Robustness Analysis and Hubble Tension

Marco Forgione

Abstract

The paper presents and discusses the Hubble tension with respect to recent results in cosmology. I shall argue that the measurements from the James Webb Space Telescope and TRGB stars calibrations allow us to infer that the estimates of H_0 with late universe methods are robust. Building on from robustness analysis, I conclude that the resolution of the tension cannot be expected to come from new systematics, but rather from new physics.

1 Introduction

Friedmann equations are among the most important equations in modern cosmology. They describe the expansion of our universe —under the assumption that our universe is homogeneous and isotropic: the so-called cosmological principle. We can obtain a simplified derivation of the first of such equations using mostly Newtonian mechanics and some general relativistic corrections. All we need is the gravitational potential, the cosmological principle, and the principle of energy conservation.

Let us first consider Newton's shell theorem: for a spherically homogeneous object with radius r, the gravitational field inside the object at a distance r from the center is the same as if the total mass was concentrated at the object's center. That is: "in a spherically symmetric distribution of matter, a particle feels no force at all from the material at greater radii, and the material at smaller radii gives exactly the force one would get if all the material was concentrated at the central point" (Liddle 2015, p. 22). Then, consider an observer to be at the center of a uniform expanding universe (this is unproblematic because of the cosmological principle) with mass density ρ. The total mass is: $M = 4\pi\rho r^3/3$, and the contributing force with respect to a test particle of mass m at distance r is: $F = G\frac{Mm}{r^2} = G\frac{4\pi\rho r m}{3}$. Because of energy conservation, one obtains that $U = 1/2m\dot{r} - 4\pi\rho r^2 m/3$, which gives the evolution of the separation r between observer and test particle.

Because of the cosmological principle and homogeneity, we can change coordinates system to track the distance between particles in relation to the expansion of the universe so that: $\vec{r} = a(t)\vec{x}$. With the change of coordinates, the object remains fixed at its position, while the distance changes in proportion to a factor that depends on time only: the scale factor of the universe $a(t)$. By substituting the previous equation with the new coordinates system one obtains: $U = 1/2m\dot{a}x^2 - 4G\pi\rho a^2 x^2 m/3$ which gives: $(\frac{\dot{a}}{a})^2 = 8/3\pi G\rho + \frac{2U}{a^2x^2m}$. Now, the more appropriate derivation of the equation ought to consider relativistic effects, and thus the mass density is replaced by the total energy density $\epsilon(t)/c^2$, and one needs to add a term for the curvature of space. The GR form of the Friedmann equation is:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\frac{\epsilon(t)}{c^2} - \frac{kc^2}{a^2} \tag{1}$$

where $kc^2 = -2U/mx^2$ and k is a constant that represents the curvature of the universe and does not change with time or space.

As I mentioned earlier, Friedmann equations are among the most important equations in modern cosmology since they help us understand the expansion of the universe, that is, they help us understand how recession velocity of cosmological objects is proportional to the distance (a relation that was originally discovered by Hubble). The recession velocity is $\vec{v} = d\vec{r}/dt$ and has the same direction as $\vec{r}$, thus we can write: $\vec{v} = \frac{|\dot{\vec{r}}|}{|r|}\vec{r}$ and, since $\vec{r} = a\vec{x}$, where $\vec{x}$ is a constant, then $\vec{r} = \frac{\dot{a}}{a}\vec{r}$ (Liddle 2015). Hubble's law tells us that $\vec{v} = H\vec{r}$ and thus $H = \frac{\dot{a}}{a}$.

The term H is called Hubble constant, but to call it a constant is at least misleading. Indeed, the term (or parameter) needs not be constant in time (since $a(t)x$ depends on time) and it is thus more appropriate to talk about H_0 as the value of the Hubble parameter at present time. Notably, this latter consideration already responds to a possible objection to what we will see later on in this paper. One could maintain that since the Hubble constant is not indeed a constant, it can vary in time, and that such a variation could account for the apparent tension between measurements of H_0 that I will discuss below. However, since the calculations of Hubble's constant already consider the temporal parameter, they already account for the values of H_0

at the time of measurement, thereby ruling out the possibility of explaining different measured values of H_0 due to changes over time.

The Hubble constant is one of the parameters of the λ Cold Dark Matter (λCDM) model whose numerical values can be determined by observations only. Since H_0 corresponds to the expansion rate of the universe, it should be relatively easy to measure: one would measure as many velocities and distances of galaxies as possible by using the corresponding redshift of spectral lines, and then infer H_0. Unfortunately, this method does not distinguish between the velocity of galaxies due to the expansion of the universe, and due to, for example, the gravitational effects of nearby cosmological objects. To avoid the mixing of these two components (the Hubble expansion and the peculiar velocities) one can measure the same quantities for galaxies in the Hubble flow, where the Hubble velocity dominates over the peculiar velocity. However, since those measurements involve objects very far away, we need an accurate estimate of the distances of those galaxies.

Since geometric methods such as parallaxes are unfeasible for measuring the distance of galaxies in the Hubble flow, cosmologists had to resort to the use of standard candles —namely, objects that can be used to calibrate distances based on their luminosity. The process, as suggested in (Liddle 2015), is similar to determining that a light bulb that is a quarter as bright as another one is twice as far away (inverse square law). This process works as long as we know that all light bulbs have the same intrinsic luminosity, and that is why we cannot use just any cosmological objects, bur rather, we need objects whose intrinsic luminosity is well-known. Once we have found such objects, we can use them as calibrators to measure the distance of even further away galaxies, and thus calculate a value for H_0 that is not affected by the peculiar velocities. This method is called: the distance ladder method.

Another method to measure Hubble's constant is to calculate it from the Cosmic Microwave Background Radiation (CMB), which is a faint thermal radiation that reaches us from the Big Bang. Measures of H_0 from the CMB are oftentimes called indirect-measurements, and they give values that are lower than those calculated with the distance ladder method. As suggested in (Liddle 2015, p. 50): "[t]here is a mild tension between these measurements which is currently under investigation, but it seems safe to conclude that H_0 [h, in the original] is now quite accurately measured to be 70 [0.70 in the original] within at most a few percent [...] The long lasting problem of determining the overall scale of the universe is therefore essentially solved".

The above mentioned 'mild-tension' has puzzled scientists for many years and, despite what was claimed above, the problem is still been discussed in both cosmology and philosophy of science communities. In this paper, I shall argue that one can use the tools of robustness analysis to show that the tension cannot be solved by appealing to some systematics, but rather could be indicative of a need for new physics. In the next section, I will briefly review the history of the H_0 tension, then, in Section 3, I will review the methods used to calculate the Hubble constant from late universe, that is, the distance ladder method. In Section 4 I will discuss some recent literature in cosmology addressing the problem between late and early universe measurements of the Hubble constant. In Section 5, I will present an argument that favors a solution to the Hubble tension based on systematics and recent measurements of TRGBs stars. There, I will use the literature on robustness and some recent analysis on the TRGBs to come to the conclusion that systematics are not a viable solution to appease the tension, but rather, we need to look for new physics and improvements on our cosmology models.

2 Historical Overview

The relationship between redshift (the velocity displacement) and galaxy distance was already demonstrated in (Hubble 1925) and the initial estimate was of $500km \cdot s^{-1}Mpc^{-1}$. However, such a large value for H_0 implied that the age of the universe was about 2 billion years, which, even at that time, was a number not sufficient to account for the astronomical observations within our solar system. We jump a few years ahead, and in 1950s, thanks to the observations of the Hale telescope, Humason, Mayall, and Sandage (1979) calculate $H_0 \sim 180km \cdot s^{-1}Mpc^{-1}$. But, shortly after, (Sandage 1958) placed the value of the Hubble constant between $50km \cdot s^{-1}Mpc^{-1}$ and $100km \cdot s^{-1}Mpc^{-1}$. In the following decade, the discussion about the discrepancy between such values continued, but, due to technological limitations, only little improvements were achieved. It was in 1970 that an updated review of the problem, and of different measurements, was presented by (Bergh 1970), who summarized the results of nine methods used at the time to calculate the H_0 constant. In those same years, the discovery of the Cosmic Microwave Background Radiation (CMB) validated the Big Bang model and it was assumed that there had to be some form of agreement with the inverse of the Hubble constant —and thus with the expansion rate of the universe. The assumption was made based on the necessary compatibility between the age of the universe and a model that describes its evolution. As most recently recalled in (Tully 2023), the 1970s and 1980s were characterized by a dispute between two classes of results: on the one hand there were those who supported a low-value H_0, with estimates that ranged between $50 km \cdot s^{-1} Mpc^{-1}$ and $57 km \cdot s^{-1} Mpc^{-1}$ with an error of ± 7 (see: (Sandage and Tammann 1974a), (Sandage and Tammann 1974b), (Sandage and Tammann 1975a). (Sandage and Tammann 1975b), (Sandage and Tammann 1982)). On the other hand, others (for example, (De Vaucouleurs 1985)) calculated H_0 to be between $90 km \cdot s^{-1} Mpc^{-1}$ and $110 km \cdot s^{-1} Mpc^{-1}$, where: "[h]e and colleagues built a ladder founded on primary novae, Cepheids, RR Lyrae, and horizontal branch stars, secondary and globular clusters, brightest blue and red supergiant stars, brightest HII loops of rings, and the velocity dispersions of HII regions, then a tertiary luminosity index calibration" (Tully 2023, p. 4). This became known as the 50-100 controversy.

In 1977, (Tully and Fisher 1977) calculated H_0 between $75 km \cdot s^{-1} Mpc^{-1}$ and $85 km \cdot s^{-1} Mpc^{-1}$ and they used a new method that made use of the correlation between the absolute luminosity of spiral galaxies and their rotational rate. The correlation will become known as the Tully-Fisher (TF) relation, which was then used by Aaronson to calculate H_0 between 90 and $95 km \cdot s^{-1} Mpc^{-1}$, see: (Aaronson et al. 1980), (Aaronson et al. 1982), (Aaronson et al. 1986). Shortly after, the discovery that distance information is encoded in the surface brightness fluctuation amplitude of dominantly old stellar populations resulted in the estimate of H_0 to be between $77 km \cdot s^{-1} Mpc^{-1}$ and $88 km \cdot s^{-1} Mpc^{-1}$ by: (Tonry and Schneider 1988), (Tonry et al. 1997), and (Tonry 1991). Thus far: "All of the methods [...] have been linked to the distance ladder that starts with stellar parallaxes and builds trough properties of stars in various parts of the Hertzprung-Russell diagram like the main sequence turnoff, red giant and horizontal branches, and variable RR Lyrae and cepheid stars" (Tully 2023, p. 7).

However, it was with the launch of the Hubble Space Telescope (HST) first, and Planck after, that the discrepancy between the values of H_0 became solidified. With respect to the former, two different teams were observing cepheids: on the one hand Sandage et al. obtained a value of $H_0 = 58 \pm 6 km \cdot s^{-1} Mpc^{-1}$ by using distances to type Ia Supernovae (SNIa), see: (Sandage et al. 1996). On the other hand, Wendy Freedman led the HST Key Project Study, which obtained $H_0 = 72 \pm 8 km \cdot s^{-1} Mpc^{-1}$ by using Cepheids, tip of the red giant branch, globular cluster, planetary nebulae, and surface brightness fluctuations observations, see: (Freedman et al. 2001). With Planck (2009-

2013), "the spatial power spectrum of fluctuations could be fit in impressive detail with a ΛCDM model" and obtained $H_0 = 67.4 \pm 0.5 km \cdot s^{-1} Mpc^{-1}$ (Tully 2023, p. 11). At that point, the difference between the distance ladder methods and CMB-based measurements was about 5σ.

Tully (2023) and others have hoped that the problem of the discrepancy of the H_0 values might be resolved if we had independent paths of confirmation for the distance ladder methods, that is, if we had an alternative to SNIa that can be applied at substantial redshifts. In other words, many believe, or perhaps hope, that unknown systematics are at the root of the H_0 tension problem.

3 Distance Ladder

Before moving on, I shall spend a few words on the so-called distance ladder method that delivered the late universe values of H_0. As I was mentioning earlier, to measure the recession velocity of galaxies we cannot use geometric methods such as parallaxes only. Indeed, we first need to find some (small) cosmological objects that are in the Hubble flow, whose luminosity is calculated from the theory, and that are bright enough to be detected by us. Since we currently lack the technology to perform a measurement of a distance that satisfies all such properties, scientists have invented a cosmic ladder method where each rung allows us to measure objects progressively farther away. So far so good. The problem is that each rung is also characterized by some systematic errors which get carried over to the next rung, and thus need to be accounted for and properly eliminated. In more technical terms: one needs to solve the expansion law: $c\delta\Lambda/\Lambda_0 = H_0 D_0$ where $\delta\Lambda/\lambda_0$ is the "redshift of the observed spectral lines of the galaxies compared to what would be expected only taking into account their distance" (Gueguen 2023, p. 35). To determine the distance, one needs to identify a standard candle, which is an object whose intrinsic luminosity (absolute magnitude) M is known.[1] On the other hand, the relative luminosity m (or relative magnitude) consists of the brightness of the object as it appears to us.

The cosmic distance ladder for the measurement of H_0 is usually based on three rungs (or steps). (1) The first rung, the absolute zero-point calibration, consists of measuring objects that typically live in the Milky Way, the Large

[1] The absolute luminosity of a an object is the luminosity we would measure if the object was 10 parsec away from us.

Magellanic Cloud (LMC), or NGC4528, and that are close enough to us to be measured by geometric methods such as trigonometric parallaxes. The second rung (2) consists of determining the absolute distance of galaxies that host Supernovae Type Ia (SNeIa) events and Cepheids stars that act as standard candles.

Cepheids have been the primary distance indicators because they are bright and reliably identifiable, since their mean luminosity depends on their pulsating periodicity (Leavitt and Pickering 1912). However, no standard candles are completely ideal since their luminosity can depend on factors such as metallicity and age. In addition, crowding and blending can also affect the precision of the distance measurements, especially for Cepheids beyond 20Mpc. One solution has been to quantify "the mean level of the local backgrouond due to brightness fluctuations using 'artificial stars' photometry" [where] artificial stars of pre-defined brightness may be randomly added to the images near Cepheids and their recovered magnitudes used to account for the mean background of real stars (Riess et al. 2023, p. 2). The result is an increase in accuracy of the mean magnitude of Cepheids, but a reduced precision for individual Cepheids. Nevertheless, the precision of Cepheids-based measurements has reached a substantial improvement over the years, as demonstrated in (Riess et al. 2019).

Another type of object that has been more recently used to determine the Hubble constant are the Tip of the Red Giant Branch (TGRB) stars. TRGB corresponds to the beginning of the helium flash (core helium burning) of a low-mass red giant (see: (Serenelli et al. 2017) for details). The theoretical background of red giant branch stellar evolution is well understood and constitutes a good empirically-based method to measure the distance of nearby galaxies. It is basically an alternative to the use of Cepheids. There are some advantages to the use of TRGB stars as standard candles, as reported in (Freedman et al. 2019): (1) TRGB stars are located in galaxy halos that suffer little reddening. (2) TRGB stars are quite isolated and thus are minimally affected by crowding and blending effects. (3) Metallicity can be identified and calibrated away (see, for example: (Mager, Madore, and Freedman 2008)). (4) The observation of TRGB stars is also more efficient, since Cepheids need many observations spread over time to determine periods, amplitudes and mean magnitude.

Finally, (3) the third step of the distance ladder method requires "high precision relative distances to a statistically significant sample of galaxies far enough into the Hubble flow so that their peculiar velocities are a small fraction of the cosmological recessional velocities (using SNe Ia)" (Freedman et al. 2019, p. 3). One point worth remarking here, and which shall come back later in this paper, is that the distance ladder method uses trigonometry and the internal working of some cosmological objects, but it does not rely on the λCDM model, unlike Barionic Acoustic Oscillations (BAO) and CMB based measurements.

4 Early and Late Universe Tension

Most recently, (Verde, Treu, and Riess 2019) have summarized a workshop held at Karli Institute for Theoretical Physics (15-17 July 2019).[2] The workshop discussed the discrepancies of H_0 values (see: Figure 1) and possible ideas to resolve the tension.

The most conservative approach to reconcile the discrepancy between measurements of H_0 from late and early universe is to invoke the existence of some systematic errors. With respect the distance ladder method, the data seem to point collectively towards a low value of H_0 and, as we shall see later, the use of observations from the Hubble Space Telescope (HST) have already lowered the uncertainties and thus the space for possible unknowns (Riess et al. 2019). Such space has been reduced even further with the observations from the James Webb Space Telescope (JWST) (Riess et al. 2023). With respect to possible systematics in the CMB predictions:

> [A]fter a thorough re-analysis and cross checks of multiple CMB observations [...] it is clear that systematic errors in CMB data cannot alone explain the tension [...] Moreover, a suite of low redshifts, different, truly independent measurements, affected by completely different possible systematics, agree with each other; it seems improbable that completely independent systematic errors affect all these measurements by shifting them all by about the same amount and in the same direction (Verde, Treu, and Riess 2019, p. 7).

Some hypotheses that have been suggested are, for example: (i) that an early dark energy component (yet to be found) is represented by a scalar field, but one would also need to verify that the additional component (and

[2] link: www.kitp.uscb.edu/activities/evervac-c19.

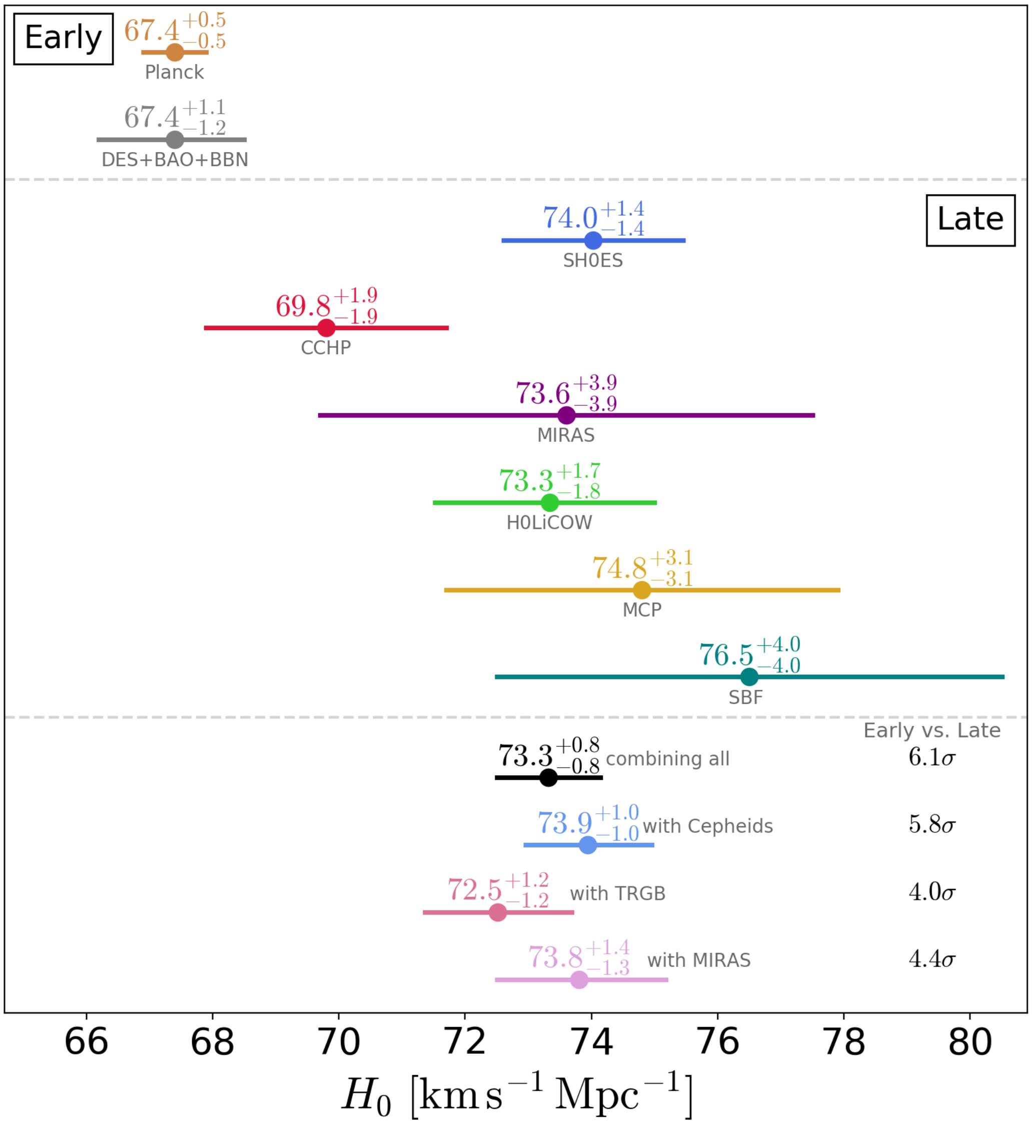

Figure 1: Measurements of Hubble constant (Verde, Treu, and Riess 2019, p. 10). Local distance ladder: SHOES (Cepheids) and SBF; CCHP (TRGBs); Miras. Early Universe measurements via CMB: Planck and BAO.

parameters thereof) is favored over the λCDM model (Poulin et al. 2019); (ii) one can extend the radiation sector of the physics of the early-universe, for example with some new neutrino physics, see: (Kreisch, Cyr-Racine, and

Doré 2020). Other alternatives that have been explored, and yet rejected, are: (1) the possibility that we live in an under-dense local void or bubble that could justify a different value of local H_0. The hypothesis was addressed by (Hoscheit and Barger 2018), even though the conclusions of their studies indicate that even if we were living in such a bubble, the effect would be too small to justify the discrepancy. (2) Another possibility is that the discrepancy is explained by some effects due to gravitational lensing. Again, the possibility has already been taken into consideration (Holz and Wald 1998), and the effects have been found to be too weak to justify the discrepancy, see: (Smith et al. 2013).

What are then the alternatives if we provisionally exclude some unknown systematics?

> The observed tension could be signaling additional fundamental new physics, beyond either the current astronomers' six-parameters λCDM model or the physicists' standard model of particle physics. At present, the dominant components of the standard model of cosmology are dark matter and dark energy, neither of which has a firm theoretical foundation (Freedman et al. 2019, p. 2).

With respect to the considerations on the measurements from CMB, these are highly model-dependent, in that they rely on the correctness of the λCDM model. Thus, there is a point to be made here, even though I shall not discuss it in this work: if new physics is what we should expect from the resolution of this tension, we would need to build a model that goes beyond λCDM, but the new model could not do worse than the previous one in describing other cosmological observations. If this is the case, then not much room is available to change the expansion history of the universe from that of λCDM, and that is because supernovae and baryonic acoustic oscillations provide quite the narrow constraints. Indeed: "[t]he early-universe H_0 determination relies on angular scales such as sound horizon [...] and matter-radiation equality. These angular scales are extremely well determined by CMB data, but they depend on a ratio of two qualitatively different quantities: the physical scale (which depends on early-time physics and background parameters) [...] and the angular size distance to the CMB (which depends on H_0 as well as other background parameters). To keep the angular scales fixed while increasing H_0, both the physical scales and the distance must decrease" (Verde, Treu, and Riess 2019, p. 8).

With respect to the late universe measurements via distance ladder methods, (Riess et al. 2019) determined H_0 from the observations collected from the Hubble Space Telescope from 70 Cepheids in the LMC. They argue that new observations with HST reduce the uncertainty of calibration of the Cepheids distance ladder method to 1.3%; and that the combination of LMC, masers in NG4528, and Milky Way parallaxes, gives $H_0 = 74.03 \pm 1.82 km \cdot s^{-1} Mpc^{-1}$, with a difference of 4.4σ with respect to the CMB observations (data inferred from Planck). The local value results of H_0 were determined by SH0ES team using luminosity of Cepheids, including NG4258 (Riess et al. 2016), eight detached eclipsing binaries, and with the most recent observations from HST scanning. They argue that while the resolution of the tension requires increasing precision and accuracy, thanks to HST "[...] we observed up to a dozen LMC Cepheids in three filters in a single orbit, obtaining HST photometry for 70 widely separated LMC Cepheids [...] This photometry establishes a new, zero-point-independent link between LMC Cepheids and those in the host of SNe Ia" (Riess et al. 2019, p. 2). The result is that "[w]hile it is difficult and perhaps debatable to identify the precise threshold at which a tension passes the point of being attributable to a fluke, the one presently involving H_0 appears to have passed that point. The higher, local value of H_0 from the distance ladder has been determined through five independent geometric absolute calibrations of the Cepheid P–L [pulse-luminosity] relation" (Riess et al. 2019, p. 11). The conclusion is that: "this discrepancy is not attributable to an error and thus cannot be attributed to systematics" (Riess et al. 2019, p. 1).

To take stock: the present problem (up to 2019) of the H_0 tension consists of the discrepancy between H_0 measures from local late universe $H_0 = 74 \pm 1.42$ (Riess et al. 2019) and those based on the CMB early universe $H_0 = 67 \pm 0.5$ (Collaboration et al. 2020). Between these two values, the estimation of H_0 based on TRGBs sits almost in between such values. Indeed, the Carnegie-Chicago Hubble Program (CCHP) provided an independent determination of H_0 based on TRGBs: "New and independent determination of the local value of the Hubble constant based on a calibration of the tip of the red giant branch (TRGB) applied to type Ia supernovae (SNe Ia). We find a value of H_0=69.8$\pm$ 0.8 ($\pm$1% stat) $\pm$1.7 ($\pm$2.4 sys) $km \cdot s^{-1} Mpc^{-1}$" (Freedman et al. 2019, p. 1).

As I mentioned earlier, there are several advantages of using TRGB stars over the use of Cepheids: low halo reddening, minimal crowding or blending, shallow metallicity effects, and no-need for multiple epochs observations.

Also, since the physics behind pulsating Cepheids and helium flash is unrelated, the two can be considered as entirely independent, despite both being calibrators for the distance ladder method.[3] I shall get back to the this in the final sections of this paper. As we shall see in the next section, the fact that not all calibrators used for the distance ladder methods allow for the same estimate of Hubble's constant seems to reinforce the conservative approach. That is, the position for which the discrepancy might indicate the presence of some unknown systematics, rather than the presence of new physics.

5 Robustness and Systematics

5.1 Robustness

The idea behind robustness analysis, at least as formulated in (Wimsatt, Brewer, and Collins 1981), is that of invariance of a given result under multiple independent determinations. When this is applied to scientific models and theories, it can play a role in determining which models make trustworthy predictions, especially if there is no comprehensive theory working in the background. In this sense, robustness analysis can show "whether a result depends on the essentials of the model or on the simplifying assumptions" (Levins 1966, p. 422). This is usually done by studying distinct similar models of the same phenomenon, and thus: "if these models, despite their different assumptions, lead to similar results, we have what we can call robust theorem that is relatively free of the details of the model" (Levins 1966, p. 423).

We can slightly turn the target of robustness analysis for our goals here. Instead of using it to separate the essentials parts of a model from its simplifying assumptions, we can apply it to determine whether a given result is trustworthy or not based on whether it is robust under different means of determination. Indeed, a more recent recent characterization that fits better with our purposes is offered in (Soler et al. 2012, p. 2):

> Robustness is defined as the use of 'multiple means of determination' to 'triangulate' the existence and the properties of

[3] The results from TRGBs observations has nonetheless been contested, see, for example, (Yuan et al. 2019), and (https://www.quantamagazine.org/cosmologists-debate-how-fast-the-universe-is-expanding-20190808)

> a phenomenon, of an object or a result. The fundamental idea is that any object (a perceptual object, a physical phenomenon, an experimental result, etc.) that is sufficiently invariant under several independent derivations (in a wide sense of the term 'derivation' including means of identification, sensorial modalities, measurement processes, tests, models, levels of description, etc.) owes its strength (i.e., its robustness) to this situation.

However, one might argue that this method of assessing the strength of a model, theory, or property, is lacking a relation with the data and with the 'empirical side of science'. Indeed, already in 1993, (Orzack and Sober 1993) criticized robustness analysis in that it is a non-empirical form of confirmations and thus should not have a place in scientific practice. Instead, they give a formal interpretation to the notion of robust theorems in terms of propositions that are logically entailed by a set of models —where the type of proposition that follows from logical entailment cannot be empirical in that it is performed on models and not on data. Orzack and Sober (1993), but this is also recalled by (Weisberg 2006) and (Gueguen 2023), list three possible scenarios in which a theory may be subject to robustness analysis: (i) theorists know that all models are false, (ii) they know that there is at least one model that is true, (iii) they do not know whether there is at least one model that is true. Then, they argue, RA yields a true theorem (T) only in case (ii), but the other two cases give no evidence about the truth of T. One could argue further that theorists might know that all models are true, but then there is no need for robustness analysis. Alternatively, we could consider that all models are false, especially since all models are idealized (or contain idealizations). But, if this were the case, robust analysis would not be able to tell us anything because logical entailment fails only when all premises are true and the conclusion is false. But if the conclusion is true and the premises are either all false or some are false and some are true, one cannot infer anything about the premises.

Even though the logic behind the argument is sound, (Weisberg 2006, p. 733) argues that: "their [Ozark and Sober] analysis is too abstract to be used for this purpose. While I do not dispute the validity of their argument [...] this result does not necessarily apply to specific subset of models or to particular kinds of logical consequences". Afterward, Weisberg provides an example of RA and a 4-step procedure for RA. The example is based on the Lotka-Volterra model for predation, which is described by two differential equations that do not have a stable equilibrium, but only an unstable one that corresponds to the time averaged size of the modeled populations. If one were to introduce an external factor like a pesticide, then the population of prey would increase based on the number of predator and on the application of the pesticide. To evaluate the robustness of this statement (Volterra principle), one needs to evaluate different alternative models. Weisberg uses: (i) the addition of a term for the prey's population carrying capacity; the result is that the model changes, but the Volterra principle still holds. (ii) The addition of a term for predation satiation, which again changes the model, yet the Volterra principle holds. Then, Weisberg concludes that the Volterra principle is a robust property because it is common to different models that make use of different assumptions.

The general version of RA is then formulated as a conditional hypothesis: "Ceteris paribus, if [common structure] obtains, then [robust property] will obtain" (Weisberg 2006, p. 738), and the robustness analysis can be described as a 4-step procedure: (1) examine group of models to determine if they all predict a common result, the robust property; (2) analyze models for common structure that generate the robust property; (3) combine (1) and (2) to formulate the robust theorem, that is, the conditional statement that links robust property and common structure; (4) "the theorists can conduct stability analysis of the robust theorem to determine what conditions will defeat the connection between common structure and robust property" (Weisberg 2006, p. 737). In what follows we shall see an application of RA to the case of the tension between different measurements of the Hubble constant.

5.2 A Matter of Systematics

In a recent paper, (Gueguen 2023) suggests to use robustness analysis ((Levins 1966), (Wimsatt, Brewer, and Collins 1981), (Orzack and Sober 1993), and others) to investigate whether the H_0 tension can be considered as a form of a crisis in cosmology. She suggests that the recent observations from TRGBs might indicate that there are unknown systematics to be accounted for in the distance ladder methods. Since TRGBs are old stars that live in isolation —and thus they are not exposed to crowding effects— and since they are not pulsating objects, they can be considered as complementary to Cepheids: "these complementary features constitute an ideal investigation also to discover new sources of uncertainties not necessarily accounted for in the report of the accuracy of these measurements. The question is thus the following:

how can we explain the fact that, TRGBs excluded, the different methods based on the local universe agree on a high value and the methods based on the early universe on a low value, whereas at the same time the two methods that are the closest to each other, the more complementary and the more likely to agree fail to do so? How can we account for this success on one hand and this failure on the other?" (Gueguen 2023, p. 42).

To do so, since we do not have a toolbox that can decide for the correctness of the measurements and observations directly, we need a tool that tackles the problem from a more abstract level. That is, ideally, we would want to evaluate the different models that are being used to calculate H_0 both within individual methods (that is: TRGBs, Cepheids, BAO, Gravitational lensing), and across different ones (such as: between distance ladder and CMB). It is in the former sense (for the most part) that (Gueguen 2023) applies robustness within the distance ladder methods and, more specifically, to the case of Cepheids and TRGBs.

At the conference in Santa Barbara (California) in 2019 four new measurements of H_0 were presented (Gueguen 2023): (i)H0LiCOW: H_0=73.3 $\pm$ 1.7 $km \cdot s^{-1} Mpc^{-1}$ which uses gravitational lensing: (Wong et al. 2020); (ii) Mira variables: 73.3$\pm$ 4 $km \cdot s^{-1} Mpc^{-1}$ which are based on a distance ladder method that uses pulsating stars (Mira) as calibrators (Huang et al. 2019); (iii) Megamaser cosmology project $H_0 = 73.9 \pm 3$ $km \cdot s^{-1} Mpc^{-1}$ which also measure H_0 directly by gravitational lensing (Pesce et al. 2020); and (iv) the cosmic ladder based on the tip of the red giant branch stars: $H_0 = 69.8 \pm 0.8$ $km \cdot s^{-1} Mpc^{-1}$ (Freedman et al. 2019).[4] As I recalled earlier, TRGBs are old stars that live in isolation, they are not as exposed to crowding effects, they are not affected by metallicity, and they are not pulsating. Because of these reasons, and because they are different calibrators in the distance ladder method, they can be compared to Cepheids measurements in the context of robust analysis. In other words, the idea is to compare the replication of the measurements of H_0 via the distance ladder method, calibrated with TRGB and Cepheids.

Yet, following (Fletcher 2021), (Zwaan et al. 2018), and (Schmidt and Oh 2016), there are four types of replication: (i) direct replication consists of the attempt to reproduce the original study on different statistical sets; (ii) methodological replication involves the re-analysis of an experiment, possi-

[4]See also: https://www.quantamagazine.org/cosmologists-debate-how-fast-the-universe-is-expanding-20190808.

bly by a different team; (iii) systematic replication consists of changing or manipulating one variable of the experiment to identify variables with causal powers and thus circumscribe causal contributions; (iv) conceptual replication consists of testing or measuring the same phenomenon with different methods. Gueguen (2023) argues that TRGBs and Cepheids can be seen as either systematic or conceptual replications. In the former case, the two methods deliver different results, yet one is not able to identify the one variable that marks the difference between the two methods. It follows that one cannot exclude that there might be some unknown systematics that are causing (or contributing) to the different results we observe. Indeed, measurements from TRGB and Cepheids deliver different values for the Hubble constant: $H_0 = 69.8 \pm 0.8$ for TRGB (Freedman et al. 2019) and $H_0 = 74 \pm 1.4$ for Cepheids (Riess et al. 2019), and "[i]n the case of the Hubble constant, the failure of the systematic replication performed on the Cepheids and TRGB results shows that the precision of these measurements, though by far the most mature techniques for determining the value of the Hubble constant, has not reached a sufficient level for robustness arguments to be telling and/or trusted" (Gueguen 2023, p. 50).

With respect to conceptual replication, this amounts to robustness analysis in that the same parameter (H_0) is being calculated using different and independent models (different standard candles). However, there are no indications on where to look at for an explanation of the discrepancy of the measurements based on TRGBs ($H_0 = 69.8 \pm 0.8$) and Cepheids ($H_0 = 74 \pm 1.42$), and thus the robustness analysis in this case fails: "[i]n other words, the comparison between the two results can be constructed such as to maximally overlap and leave only the choice of standard candle as the variable explored —which amounts to the perfect picture of a systematic replication, or to be fully independent, which would amount in the case of an agreement to a perfect conceptual replication, inasmuch as the standard candle used is no longer considered a mere variable but a method" (Gueguen 2023, p. 50).

The conclusion is that one should not apply robustness analysis (conceptual replication) too early. If there is a failure of systematic replication, one should look for new systematics by refining measurements and observations. To do so, one makes use of models that are complementary to one another, rather than being independent. It is in this sense that the James Webb Space Telescope might contribute to indicate where to look for such new systematics:

> Freedman proceeded to a detailed systematic comparison between TRGB stars and Cepheids that we now have a better idea about where to look for possible unidentified unknown unknowns. While the two methods show excellent agreement on the distance modulus to 28 galaxies for instance, the study shows that this agreement no longer holds when comparing the distance to the 10 SNIa host galaxies that the two have in common. Future observational campaigns with much higher resolution, notably thanks to the JWST telescope, might be in a position to elucidate this disagreement (Gueguen 2023, p. 50).

However, as we shall see in the last section of this paper, the new measurements from the James Webb Space Telescope have greatly reduced the possibility of new systematics, but they have not resolved the tension.

What about the distance measures calculated based on CMB and BAO? The results from (Wong et al. 2020) were considered as further evidence of a crisis. Indeed, since H0LiCOW consists of gravitational lensing measures, and Cepheids are used in the distance ladder methods, they should be as independent as possible, since theories, principles, and analysis are unrelated. But, argues Gueguen, H0LiCOW is young and more work needs to be done to evaluate how different assumptions might distort the results. For example, (Birrer et al. 2023) discuss how different mass assumptions and lenses samples can vary H_0 from $\sim$73 to $\sim$67, a value much closer to CMB measures. This seems to suggest that based on different tweaks and assumptions, the values of H_0 obtained from the observations may vary significantly, thereby reinforcing the need to evaluate the assumptions used by method and models thereof.

5.3 Recent Results and Robustness Analysis

There is a difference in applying robustness analysis between Cepheids and TRGBs, and between Cepheids and H0LiCOW: in the former case the analysis runs on two different calibrators for a distance ladder method, while in the latter case the methods are almost entirely unrelated. This does not necessarily invalidate Gueguen's point, for which we cannot apply conceptual replication (robustness analysis) before systematic replication. Yet, it remains that we should distinguish two possible ways of applying robustness analysis: intra-methods and inter-methods. Is this difference relevant to the case of the Hubble Tension? I think so, since it impacts the context in which we can (or we cannot) apply robustness analysis. The intra-method consists of the use of (or fail to use) robustness analysis (RA) between TRGB and Cepheids. As mentioned above, Gueguen suggests that this amounts to a failure of RA, since the use of different calibrators yields different results, even though it is not clear which variable is responsible for the discrepancy. It is argued that there might be some systematics that taint the observables and explain the discrepancy. Similarly, RA does not apply to the case of gravitational lensing and Cepheids because the assumptions of the former ought to be further studied, and also the discrepancy in the results does not point towards a single causal variable either. This latter point seems to be more trivial, since there are far many more differences between gravitational lensing and distance ladder methods, than there are between H_0 determination with different calibrators. It is a much stronger condition to say that robustness analysis should provide a one-variable culprit for the discrepancy between Cepheids and H0LiCOW, as compared to the discrepancy between Cepheids and TRGB.

Let us begin with some considerations on the intra-model application of robustness analysis, and, more specifically, on some recent results from the calibration of TRGBs and from the James Webb Telescope (JWT). With respect to the former, I have briefly introduced how TRGB is a well understood phase of stellar evolution. In addition, TRGB stars are oftentimes targeted for distance measurements since (among other things) they are considered as non-variable stars. However, as most recently reported in (Anderson, Koblischke, and Eyer 2023), the variability of red giant stars has been known and studied for a 'long' time (Stebbins and Huffer 1928), as well as red giants period-luminosity sequences (Wood et al. 1999). More recently, the Optical Gravitational Lensing Experiment (OGLE) (Udalski et al. 2008) contributed to the discovery of a large population of red giant stars with small amplitude (OSARG) (Wray, Eyer, and Paczyński 2004) and multi-periodic variability (Kiss and Bedding 2003). The calibration that results from accounting for this variability suggests a 2.9% increase in TRGB distances, and a measure of H_0=71.8±1.5 $km \cdot s^{-1} Mpc^{-1}$. The work by (Anderson, Koblischke, and Eyer 2023, p. 10) concludes: "[w]e have shown that virtually all stars near the TRGB exhibit OSARG variability. This is a feature of RGB stars, which become more variable at higher luminosity [...] Applying our calibration to the CCHP H_0 results [(Freedman 2021)] yields H_0=71.8±1.5 $km \cdot s^{-1} Mpc^{-1}$, consistent with DL [distance ladder] calibrated using Cepheids [(Riess et al.

2022)] and in 2.8σ tension with the early Universe value from [(Collaboration et al. 2020)] that assumes λCDM model. Thus, our OSARGs LMC calibration improves among stellar standard candles and corroborates the need to find astrophysical solutions to the Hubble discord". In addition, (Yuan et al. 2019) address the HST archival observations for two ground based surveys (Optic Gravitational Lensing Experiment (OGLE-III) and Magellanic Cloud Photometry Survey (MCPS)), and show that MCPS data are less suitable for precise TRGB studies due to biases caused by blending. They find a $\sim 0.1mag$ offset in the magnitude of TRGB determined from MCPS data, and a smaller offset for OGLE-III. They show that the offset is due to the: "limited resolving power of ground instruments, as well a different filter responses for red stars" (Yuan et al. 2019, p. 2). Then, they apply the corrections to TRGB from (Freedman et al. 2019) and obtain H_0=72.4$\pm 2km \cdot s^{-1}Mpc^{-1}$ for TRGB and SNeIa distance ladder, which is a value compatible with the results obtained from distance ladder method with Cepheids calibrators.

If we believe these recent results, and if we consider the use of different calibrators as a one-variable-difference in the use of models to verify the value of H_0, then robustness analysis would seem to suggest that we have justified reasons to believe the concordance of such measurements. Following (Weisberg 2006), the conditional hypothesis would read: 'ceteris paribus, if the distance ladder method with Cepheids gives a value of H_0 $\sim$74, then H_0 $\sim$74'. To evaluate the robustness of the statement we have changed the calibrators (now, TRGB) in the distance ladder method, that is, we are evaluating inter-model robustness, and obtained that H_0 is still approximately equal to 74. If we apply the more recent version of robustness analysis offered by (Soler et al. 2012), and which I have mentioned earlier, the use of different calibrators amounts to inter-model differences that, since they give the same values for H_0, warrant the strength of their converging results.[5]

It is also worth mentioning that recent results from the James Webb Space Telescope (JWT), (Riess et al. 2023), (Riess et al. 2024), have eliminated the possibility of resolving the tension between TRGBs and Cepheids by appealing to systematics. The JWST, (Gardner et al. 2023) and (Rigby et al. 2023) has provided us with a much improved resolution to our observations,

[5]This conclusion is further strengthen if we broaden the inter-model perspective to late universe methods. Indeed, H0LiCOW gives similar results for H_0, but uses a different physics background for such calculations: gravitational lensing and thus general relativity.

so that now we are capable of overcoming crowding effects, and thus enhance the observation precision of individual Cepheids.[6] What Riess et al. (2023) deliver is a replication of the distance ladder methods with Cepheids, which was originally based on HST observations, with the more accurate measurements from JWST. One of the most relevant aspects of the study is the comparison between the relative distance between two galaxies (NGC4528 and NGC5584) obtained via observations from both space telescopes:

> This comparison involves large Cepheids samples (325 with JWST and 560 with HST); because both are compared *directly* Cepheids-to-Cepheids and *internally* to their respective photometric systems, uncertainties related to Cepheid calibration, photometric zero points, and SNe IA properties are irrelevant in this comparison. The number of Cepheids is sufficient to compare their relative distances to an accuracy σ=0.04Mag [...] this provides the strongest indicator to date that crowding does not play a role in the $\sim$ 0.18Mag Hubble tension (Riess et al. 2023).

What about the intra-models considerations on robustness? Here, we shall assume that the H_0 values obtained from λCDM model are internally robust. By internally, I mean that both methods (CMB and BAO) used for the calculation strengthen the trustfulness of the results within the context of the λCDM model. Then, what happens when we run the robust analysis against the measures of H_0 from λCDM model and distance ladder? The two measures do not converge (therefrom the well-discussed crisis), even though we are measuring the same parameter. Again, following (Weisberg 2006)'s conditional statement: 'ceteris paribus, if the universe is expanding, then H_0 obtains'. Obviously, the problem here is not whether the universe is expanding or not, but the fact that the use of different models deliver different values of the expansion-rate. Thus far, I have argued that inter-model robustness analysis, together with the many recent experimental results, justifies our beliefs in the distance ladder method. Granted that my argument holds, it follows that the culprit of the discrepancy seems to be in camp of our cosmological model. Yet, what this amounts to is far from clear. First of all, I have not offered here an equivalent assessment of the robustness of the measures from CMB and BAO, nor I have discussed their assumptions. Second, while the λCDM model makes use of general relativity, which is very well-verified

[6] More technical details are specified in (Riess et al. 2023) and references therein.

especially at cosmological scale, it also assumes the existence of dark matter and dark energy as substantial constituents of our universe.

More specifically: "[o]n cosmological scales, evidence for the existence of non-baryonic dark matter comes from observations of the Cosmic Microwave Background (CMB), Baryonic Acoustic Oscillations (BAO), primordial element abundances, and large-scale structure" (De Baerdemaeker and Dawid 2022, p. 2). Since non-baryonic (dark) matter and baryonic matter interact differently with radiation, the former might have had unforeseen effects on the power spectrum of the CMB anisotropies. Similarly, "baryonic matter is subject both to gravitational collapse and outward radiation pressure, dark matter only contributes to gravitational collapse. The BAO amplitude is too high to be generated by baryonic matter alone, thus, again, providing evidence for non-baryonic dark matter" (De Baerdemaeker and Dawid 2022, p. 2). Also, the abundance of the lightest elements of the universe (produced during the big bang nucleosynthesys) is calculated based on photon density estimates from the CMB, but the estimates based on baryonic matter (only) are too low to justify a flat geometry of the universe (Reeves et al. 1973). Finally, baryonic matter is also insufficient to account for the amount of large-scale structure in the universe (Blumenthal et al. 1984).[7]

The presence of such evidences, and yet the lack of a model that tells us more about the physics of dark matter and dark energy, constitutes a reasonable indication that something is amiss in our standard model of cosmology, thereby validating (Verde, Treu, and Riess 2019, p. 3)'s statement for which: "something is not well understood in the relation between CMB anisotropies and the growth of structures, and this could perhaps be a hint towards new physics". It remains that a stronger conclusion, at least in the context of robustness analysis, requires investigating the assumptions on the background of the methods based on λCDM model, and of their the derivations of H_0. What we can conclude is that robustness analysis does justify our trust in the late universe measures of Hubble's constant, and that the tension can not be dismissed as a matter of systematics.

[7]For a philosophical discussion see, among others: (De Baerdemaeker and Dawid 2022) and references therein.

6 Conclusions

In this paper I first reviewed the history and the recent literature on the so-called H_0 tension. Afterward, I used the most recent results from calibrations of TRGB stars, and observations from the James Webb Space Telescope, to show that we are indeed justified in believing that the calculation of H_0 with late universe methods (mostly, distance ladder) is robust. Building on from such a conclusion, I have suggested that the resolution of the tension can not be expected from additional systematics —altough a thorough analysis of the assumptions behind the calculation of H_0 with BAO and CMB (that is, the λCDM model) is still needed. It follows that, granted the robustness of late universe methods, the culprit of the tension is to be searched in the standard model of cosmology, and given the many evidences we have for dark matter and dark energy, this conclusion seems justified.

References

Aaronson, M et al. (1986). "A distance scale from the infrared magnitude/HI velocity-width relations. V-Distance moduli to 10 galaxy clusters, and positive detection of bulk supercluster motion toward the microwave anisotropy". In: *The Astrophysical Journal* 302, pp. 536–563.

Aaronson, Marc et al. (1980). "A distance scale from the infrared magnitude/HI velocity-width relation. III-The expansion rate outside the local supercluster". In: *Astrophysical Journal, Part 1, vol. 239, July 1, 1980, p. 12-37. Research supported by the University of Washington* 239, pp. 12–37.

Aaronson, Marc et al. (1982). "The velocity field in the local supercluster". In: *Astrophysical Journal, Part 1, vol. 258, July 1, 1982, p. 64-76.* 258, pp. 64–76.

Anderson, Richard I, Nolan W Koblischke, and Laurent Eyer (2023). "Reconciling astronomical distance scales with variable red giant stars". In: *arXiv preprint arXiv:2303.04790*.

Bergh, Sidney Van Den (1970). "Extra-galactic distance scale". In: *Nature* 225.5232, pp. 503–505.

Birrer, S. et al. (2023). *Time-Delay Cosmography: Measuring the Hubble Constant and other cosmological parameters with strong gravitational lensing.* arXiv: 2210.10833 [astro-ph.CO].

Blumenthal, George R et al. (1984). "Formation of galaxies and large-scale structure with cold dark matter". In: *Nature* 311.5986, pp. 517–525.

Collaboration, Planck et al. (2020). "Planck 2018 results. VI. Cosmological parameters". In.

De Baerdemaeker, Siska and Richard Dawid (2022). "MOND and meta-empirical theory assessment". In: *Synthese* 200.5, p. 344.

De Vaucouleurs, G (1985). "Tycho's supernova and the Hubble constant". In: *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 289, Feb. 1, 1985, p. 5-9.* 289, pp. 5–9.

Fletcher, Samuel C (2021). "The role of replication in psychological science". In: *European Journal for Philosophy of Science* 11.1, p. 23.

Freedman, Wendy L (2021). "Measurements of the Hubble constant: tensions in perspective". In: *The Astrophysical Journal* 919.1, p. 16.

Freedman, Wendy L et al. (2001). "Final results from the Hubble Space Telescope key project to measure the Hubble constant". In: *The Astrophysical Journal* 553.1, p. 47.

Freedman, Wendy L et al. (2019). "The Carnegie-Chicago Hubble program. VIII. An independent determination of the Hubble constant based on the tip of the red giant branch". In: *The Astrophysical Journal* 882.1, p. 34.

Gardner, Jonathan P et al. (2023). "The James Webb Space Telescope Mission". In: *Publications of the Astronomical Society of the Pacific* 135.1048, p. 068001.

Gueguen, Marie (2023). "A Crack in the Track of the Hubble Constant". In: *Philosophy of Astrophysics: Stars, Simulations, and the Struggle to Determine What is Out There*. Springer International Publishing Cham, pp. 33–55.

Holz, Daniel E and Robert M Wald (1998). "New method for determining cumulative gravitational lensing effects in inhomogeneous universes". In: *Physical Review D* 58.6, p. 063501.

Hoscheit, Benjamin L and Amy J Barger (2018). "The KBC void: Consistency with supernovae type Ia and the kinematic SZ effect in a ΛLTB model". In: *The Astrophysical Journal* 854.1, p. 46.

Huang, CD et al. (2019). "Hubble Space Telescope Observations of Mira Variables in the Type Ia Supernova Host NGC 1559: An Alternative Candle to Measure the Hubble Constant doi: 10.3847/1538-4357/ab5dbd". In: *arXiv preprint arXiv:1908.10883*.

Hubble, Edwin P (1925). "Cepheids in spiral nebulae". In: *Pop. Astr.; Vol. 33; Page 252-255* 33.

Humason, Milton L, Nicholas U Mayall, and Allan R Sandage (1979). "Redshifts and magnitudes of extra-galactic nebulae". In: *A Source Book in Astronomy and Astrophysics, 1900–1975*. Harvard University Press, pp. 753–762.

Kiss, LL and TR Bedding (2003). "Red variables in the OGLE-II data base–I. Pulsations and period–luminosity relations below the tip of the red giant branch of the Large Magellanic Cloud". In: *Monthly Notices of the Royal Astronomical Society* 343.3, pp. L79–L83.

Kreisch, Christina D, Francis-Yan Cyr-Racine, and Olivier Doré (2020). "Neutrino puzzle: Anomalies, interactions, and cosmological tensions". In: *Physical Review D* 101.12, p. 123505.

Leavitt, Henrietta S and Edward C Pickering (1912). "Periods of 25 Variable Stars in the Small Magellanic Cloud." In: *Harvard College Observatory Circular, vol. 173, pp. 1-3* 173, pp. 1–3.

Levins, Richard (1966). "The strategy of model building in population biology". In: *American scientist* 54.4, pp. 421–431.

Liddle, Andrew (2015). *An introduction to modern cosmology*. John Wiley & Sons.

Mager, Violet A, Barry F Madore, and Wendy L Freedman (2008). "Metallicity-corrected tip of the red giant branch distance to NGC 4258". In: *The Astrophysical Journal* 689.2, p. 721.

Orzack, Steven Hecht and Elliott Sober (1993). "A critical assessment of Levins's The strategy of model building in population biology (1966)". In: *The Quarterly Review of Biology* 68.4, pp. 533–546.

Pesce, DW et al. (2020). "The megamaser cosmology project. XIII. Combined Hubble constant constraints". In: *The Astrophysical Journal Letters* 891.1, p. L1.

Poulin, Vivian et al. (2019). "Early dark energy can resolve the Hubble tension". In: *Physical review letters* 122.22, p. 221301.

Reeves, Hubert et al. (1973). "On the origin of light elements". In: *Astrophysical Journal, Vol. 179, pp. 909-930 (1973)* 179, pp. 909–930.

Riess, Adam G et al. (2016). "A 2.4% determination of the local value of the Hubble constant". In: *The Astrophysical Journal* 826.1, p. 56.

Riess, Adam G et al. (2019). "Large Magellanic Cloud Cepheid standards provide a 1% foundation for the determination of the Hubble constant and stronger evidence for physics beyond ΛCDM". In: *The Astrophysical Journal* 876.1, p. 85.

Riess, Adam G et al. (2022). "A comprehensive measurement of the local value of the Hubble constant with 1 km s- 1 Mpc- 1 uncertainty from the Hubble Space Telescope and the SH0ES team". In: *The Astrophysical journal letters* 934.1, p. L7.

Riess, Adam G et al. (2023). "Crowded No More: The Accuracy of the Hubble Constant Tested with High-resolution Observations of Cepheids by JWST". In: *The Astrophysical Journal Letters* 956.1, p. L18.

— (2024). "JWST Observations Reject Unrecognized Crowding of Cepheid Photometry as an Explanation for the Hubble Tension at 8σ Confidence". In: *The Astrophysical Journal Letters* 962.1, p. L17.

Rigby, Jane et al. (2023). "The science performance of JWST as characterized in commissioning". In: *Publications of the Astronomical Society of the Pacific* 135.1046, p. 048001.

Sandage, Allan (1958). "Current Problems in the Extragalactic Distance Scale." In: *Astrophysical Journal, vol. 127, p. 513* 127, p. 513.

Sandage, Allan and GA Tammann (1974a). "Steps toward the Hubble constant. III-The distance and stellar content of the M101 group of galaxies". In: *Astrophysical Journal, vol. 194, Dec. 1, 1974, pt. 1, p. 223-243. Research supported by the Swiss National Science Foundation* 194, pp. 223–243.

— (1974b). "Steps toward the Hubble constant. IV-Distances to 39 galaxies in the general field leading to a calibration of the galaxy luminosity classes and a first hint of the value of H sub 0". In: *Astrophysical Journal, vol. 194, Dec. 15, 1974, pt. 1, p. 559-568. Research supported by the Swiss National Science Foundation* 194, pp. 559–568.

— (1975a). "Steps toward the Hubble constant. V-The Hubble constant from nearby galaxies and the regularity of the local velocity field". In: *Astrophysical Journal, vol. 196, Mar. 1, 1975, pt. 1, p. 313-328. Research supported by the Swiss National Science Foundation* 196, pp. 313–328.

— (1975b). "Steps toward the Hubble constant. VI-The Hubble constant determined from redshifts and magnitudes of remote SC I galaxies: The value of Q sub 0". In: *Astrophysical Journal, vol. 197, Apr. 15, 1975, pt. 1, p. 265-280.* 197, pp. 265–280.

— (1982). "Steps toward the Hubble constant. VIII-The global value". In: *Astrophysical Journal, Part 1, vol. 256, May 15, 1982, p. 339-345. Research supported by the Swiss National Science Foundation;* 256, pp. 339–345.

Sandage, Allan et al. (1996). “Cepheid calibration of the peak brightness of Type Ia Supernovae: calibration of SN 1990N in NGC 4639 averaged with six earlier Type Ia Supernova calibrations to give H0 directly”. In: *The Astrophysical Journal* 460.1, p. L15.

Schmidt, Frank L and In-Sue Oh (2016). “The crisis of confidence in research findings in psychology: Is lack of replication the real problem? Or is it something else?” In: *Archives of Scientific Psychology* 4.1, p. 32.

Serenelli, Aldo et al. (2017). “The brightness of the red giant branch tip-Theoretical framework, a set of reference models, and predicted observables”. In: *Astronomy & Astrophysics* 606, A33.

Smith, Mathew et al. (2013). “The effect of weak lensing on distance estimates from supernovae”. In: *The Astrophysical Journal* 780.1, p. 24.

Soler, Léna et al. (2012). *Characterizing the robustness of science: After the practice turn in philosophy of science*. Vol. 292. Springer Science & Business Media.

Stebbins, Joel and Charles Morse Huffer (1928). “The Constancy of the Light of Red Stars.” In: *Publications of the Washburn Observatory, vol. 15, pp. 137-174* 15, pp. 137–174.

Tonry, John and Donald P Schneider (1988). “A new technique for measuring extragalactic distances”. In: *Astronomical Journal (ISSN 0004-6256), vol. 96, Sept. 1988, p. 807-815.* 96, pp. 807–815.

Tonry, John L (1991). “Surface brightness fluctuations-A bridge from M31 to the Hubble constant”. In: *Astrophysical Journal, Part 2-Letters (ISSN 0004-637X), vol. 373, May 20, 1991, p. L1-L4.* 373, pp. L1–L4.

Tonry, John L et al. (1997). “The SBF survey of galaxy distances. I. Sample selection, photometric calibration, and the Hubble constant”. In: *The Astrophysical Journal* 475.2, p. 399.

Tully, R Brent (2023). “The Hubble Constant: A Historical Review”. In: *arXiv preprint arXiv:2305.11950*.

Tully, R Brent and J Richard Fisher (1977). “A new method of determining distances to galaxies”. In: *Astronomy and Astrophysics, vol. 54, no. 3, Feb. 1977, p. 661-673.* 54, pp. 661–673.

Udalski, Andrzej et al. (2008). “The optical gravitational lensing experiment. final reductions of the OGLE-III Data”. In: *arXiv preprint arXiv:0807.3884*.

Verde, Licia, Tommaso Treu, and Adam G Riess (2019). “Tensions between the early and late Universe”. In: *Nature Astronomy* 3.10, pp. 891–895.

Weisberg, Michael (2006). “Robustness analysis”. In: *Philosophy of science* 73.5, pp. 730–742.

Wimsatt, William C, M Brewer, and B Collins (1981). “Robustness, reliability, and overdetermination”. In: *Characterizing the Robustness of Science. Boston Studies in the Philosophy of Science* 292, pp. 61–87.

Wong, Kenneth C et al. (2020). “H0LiCOW–XIII. A 2.4 per cent measurement of H 0 from lensed quasars: 5.3 σ tension between early-and late-Universe probes”. In: *Monthly Notices of the Royal Astronomical Society* 498.1, pp. 1420–1439.

Wood, PR et al. (1999). “MACHO observations of LMC red giants: Mira and semi-regular pulsators, and contact and semi-detached binaries”. In: *Symposium-International Astronomical Union.* Vol. 191. Cambridge University Press, pp. 151–158.

Wray, JJ, L Eyer, and B Paczyński (2004). “OGLE small-amplitude variables in the Galactic bar”. In: *Monthly Notices of the Royal Astronomical Society* 349.3, pp. 1059–1068.

Yuan, Wenlong et al. (2019). “Consistent calibration of the tip of the red giant branch in the large magellanic cloud on the hubble space telescope photometric system and a redetermination of the hubble constant”. In: *The Astrophysical Journal* 886.1, p. 61.

Zwaan, Rolf A et al. (2018). “Making replication mainstream”. In: *Behavioral and Brain Sciences* 41, e120.