesis. Similarly, in the High-z case, the group ruled out possible error scenarios including evolution, extinction, selection bias and gravitational lensing, using variations in the data set, or the light curve fitting method, as explained above.[50] Thus, for both groups, the rationale of resorting to considerations of robustness seems to be the intention to eliminate possible sources of error.

In my discussion of the two teams' construction of a robust data acquisition procedure, I noted its compatibility with Basso's account. It therefore seems that several forms of robustness reasonings are involved in different aspects of the argument for the accelerating universe. A more detailed analysis than can be given here (incorporating several accounts) is necessary to deal with this question in a satisfactory manner.

7 Conclusion

In this article, I argued that epistemic iteration can account for the supernova evidence for the accelerating universe. I introduced a particular form of iterative progress, namely, endorsement, and argued that, in order to validate experimental results, supernova scientists needed to build and endorse a measurement system through various iterative means. I illustrated these ideas through a historical examination of the discovery based on oral-history interviews, and textual analysis of published material.

As I have shown, both of the teams that established the acceleration of the universe did so by developing an endorsed measurement system, or a *measurement protocol.* Once

---

[50]See Section (3.5.2).

the protocol was in place, scientists conducted measurements with a robust data acquisition procedure, and the result was confirmed due to the consistency of this set of robust results. As we have seen above, only at the end of such procedure, a confident assertion of an experimental fact appears.

To what extent is this account generalizable to other cases of experimental validation? Since the pioneering work of Ian Hacking, who argued that experiment "has a life of its own," many case studies of experiment have been produced (Hacking, 1983, 150). These studies range from the panoramic approaches analyzing the social and epistemic aspects of belief formation in an experimental community ((Galison, 1987), (Collins, 2004), (Steinle, 2016)), to more focused approaches that analyze the historical development of the argumentation of a primary evidence paper, as in (Staley, 2004b), or the epistemology of a single experiment (Franklin, 1986, Chapter 3).

What I aimed to accomplish in this article was to steer a middle ground between these two approaches, and examine whether experimental validation has structural elements that can be discerned via a historical study of the formation of a measurement protocol. This approach is in line with more recent work in the philosophy of science in practice, and in my paper I drew on, or indicated accordances with, case-studies from fields as diverse as plant biology, particle physics, and neurophysiology. In the case of the discovery of the accelerating universe, I identified the iterative construction of a measurement protocol, and the requirement of several measurements to ensure robust consistency to be such elements. It is to be noted that experiments in astronomy are non-interventional, and the structure of experimental validation in astronomy may not be immediately generalizable to more familiar laboratory contexts due to this reason. Hence, further comparative studies are needed to assess whether the present account could be adapted to other cases as well. In an article examining the experimental practices of neuroscientists, Jacqueline Sullivan argued that the diversity exhibited by the experimental protocols in neuroscience gives us good reasons to doubt the possibility of a global account of neuroscientific explanations (Sullivan, 2008). In a similar vein, it would be too optimistic to hope that an account derived from a supernova cosmology measurement to be readily applicable in other cases. Still, a common aspect of the studies that I argued to be in accordance with my treatment of the accelerating universe case is that they all focus on the processes of data production, assessment and analysis. Hence, to the extent that cases of experimental validation across disciplines is found mainly to depend on these *practices of data,* categories such as endorsement and robust consistency might prove to be prevalent beyond the domain of empirical cosmology.

References

Abbott, T. M. C. et al. (2018). Dark Energy Survey year 1 results: Cosmological constraints from galaxy clustering and weak lensing. *Phys. Rev. D*, 98:043526.

APJ. The Astrophysical Journal Letters. `http://iopscience.iop.org/journal/2041-8205`. Accessed: 2018-12-318.

Barwich, A.-S. and Chang, H. (2015). Sensory measurements: Coordination and standardization. *Studies in History and Philosophy of Science Part A*.

Basso, A. (2017). The appeal to robustness in measurement practice. *Studies in History and Philosophy of Science Part A*, 65-66:57 – 66. The Making of Measurement.

Brandom, R. B. (1998). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

Burian, R. M. (1997). Exploratory Experimentation and the Role of Histochemical Techniques in the Work of Jean Brachet, 1938-1952. *History and Philosophy of the Life Sciences*, 19(1):27–45.

Calcott, B. (2011). Wimsatt and the robustness family: Review of wimsatt's re-engineering philosophy for limited beings. *Biology and Philosophy*, 26(2):281–293.

Chang, H. (2004). *Inventing Temperature: Measurement and Scientific Progress*. Oxford University Press, Oxford New York.

Chuang, C.-H., Prada, F., Cuesta, A. J., Eisenstein, D. J., Kazin, E., Padmanabhan, N., Sànchez, A. G., Xu, X., Beutler, F., Manera, M., Schlegel, D. J., Schneider, D. P., Weinberg, D. H., Brinkmann, J., Brownstein, J. R., and Thomas, D. (2013). The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: single-probe measurements and the strong power of $f(z)\sigma 8(z)$ on constraining dark energy. *Monthly Notices of the Royal Astronomical Society*, 433(4):3559–3571.

Collins, H. (2004). *Gravity's Shadow: The Search for Gravitational Waves*. University of Chicago Press.

Franklin, A. (1986). *The Neglect of Experiment*. Cambridge University Press.

Franklin, A. (2002). *Selectivity and Discord: Two Problems of Experiment*. University of Pittsburgh Press, Pittsburgh, Pa.

Franklin, L. (2005). Exploratory Experiments. *Philosophy of Science*, 72(5):888–899.

Galison, P. (1987). *How Experiments End*. University of Chicago Press.

Gandenberger, G. S. (2010). Producing a robust body of data with a single technique*. *Philosophy of Science*, 77(3):381–399.

Garnavich, P. M., Kirshner, R. P., Challis, P., Tonry, J., Gilliland, R. L., Smith, R. C., Clocchiatti, A., Diercks, A., Filippenko, A. V., Hamuy, M., Hogan, C. J., Leibundgut, B., Phillips, M. M., Reiss, D., Riess, A. G., Schmidt, B. P., Schommer, R. A., Spyromilio, J., Stubbs, C., Suntzeff, N. B., and Wells, L. (1998). Constraints on Cosmological Models from Hubble Space Telescope Observations of High-z Supernovae. *Astrophysical Journal Letters*, 493:L53–L57.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science.* Cambridge University Press.

Hamuy, M., Maza, J., Phillips, M. M., Suntzeff, N. B., Wischnjewsky, M., Smith, R. C., Antezana, R., Wells, L. A., Gonzalez, L. E., Gigoux, P., Navarrete, M., Barrientos, F., Lamontagne, R., della Valle, M., Elias, J. E., Phillips, A. C., Odewahn, S. C., Baldwin, J. A., Walker, A. R., Williams, T., Sturch, C. R., Baganoff, F. K., Chaboyer, B. C., Schommer, R. A., Tirado, H., Hernandez, M., Ugarte, P., Guhathakurta, P., Howell, S. B., Szkody, P., Schmidtke, P. C., and Roth, J. (1993a). The 1990 Calan/Tololo Supernova Search. *The Astronomical Journal*, 106:2392–2407.

Hamuy, M., Phillips, M. M., Wells, L. A., and Maza, J. (1993b). K Corrections for Type IA supernovae. *Publications of the Astronomical Society of the Pacific*, 105:787–793.

Hansen, L., Jorgensen, H. E., and Norgaard-Nielsen, H. U. (1987). Search for Supernovae in Distant Clusters of Galaxies. *The Messenger*, 47:46–49.

Hudson, R. G. (1999). Mesosomes: A study in the nature of experimental reasoning. *Philosophy of Science*, 66(2):289–309.

Hudson, R. G. (2009). The methodological strategy of robustness in the context of experimental wimp research. *Foundations of Physics*, 39(2):174–193.

Hudson, R. G. (2013). *Seeing Things: The Philosophy of Reliable Observation.* Oxford University Press.

Karaca, K. (2013). The Strong and Weak Senses of Theory-Ladenness of Experimentation: Theory-Driven versus Exploratory Experiments in the History of High-Energy Particle Physics. *Science in Context*, pages 93–136.

Karaca, K. (2018). Two Senses of Experimental Robustness: Result Robustness and Procedure Robustness. *The British Journal for the Philosophy of Science.*

Knop, R. A., Aldering, G., Amanullah, R., Astier, P., Blanc, G., Burns, M. S., Conley, A., Deustua, S. E., Doi, M., Ellis, R., Fabbro, S., Folatelli, G., Fruchter, A. S., Garavini, G., Garmond, S., Garton, K., Gibbons, R., Goldhaber, G., Goobar, A., Groom, D. E., Hardin, D., Hook, I., Howell, D. A., Kim, A. G., Lee, B. C., Lidman, C., Mendez, J., Nobili, S., Nugent, P. E., Pain, R., Panagia, N., Pennypacker, C. R., Perlmutter, S., Quimby, R., Raux, J., Regnault, N., Ruiz-Lapuente, P., Sainton, G., Schaefer, B., Schahmaneche, K., Smith, E., Spadafora, A. L., Stanishev, V., Sullivan, M., Walton, N. A., Wang, L., Wood-Vasey, W. M., and Yasuda, N. (2003). New Constraints on Ω_M, Ω_Λ, and w from an Independent Set of 11 High-Redshift Supernovae Observed with the Hubble Space Telescope. *Astrophysical Journal*, 598:102–137.

Kuorikoski, J. and Marchionni, C. (2016). Evidential diversity and the triangulation of phenomena. *Philosophy of Science*, 83(2):227–247.

Leonelli, S. (2019). What distinguishes data from models? *European Journal for Philosophy of Science*, 9(2):22.

Malmquist, K. G. (1922). On Some Relations in Stellar Statistics. *Meddelanden fran Lunds Astronomiska Observatorium Serie I*, 100:1–52.

Mayo, D. (2005). *Evidence as Passing Severe Tests: Highly Probable versus Highly Probed Hypotheses*, pages 95–128. Johns Hopkins University Press, Baltimore and London.

NobelPrize.org. Advanced Information. NobelPrize.org. Nobel Media AB 2019. `https://www.nobelprize.org/uploads/2018/06/advanced-physicsprize2011-1.pdf`. Accessed: 09-15-2019.

Norgaard-Nielsen, H. U., Hansen, L., Jorgensen, H. E., Aragon Salamanca, A., and Ellis, R. S. (1989). The Discovery of a Type IA Supernova at a Redshift of 0.31. *Nature*, 339:523–525.

Nugent, P., Kim, A., and Perlmutter, S. (2002). K-Corrections and Extinction Corrections for Type Ia Supernovae. *Publications of the Astronomical Society of the Pacific*, 114:803–819.

Perlmutter, S., Aldering, G., della Valle, M., Deustua, S., Ellis, R. S., Fabbro, S., Fruchter, A., Goldhaber, G., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Knop, R. A., Lidman, C., McMahon, R. G., Nugent, P., Pain, R., Panagia, N., Pennypacker, C. R., Ruiz-Lapuente, P., Schaefer, B., and Walton, N. (1998). Discovery of a Supernova Explosion at Half the Age of the Universe. *Nature*, 391:51.

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R., Nugent, P., Castro, P., Deustua, S., Fabbro, S., Goobar, A., Groom, D., et al. (1999). Measurements of Ω and Λ from 42 High-redshift Supernovae. *The Astrophysical Journal*, 517(2):565.

Perlmutter, S., Gabi, S., Goldhaber, G., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Pain, R., Pennypacker, C. R., Small, I. A., Ellis, R. S., McMahon, R. G., Boyle, B. J., Bunclark, P. S., Carter, D., Irwin, M. J., Glazebrook, K., Newberg, H. J. M., Filippenko, A. V., Matheson, T., Dopita, M., and Couch, W. J. (1997). Measurements of the Cosmological Parameters Ω and Λ from the First Seven Supernovae at $z \geq 0.35$. *Astrophysical Journal*, 483:565–581.

Planck Collaboration, Aghanim, N., et al. (2018). Planck 2018 results. VI. Cosmological parameters. *ArXiv e-prints*.

Rheinberger, H.-J. (1997). *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube (Writing Science)*. Stanford University Press.

Riess, A. G. (2011). Nobel Lecture: My Path to the Accelerating Universe. `https://www.nobelprize.org/nobel_prizes/physics/laureates/2011/riess_lecture.pdf`. [Online; accessed October, 2017].

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., and Scolnic, D. (2019). Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond ΛCDM. *The Astrophysical Journal*, 876(1):85.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., and Tonry, J. (1998). Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *The Astronomical Journal*, 116(3):1009.

Riess, A. G., Filippenko, A. V., Liu, M. C., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., Tonry, J., Woudt, P., Brunner, R. J., Dey, A., Gal, R., Graham, J., Larkin, J., Odewahn, S. C., and Oppenheimer, B. (2000). Tests of the Accelerating Universe with Near-Infrared Observations of a High-Redshift Type IA Supernova. *The Astrophysical Journal*, 536:62–67.

Riess, A. G., Nugent, P. E., Gilliland, R. L., Schmidt, B. P., Tonry, J., Dickinson, M., Thompson, R. I., Budavári, T., Casertano, S., Evans, A. S., Filippenko, A. V., Livio, M., Sanders, D. B., Shapley, A. E., Spinrad, H., Steidel, C. C., Stern, D., Surace, J., and Veilleux, S. (2001). The Farthest Known Supernova: Support for an Accelerating Universe and a Glimpse of the Epoch of Deceleration. *The Astrophysical Journal*, 560:49–71.

Riess, A. G., Strolger, L.-G., Tonry, J., Casertano, S., Ferguson, H. C., Mobasher, B., Challis, P., Filippenko, A. V., Jha, S., Li, W., Chornock, R., Kirshner, R. P., Leibundgut, B., Dickinson, M., Livio, M., Giavalisco, M., Steidel, C. C., Benítez, T., and Tsvetanov, Z. (2004). Type Ia Supernova Discoveries at $z > 1$ from the Hubble Space Telescope: Evidence for Past Deceleration and Constraints on Dark Energy Evolution. *The Astrophysical Journal*, 607:665–687.

Sandage, A. (1961). The Ability of the 200–INCH Telescope to Discriminate Between Selected World Models. *Astrophysical Journal*, 133:355.

Schmidt, B. (2007). Interview of Brian Schmidt by Ursula Pavlish. Niels Bohr Library & Archives, American Institute of Physics, College Park, MD USA. Conducted at the California Institute of Technology, July 25.

Schmidt, B. P., Suntzeff, N. B., Phillips, M. M., Schommer, R. A., Clocchiatti, A., Kirshner, R. P., Garnavich, P., Challis, P., Leibundgut, B., et al. (1998). The High-z Supernova Search: Measuring Cosmic Deceleration and Global Curvature of the Universe Using Type Ia Supernovae. *The Astrophysical Journal*, 507(1):46.

Schupbach, J. N. (2018). Robustness Analysis as Explanatory Reasoning. *British Journal for the Philosophy of Science*, 69:275 – 300.

Soler, L., Nickles, T., Trizio, E., and Wimsatt, W. C. (2012). *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science.* Dordrecht: Springer.

Spitzer, Jr., L. (1979). History of the Space Telescope. *Quarterly Journal of the Royal Astronomical Society*, 20:29.

Staley, K. (2004a). Robust Evidence and Secure Evidence Claims. *Philosophy of Science*, 71(4):467–488.

Staley, K. (2004b). *The Evidence for the Top Quark: Objectivity and Bias in Collaborative Experimentation.* Cambridge University Press, Cambridge, New York.

Staley, K. (2018). Securing the Empirical Value of Measurement Results. *The British Journal for the Philosophy of Science.*

Stegenga, J. and Menon, T. (2017). Robustness and Independent Evidence. *Philosophy of Science*, 84:414 – 435.

Steinle, F. (1997). Entering New Fields: Exploratory Uses of Experimentation. *Philosophy of Science*, 64, Supplement. Proceedings of the 1996 Biennial Meetings of the Philosophy of Science Association:S65–S74.

Steinle, F. (2016). *Exploratory Experiments: Ampère, Faraday, and the Origins of Electrodynamics.* University of Pittsburgh Press.

Sullivan, J. A. (2008). The multiplicity of experimental protocols: a challenge to reductionist and non-reductionist models of the unity of neuroscience. *Synthese*, 167(3):511.

Tal, E. (2016). Making Time: A Study in the Epistemology of Measurement. *The British Journal for the Philosophy of Science*, 67(1):297–335.

Tammann, G. A. (1979). Cosmology with the Space Telescope. In Macchetto, F., Pacini, and Tarenghi, M., editors, *ESA/ESO Workshop on Astronomical Uses of the Space Telescope.*

Wimsatt, W. C. (2007). *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality.* Harvard University Press.

Woodward, J. (2000). Data, phenomena, and reliability. *Philosophy of Science*, 67:S163–S179.