The Hubble Constant: Case Study of a Cosmological Tension

Bridget Falck

Before the launch of the Hubble Space Telescope in 1990, there was significant disagreement in the reported measurements of the Hubble constant. This was seen as a controversy or tension because the difference between the values obtained by different groups were well outside the reported error bars. Ashman (2001) portrays the controversy as a dispute between two research groups: one led by de Vaucouleurs which consistently obtained high values around 100 km/s/Mpc, and the other by Sandage and Tammann which consistently obtained values around 50 km/s/Mpc. However, Hodge (1981) reports values published by many groups, and not all of them represent this dichotomy.

At the time, all measurements of the Hubble constant were based on determining distances and redshifts to galaxies in the Hubble flow, where velocities are dominated by the expansion of the Universe rather than peculiar motion due to gravity. This process involved connecting several distance indicators – astrophysical phenomena for which it is possible to determine a distance independently of its redshift – that span distances to our Galaxy, the local group, nearby groups, the Virgo cluster, and beyond. Distance indicators such as Cepheid variables, RR Lyrae variables, H II regions and H I line widths represent very different astrophysical processes, and using them to obtain a value of H_0 involves a complex chain of inference from the collected data, to the astrophysical phenomena the data are said to represent, to interpretations about how these data can be combined, to statements about the expansion rate of the Universe.

This inference chain follows what Pinch (1985) calls *levels of externality* (and see Hoeppe 2014): observational reports are externalized from claims about the local data collection to claims about universal phenomena. In reports with higher externality, "more of the observational situation must be encompassed" (Pinch, p. 9), which involves specifying the *evidential context*, i.e. a body of knowledge, theory, hypothesis, or set of observations. In Pinch's example, the observational reports of a) "splodges on a graph", b) Ar37 atoms, and c) solar neutrinos proceed from low to high externality. Note that this relies on a distinction between data and phenomena (Bogen, 2017): data are "idiosyncratic to particular experimental contexts" (e.g., splodges on a graph), while phenomena have "stable, repeatable characteristics" and depend on the interpretation of the data (e.g., Ar37 atoms, solar neutrinos, or some competing explanation).

When disputes occur, such as in the case of the Hubble controversy, this chain of inference may be attacked at different levels: the observed magnitudes may be in error or improperly calibrated, the astrophysical process used as a distance indicator (e.g., the period-luminosity relation of Cepheids) may be unreliable, the assumption of homogeneity (and thus the use of the FLRW metric) may not apply as expected at the distance scales being probed, or the cosmological model (e.g., LCDM) may be wrong. This process recalls the *Duhem-Quine thesis* that observational statements depend on a network of assumptions, any one of which can be called into question. If a challenge to a measurement or inference is successful, it can no longer support claims at a higher level of externality, with the effect of changing the level of externality and evidential context of the report (Pinch, pp. 15-16). For example, if the assumption of the non-evolution of Supernovae Ia luminosities were to be successfully challenged, then the evidential context of the supernova Hubble diagram shifts from cosmology to stellar and galactic astrophysics: the observational report of cosmic acceleration would no longer be justified, though reports of lower externality may still be valid.

The Hubble 'tension' was a confluence of two main factors according to Ashman: corrections to measurements which tended to act in opposite directions, conflating the difference between the low and high values of H_0 measured by the different groups, and underestimated errors, which turned understandable disagreements into a tension. Astrophysicists take great pains to estimate the level of uncertainty in their measurements, but these error bars may themselves be in error. Observational reports which include an estimate of uncertainty can be called into question both for the values and for the reported errors. Increasing the uncertainty doesn't make a disagreement go away but reduces it to a level that is to be expected without controversy or tension, which turned out to be the case with the previous Hubble controversy. Hodge was prophetic when he assumed "that the existing disastrous disagreements and complicated controversy are merely symptoms of the present adolescent nature of the field" (1981, p. 357).

In addition to the chain of inference which contains observational reports at varying levels of externality, leading from idiosyncratic data to cosmological statements, experimental results also rest on a set of *background assumptions*, not all of which are (or can be made) explicit. In the early days of Hubble's observations, the existence of galaxies other than our own was not generally known or accepted, and the distance scales now taken for granted were not yet even fathomed. Pinch notes that "it is only during a controversy… that such background assumptions come into the foreground" (1985, p. 14), a process which Feest calls "epistemically productive" (2016, p. 37). In this example, the shift in evidential context from 'nebulae' to 'galaxies' was very productive indeed: it increased the size of the known Universe by a shit-ton (scientifically speaking).

The process of accounting for sources of systematic error and building an 'error budget' is one of the *epistemological strategies* that scientists have developed to establish the validity of an experimental result (Franklin 1989, in Feest 2016). These strategies are an attempt to escape what Collins (1985) calls an *experimenter's regress*, which is a circularity between judgments of the validity of a measuring device and those of the measurement result. To Collins, since "experimentation is a matter of skillful practice" (1985, p. 2) which is implicit and non-transferrable, there is no way to resolve disagreements between experiments which claim to be able to measure the same phenomenon, an argument which Feest calls "a modern-day version of Duhemian underdetermination" (p. 35). Franklin, however, points to scientists' epistemological strategies as ways of establishing the validity of an experiment – though tacit knowledge is involved in resolving disputes, the process is nevertheless rational.

Feest frames the debate in terms of *operational definitions* of experimental concepts, which are "rules that lay down paradigmatic experimental parameters" (2016, p. 37). As far as I understand, the description of H_0 as the global expansion rate of the Universe at this point in cosmological time defines the *phenomenon*, while an operational definition would describe typical measurement procedures in which this phenomenon reveals itself. So, one operational definition would involve the relation between distance and redshift as measured by distance indicators, and a more recent definition involves the angular power spectrum of CMB photons. Disagreements may focus on whether an operational definition adequately refers to the phenomenon or on whether a given experiment has applied that definition correctly. The second type of disagreement seems to account for questions of statistical or systematic uncertainties in the measurements, while the first questions theoretical frameworks, though this may be an over-simplification.

Some of the inferential steps involved in measurements necessarily rest on background assumptions. Observational disputes, or general skepticism regarding experimental results, serve to draw out and explicate these hidden auxiliary hypotheses (Feest 2016, p. 39). In the case of the Hubble controversy, one such background assumption was that the Universe must be older than the objects in it. This quite reasonable assumption motivated what Ashman called a "rather non-scientific" statement by his professor that though the methods producing a high value of H_0 "appear to be more sound," the low value is probably correct (2001, p. 107). This is only illogical without acknowledging the background assumption stated above, since H_0 ~ 100km/s/Mpc produces a (linearly extrapolated) age of the Universe that is younger than some reported ages of objects (globular clusters IIRC). So, if you trust these reported ages, and you have no reason to assume anything other than a linear extrapolation – that the Hubble constant does not itself evolve in time – then you are less likely to trust the reported high values, despite their apparently greater trustworthiness. (This example can quite easily be interpreted with Bayesian logic.) Since these prior assumptions influence the inference, they themselves may be challenged; as it turns out, the Hubble constant does indeed evolve with time.

The Hubble controversy highlights the importance of *replication* in science – though in practice, replication is often only attempted for new discoveries or surprising results, "the acceptance of replicability can and should act as a demarcation criterion for objective knowledge" (Collins 1985, p. 19). This *objectivity* has a social quality: the attempt to reproduce results is a way of acknowledging and potentially uncovering individual bias and "subjective preference" (Longino 1990, p. 73). According to Longino, objectivity in science is both the result of a social process and a matter of degree: "A method of inquiry is objective to the degree that it permits *transformative* criticism" (1990, p. 76). Collins points out that successful replication has more confirmatory power when the experiments (or in our case, observational programs) are different: so, distances indicators are sought which do not rely on the same set of assumptions and are not likely to be subject to the same unknown systematic errors, but which nevertheless are still able to be included in the measurement of H_0.

To summarize, observational disputes may be viewed through a number of lenses.

1. The steps along the chain of inference from data to phenomenon may be questioned at different *levels of externality*, e.g., "magnitudes are incorrectly calibrated" (low), or "the local Universe is too lumpy to be adequately described by the FLRW metric" (higher), or "the LCDM model is wrong" (high). Reports of greater externality must specify more of the *evidential context* than those at lower levels.
2. The question of whether an *operational definition* of the phenomenon in question has been applied correctly by a given experiment (or observational program) can be asked, a process which includes accounting for possible sources of error. For example, did a given group measure the Hubble constant, or did they measure the effect of variability in the period-luminosity relation of Cepheids, or the effect of Galactic dust, or statistical noise, or some combination of these processes? By not adequately accounting for systematics, the experiment does not actually measure the desired phenomenon. It can also be questioned whether a given operational definition adequately refers to the phenomenon.

3. Disagreements between experimental results, or disputes about the validity of a measurement, often draw out or explicate *background assumptions* which may involve tacit knowledge about how an experiment works or encompass the theoretical framework within which the measurement makes sense. It is possible, however, that not all background assumptions can be explicated or tested, so there may always be an unresolved source of systematic error or a theoretical resolution yet to be uncovered.

All of these lenses can be seen as expressions of the *underdetermination* of theory by data (see Stanford 2017), the idea that scientific knowledge rests on a complex web of interconnected beliefs, any number of which may be adjusted to resolve a dispute.

References

Ashman (2001) Measuring the Hubble constant: Objectivity under the telescope

Bogen (2017) Theory and Observation in Science. *The Stanford Encyclopedia of Philosophy*

Collins (1985) Changing Order: Replication and Induction in Scientific Practice

Feest (2016) The experimenter's regress reconsidered: Replication, tacit knowledge, and the dynamics of knowledge generation. *Studies in History and Philosophy of Science*

Hodge (1981) The Extragalactic Distance Scale. ARA&A

Hoeppe (2014) Working data together: The accountability and reflexivity of digital astronomical practice. *Social Studies of Science*

Longino (1990) Science as Social Knowledge

Pinch (1985) Towards an Analysis of Scientific Observation: The Externality and Evidential Significance of Observational Reports in Physics. *Social Studies of Science*

Stanford (2017) Underdetermination of Scientific Theory. *The Stanford Encyclopedia of Philosophy*