The Reasoner 19(4) , 2005. https://doi.org/10.54103/1757-0522/30052

The Disunity of Robustness: Hubble Tension

:: **Marco Forgione**

Abstract
The paper examines how conceptual disunity in robustness analysis (RA) can generate contradictory interpretations of the Hubble tension, the discrepancy between independent measurements of the universe's expansion rate. I will consider different philosophical accounts of robustness and how they apply to cosmological practice. I then demonstrate how each framework validates conflicting conclusions. More specifically, I show how Levins' model comparison, Woodward's measurement invariance, and Weisberg's representational accounts each justify different interpretations of whether systematic errors or new physics explain the tension.

Keywords
Robustness; Hubble tension

Submitted: 07/07/25
Revised: 07/11/2025
Accepted: 07/12/2025

How to Cite
Marco Forgione. *The Disunity of Robustness: Hubble Tension*. The Reasoner 19(4) 2025. https://doi.org/10.54103/1757-0522/30052

The Reasoner 19(4), 2005. https://doi.org/10.54103/1757-0522/30052

1. Introduction: The Paradox of Persistent Disagreement

In contemporary cosmology, a puzzling disagreement persists about how fast our universe is expanding. Astronomers call this the 'Hubble tension' – a long-lasting discrepancy between two sets of measurements of the Hubble constant (H_0), the number representing the current expansion rate of the universe. See: Di Valentino, Said, et al. (2025: *Addressing observational tensions in cosmology with systematics and fundamental physics, Physics of the Dark Universe, 1-416*). What makes this tension philosophically interesting (among other reasons) is that both sides appeal to *robustness* to defend their results. Proponents of higher expansion rates (approximately 73 km/s/Mpc) argue that their measurements are robustly consistent across different observational techniques. Advocates of lower rates (approximately 67 km/s/Mpc) maintain that their models are robustly validated by multiple theoretical evidences. This creates a paradox: Although both camps employ robustness arguments, they also reach contradictory conclusions.

Here, I suggest that, together with the observational challenges of determining the Hubble constant, philosophers should acknowledge the lack of a unique and agreed-upon definition of robustness – what I call: *conceptual disunity*. Indeed, different philosophical understandings of what constitutes robustness, either agreement between models, stability under parameter changes, or measurement convergence, allow scientists to draw incompatible inferences from the same data. This paper explores how this disunity affects the Hubble tension debate and points toward the need for a comparative framework to evaluate competing robustness claims.

2. The Hubble Tension and Competing Robustness Claims

Two Approaches to Cosmic Measurement To understand the Hubble tension, we must distinguish two approaches to measuring cosmic expansion. The first, known as the 'local approach', builds what astronomers call the cosmic distance ladder. This is a series of interdependent methods aimed at calculating the distances of cosmological objects that cannot be directly measured with a ruler. Roughly speaking, the distance ladder can be described by three main rungs:

- First rung: The zero-point calibration uses geometric parallaxes to determine the distance of nearby objects with extreme precision within the Milky Way, or in the Large Magellanic Cloud.
- Second Rung: Measurements in the second rung extend our reach well beyond the local group by leveraging objects with predictable intrinsic brightness. For example: variable stars called Cepheids, whose pulsation periods correlate with their true luminosity.
- Third rung: One uses the same principles applied to

The Reasoner 19(4), 2005. https://doi.org/10.54103/1757-0522/30052

the second rung and compares the intrinsic and apparent brightness of different standard candles such as Supernovae Type Ia (SNeIa) to measure their distances and the distance to their host galaxies. From the redshift of such distant galaxies one can use Hubble's law $\vec{v} = H_0\vec{r}$ to infer the Hubble constant.

The second approach, known as the 'global method', examines the cosmic microwave background (CMB) – the faint radiation permeating space that originated when the universe became transparent to light – see: Liddle (2015: *An introduction to modern cosmology, John Wiley & Sons*). This radiation contains small temperature variations that cosmologists analyze using sophisticated models of cosmic evolution, primarily the ΛCDM (Lambda Cold Dark Matter) model. By fitting these models to CMB data, scientists can predict the universe's current expansion rate – see: Agahim, Akrami, et al. (2020: *Planck 2018 results. VI. Cosmological parameters, Astronomy & Astrophysics, 641, A6*).

The Robustness Impasse Both the local and global methods appear to be internally robust, and yet they deliver different values for the Hubble constant. Local measurements consistently converge around higher values (H_0 ~73 km/s/Mpc) using different techniques including Cepheid-calibrated supernovae – see, among others: Riess, Yuan, et al. (2022: *A comprehensive measurement of the local value of the Hubble constant with 1 km s^{-1} Mpc^{-1} uncertainty from the Hub-*

The Reasoner 19(4) , 2005. https://doi.org/10.54103/1757-0522/30052

The Reasoner 19(4), 2005. https://doi.org/10.54103/1757-0522/30052

ble Space Telescope and the SH0ES Team, The Astrophysical Journal, 934(1), L7)– , gravitational lensing time delays – see: Wong, Suyu, et al. (2019: *H0LiCOW XIII. A 2.4% measurement of H_0 from lensed quasars: 5.3σ tension between early and late-Universe probes, Monthly Notices of the Royal Astronomical Society, 498(1), 1420-1439*)– , and water maser observations – see: Pesce, Braatz, et al. (2020: *The Megamaser Cosmology Project. VIII. A geometric distance to NGC 5765b, The Astrophysical Journal, 891(1), L1*). Global methods yield consistent lower values (H_0 ~67 km/s/Mpc) through various analyses of CMB data and baryon acoustic oscillations – frozen sound waves from the early universe imprinted in galaxy distributions. For example: Cuceu, Farr et al. (2019: *Baryon acoustic oscillation and the Hubble constant: past, present and future. Journal of Cosmology and Astroparticle Physics, 2019(10), 044*).

A further distinction that helps clarify the local-global discrepancy is between precision (the tightness of statistical uncertainties) and accuracy (closeness to the true value, accounting for possible systematic biases). Precision can be high in both approaches, but it is achieved under different sets of dependencies: local results rely on calibration chains (distance anchors, stellar-population assumptions, crowding and metallicity treatments), whereas global results depend on cosmological modeling (the structure of ΛCDM, sound-horizon physics, recombination modeling, and priors). Accuracy, in turn, is also conditional: for late-time methods it is calibration and population relative, and for early-time methods it is model and prior-relative.

Taken together, these differences lead to a 'robustness impasse': each approach satisfies internal robustness criteria while contradicting the other at a statistically significant level (approximately 5 km/s/Mpc difference, exceeding 4σ confidence). Philosophers traditionally view robustness as a truth-converging mechanism, that is: when multiple independent methods agree, confidence in the result increases. Then, the impasse here is that while the individual methods (distance ladder and ΛCDM-based methods) for calculating H_0 are deemed robust, the same robustness fails when we compare the results across different methods.

3. Philosophical Frameworks of Robustness

Divergent Accounts Philosophical accounts of robustness provide distinct lenses for interpreting the Hubble tension. For example, Richard Levins focuses on the idea of a robust theorem, which he develops by deliberately comparing a collection of simplified models – see: Levins (1966: *The strategy of model building in population biology, American Scientist, 54(4), 421-431*). He observes that each model involves its own artificial assumptions, so one can never be entirely sure whether a given result depends on the genuine features of the system or on the model's simplifications. To address this, he recommends approaching the same problem

The Reasoner 19(4) , 2005. https://doi.org/10.54103/1757-0522/30052

with several alternative models that share a core assumption but differ in their simplifying details. If all of these models – despite their varied assumptions – converge on the same outcome, then that outcome qualifies as a 'robust theorem', since it depends primarily on the shared core and is largely insensitive to the arbitrary details of any one model. As Levins (1966, p.423) famously puts it: "Hence our truth is the intersubsubsection of independent lies". Yet, while Levins' framework might work for individual methods to calculate H_0, it offers no mechanism for comparing the results obtained from the cosmic distance ladder with those from the ΛCDM model.

A different approach suggested by Woodward (2006: *Sensitive and robust scientific inference, Philosophy of Science, 13(2), 219-240*) distinguishes several varieties of robustness: *inferential robustness* involves the insensitivity of an inference to varying assumptions. *Derivational robustness* occurs when a theoretical result or prediction remains stable under different parameters in the model. *Causal robustness* focuses on the stability of causal relations under different interventions. Finally, *measurement robustness* focuses on the agreement across different measurement techniques or instruments. Notably, local measurements of H_0 primarily emphasize measurement robustness, for example: the Hubble Space Telescope (HST) and the James Webb Space Telescope (JWST) deliver consistent Cepheid results – see: Riess, Scolnic, et al., (2024: *JWST Validates HST Distance*

The Reasoner 19(4), 2005. https://doi.org/10.54103/1757-0522/30052

Measurements: Selection of Supernova Subsample Explains Differences in JWST Estimates of Local H0. The Astrophysical Journal, 977(1), 120). On the other hand, global measurements tend to prioritize derivational robustness – the stability of H_0 under variations of the parameters within ΛCDM. While Woodward cautions about distinguishing different types of robustness in order to avoid confusion and misapplications, he does not discuss whether and how different types of robust results can be compared with one another. Without this cross-examination, the robustness of the cosmic distance ladder remains philosophically disconnected from that of the cosmological model.

Another taxonomy of robustness definitions is suggested by Weisberg and Reisman (2008: *Robustness analysis, philosophy of science, 75(1), 106-131*), where they distinguish between three categories. (i) *Parameter robustness*: investigates whether the behavior of a model remains consistent across a range of parameter values. In the case of the Hubble tension, *parameters sweeps* – searching the parameter space for any combination of parameters that would resolve the tension– might expose that no choice of parameters fixes the H_0 tension. Yet, this would not discriminate the existence of some new physics outside the ΛCDM model, or some hidden biases in the datasets. (ii) *Structural robustness*: examines the effects of altering the causal or mathematical structure of the model. But, even if a structural extension could solve the H_0 tension, the risk is to mistake

The Reasoner 19(4), 2005. https://doi.org/10.54103/1757-0522/30052

an over-parameterized solution for a genuine resolution of the tension, thereby leaving the fundamental question (new physics vs. hidden systematic) still unanswered. Finally (iii), *representational robustness* evaluates whether predictions hold under different model representations. The limitation is that 'independent' methods often share underlying assumptions (stellar population models, calibrations, etc.). Because of this overlapping, it is difficult to pin down which assumption might be'responsible' for unknown systematics or new-physics.

Orzack and Sober's critique to Levins further complicates matters by arguing that robustness has heuristic value but doesn't guarantee truth – see: Orzack and Sober (1993: *How to be a successful error theorist: the case of Levins' model of robustness, Philosophy of Science, 60(4), 531-550*). They argue that different models can contain false assumptions and yet converge on the same conclusion. Yet, "if every model contains false assumptions, how can we ever hope to discover what is true?" (Orzack and Sober 1993, p.538). Once again, consider how robustness fails in evaluating the results for H_0 using different methods. At best, one can evaluate the robustness of the results of each method individually, but, since each method requires false assumptions and idealizations, it is not possible to determine which method delivers the true value of the Hubble constant.

Finally, Wimsatt (1981: *Robustness, reliability, and overdetermination, in: Characterizing the robustness of science:*

The Reasoner 19(4) , 2005. https://doi.org/10.54103/1757-0522/30052

After the practice turn in philosophy of science (2012), 61-87, Springer.) views robustness as the redundant support of a theoretical framework. That is, whenever there is an inconsistency in a theory due to conflicting idealizations or simplifying assumptions, robustness guarantees that there are alternative derivations of the same conclusions. These alternative (redundant) derivations prevent the inconsistency from propagating to the entire theoretical structure. Unfortunately, this account also provides ambiguous guidance in our case. The local measures of H_0 appear robust through redundant measurement techniques, while the global measures show robustness through model self-consistency. Yet, neither framework decisively resolves which should take precedence when they conflict.

In conclusion, these divergent accounts can be used to validate both interpretations of the Hubble tension: that local measurements contain hidden systematic errors, or that global methods require new physics. Both remain philosophically defensible through selective application of robustness frameworks, demonstrating how disunity permits contradictory conclusions to coexist.

4. The Tension Within the Tension: the TRGB Case Study

A revealing case study emerges from an anomaly within local measurement methods. While most techniques (Cepheid-calibrated supernovae, gravitational lensing) con-

The Reasoner 19(4) , 2005. https://doi.org/10.54103/1757-0522/30052

The Reasoner 19(4), 2005. https://doi.org/10.54103/1757-0522/30052

verge around $H_0 \sim 73$ km/s/Mpc, one method using 'tip of the red giant branch' (TRGB) stars yields a significantly lower value (~69 km/s/Mpc) – see: Freedman, Madore, 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, The Astrophysical Journal, 882(1), 34*). TRGB stars are low-mass stars at a specific evolutionary stage where they undergo a helium flash, creating a recognizable brightness cutoff in astronomical observations.

This anomaly creates what might be called 'the tension within the tension' – a disagreement within the local measurement cluster that challenges simple robustness narratives. Indeed, different philosophical frameworks interpret this anomaly differently. For example, a Levinsian perspective might view TRGBs as representing a different 'model' of distance measurements. This could suggest that that the use of different standard candles leads to some fragility when it comes to measurements of the Hubble constant. Similarly, when adopting Woodward's *measurement robustness*, the discrepancy might suggest insufficient agreement across techniques, pointing toward hidden systematic errors rather than new physics. Indeed, if different measurement methods yield different results, this undermines claims of robustness for the entire local approach.

Yet, Weisberg's taxonomy offers multiple interpretations: the discrepancy might reflect representational issues (different stellar physics employed), structural differences (alternative calibration pathways), or parameter sensitivities (variations in how metallicity affects brightness). Each suggests different approaches to resolving the tension within the tension.

Most recently, the James Webb Space Telescope, with its unprecedented infrared capabilities, was expected to resolve such discrepancies – see: Gardner, Mather, et al. (2023: *The James Webb Space Telescope Mission, Publications of the Astronomical Society of the Pacific, 135(1048), 068001*). To see how, it is useful to separate again statistical stability (precision) from systematic stability (accuracy), and to distinguish between direct from conceptual replication. The former evaluates precision and reliability by repeating an experiment with (approximately) the same methods but different statistical sample. The latter probes accuracy by altering experimental methods to expose systematic biases – see: Matarese, McCoy (2024: *When "replicability" is more than just "reliability": The Hubble constant controversy, Studies in History and Philosophy of Science, 107, 1-10*). In these terms, recent observations of Cepheids from JWST provide high-precision direct replication of previous HST measurements (same indicator, improved instrument), mitigating crowding and blending concerns. At the same time, TRGBs function as conceptual replications of the distance ladder: they involve different stellar physics, calibration anchors, and selection effects, thus probing accuracy in a way that is partly orthogonal to Cepheids.

The Reasoner 19(4), 2005. https://doi.org/10.54103/1757-0522/30052

The Reasoner 19(4), 2005. https://doi.org/10.54103/1757-0522/30052

JWST has confirmed the HST measurements of Cepheid variables with remarkable precision, strengthening claims of measurement robustness for this method – see: Riess, Anand, et al. (2024: *JWST observations reject unrecognized crowding of cepheid photometry as an explanation for the hubble tension at 8σ confidence, The Astrophysical Journal Letters, 962(1), L17*). That is, the improved resolution of JWST indeed resolved possible concerns about stellar crowding, and other systematics, seemingly validating the higher H_0 values. But, at the same time, JWST observations of TRGB stars validated lower H_0 values, preserving the disagreement within local methods – see: Freedman, Madore, et al. (2024: *Status report on the Chicago-Carnegie Hubble Program (CCHP): Three independent astrophysical determinations of the Hubble constant using the James Webb Space Telescope, The Astrophysical Journal, 985(2), 203*).

From the perspective of Levins and Woodward, it seems rational to acknowledge that different robustness dimensions have strengthened the robustness of H_0 estimates within each standard candle method (Cepheids and TRGBs), but the disagreement between the two standard candles remains unsettled – as well as the broader tension between local and global methods.

Following Weisberg's taxonomy, the three notions of robustness can lead toward different conclusions about the Cepheid-TRGB tension within the tension. If modest changes in parameters (e.g., metallicity corrections, extinction models) can reconcile Cepheid and TRGB results, one might infer hidden systematics. Conversely, if no plausible parameter adjustments can bridge the gap, parameter robustness would leave us with an underdetermined conclusion. At the same time, resorting to different calibrators, data processing methods, or representations, could indicate either the distance ladder's fragility or the ΛCDM model's limitations.

To take stock, the coexistence of the strengthened precision for Cepheids and the persistent disagreement with TRGBs demonstrates how technological advances don't automatically resolve robustness disunity. Indeed, they can reinforce existing interpretive frameworks by providing higher-precision data that remains subject to different philosophical interpretations.

5. Navigating Robustness Disunity

The TRGB case above shows that higher precision can strengthen measurement robustness within a method without resolving cross-method disagreement. Therefore, I suggest that the way forward is to structure comparisons across robustness notions so that different robustness claims can be weighed 'side by side'.

For example, both Cepheid and TRGB distances should be compared to multiple independent anchors, and the resulting calibrations should be checked against independent H_0

The Reasoner 19(4) , 2005. https://doi.org/10.54103/1757-0522/30052

The Reasoner 19(4), 2005. https://doi.org/10.54103/1757-0522/30052

determinations to expose any remaining systematics in the cosmic distance ladder. Examples of such independent H_0 measurements include geometric megamasers (which also serve as anchors), quasar lensing, and prospective 'standard sirens' from gravitational waves – see: Perivolaropoulos (2024: *Hubble tension or distance ladder crisis?, Physical Review D, 110(12), 123518*). Results from different anchors should then be reported and compared directly, while holding fixed the sample of Type Ia supernovae and data reduction techniques. By swapping only the rungs of the distance ladder that differ between the Cepheid and TRGB methods, any shift in the inferred Hubble constant can be traced to a specific analytical step. Furthermore, to avoid overstating independence, one should make explicit the shared assumptions and statistical priors used across methods in order to keep track of how calibration steps interrelate. While these tests apply to TRGBs, Cepheids and to the distance ladder, a similar approach should be applied to the ΛCDM model and its inferences for H_0.

With respect to robustness, these tests adopt the form of a bottom-up strategy: (i) start within a single standard candle (e.g., Cepheids or TRGB) and vary likely systematics (crowding, metallicity, photometry) to test measurement robustness; (ii) compare across standard candles to probe representational robustness; (iii) compare across independent distance methods (ladders, masers, lensing, standard sirens) to assess methodological independence; and (iv) finally confront local and global determinations under explicit model and calibration choices to test derivational/model robustness. At each stage, ask whether a result persists under the corresponding stress test. The more layers a claim survives, the more comparably robust it is across these distinct notions, while a failure at a given layer should indicate where disagreement enters.

A note of caution: this method is only provisional and it will not resolve the Hubble tension by itself. The scarcity of truly independent astronomical anchors, despite JWST's power, means the comparative analysis may still rest on a limited and potentially problematic foundation. Similarly, the high statistical uncertainties of methods like megamasers or gravitational lensing limit their power to discriminate between subtle systematic errors in the primary distance ladder techniques. These practical limitations are not merely logistical but are philosophically significant, since they concretely manifest the dependency problem at the heart of robustness disunity. Acknowledging this is crucial to understanding that resolving the tension will require simultaneous progress on both empirical and conceptual fronts.

Conclusion

In this paper I demonstrated how the disunity of robustness allows for contradictory interpretations of the Hubble tension. Yet, this disunity can suggest a productive path for-

The Reasoner 19(4) , 2005. https://doi.org/10.54103/1757-0522/30052

ward for philosophical work: rather than seeking a single, monolithic account of robustness, the focus should shift to developing a comparative framework – a bottom-up strategy – for evaluating different types of robustness claims against one another. Such a framework requires articulating the specific dependencies and idealizations inherent in each method (e.g., the calibration chain for Cepheids, the sound horizon physics for ΛCDM) and then probe those dependencies in controlled ways. The goal is not to eliminate disunity, but to create a structured dialogue between competing robustness standards.

MARCO FORGIONE
https://orcid.org/0000-0002-4617-3500
Università degli Studi di Milano

THE REASONER 19(4), 2005. https://doi.org/10.54103/1757-0522/