

# Figures of Merit for ELTs

*Charles Jenkins*<sup>1</sup>

<sup>1</sup>Research School of Astronomy and Astrophysics, Mount Stromlo Observatory, Cotter Road, Weston 2611, Australia; E-mail: charles.jenkins@anu.edu.au

**Abstract:** In discussing the scientific cases for Extremely Large Telescopes, we are discussing science that may be done in a decade or more. This is a long time to look ahead in any field of human endeavor. Can we make very general statements about the scientific productivity of very large telescopes, that will justify the great investment of both people and money? This paper reviews some of the broad-brush attempts to answer this question based on general information-theory or statistical considerations. It seems clear that the gains with aperture are much bigger for some research areas than others, which raises the pointed question of why we think of ELTs as general-purpose instruments.

**Keywords:** Extremely large telescopes, information theory, Bayesian theory

## 1. INTRODUCTION

When we look as far ahead as a telescope that may only be built two decades from now, we are looking across the time span of most activity in a career (before the descent into fame and administration) and trying to guess if we are proposing the right thing. Projects the size of an ELT will engage whole communities and certainly very large amounts of other resources. Compared to other things we might do – build many more 8-m telescopes, for example – the choices are acute and important. Of course we do not have the funding “in the bank” and it may be that we, as a community, are more likely to be able to raise the money for a landmark project than for something that appears to be business as usual, only more so. Then we need to be sure that we are building a scientific as well as a physical landmark, and consider the effect of this venture on the overall health of our community.

The case for a large telescope (doubts about adaptive optics being suspended) is, simply, more signal-to-noise and more resolution. Neither of these, in many cases, scales very fast with aperture. In many interesting areas of research, we have a hazy idea of what we are looking for and we are interested in discovery, rather than incremental improvement in knowledge about known phenomena. Can we quantify the likely discovery rate with large telescopes compared to smaller ones? Of course there are some areas where we know now that an ELT will carry us into new territory; being able to see the main sequence in Virgo, for instance, or detect Earth-like planets around nearby star. But the main sequence may turn out to be exactly the same as in the local group, or there may be no planets to be seen.

Most of us feel instinctively that the jump in angular resolution that an ELT could bring is a stronger indicator of discovery potential than the increase in light-gathering power. This is because of a sense that there is more “phase space” out there in angular terms than there is in mere signal-to-noise; resolution has been a severe and fixed limitation for all of modern astrophysics in the optical/IR.

From the time that telescopes became genuinely astrophysical instruments, increasing the photon rate at the detector has been vital – firstly to overcome reciprocity failure in photographic emulsions, then to overcome detector noise. We are now moving to an area where detectors are

so good that in many cases we can regard long exposures at low photon rate (smaller telescope) as being equivalent to short exposures at high photon rate. The scientific effect of the aperture then separates off from the photon-gathering aspect. Do we then need large telescopes simply so we can do things more quickly than the competition, or because we have more mouths to feed with data?

One relatively simple figure of merit for a large telescope is its cost effectiveness. The cost of telescopes has historically risen as a rather steep power of the aperture  $D$ , at around  $D^{2.5}$ . (See Weaver 2003 for a summary of the arguments and references to earlier work.) One possible reason for this scaling is that so much of the cost of a telescope is in steel and concrete, and the scaling is essentially in volume. What about effectiveness? The advocates of small telescopes have worked at this one and claim that the number of citations rises only as  $D^{1.5}$ . Of course citations should be weighed, not counted, but I have no doubt that we would have made this qualification more weakly had the cost-effectiveness argument gone strongly in favour of ELTs.

Martin Harwit's book *Cosmic Discovery*, although two decades old, is a valuable examination of the history of discovery in astronomy and so gives some clues to what may happen in the ELT era. Harwit defines discovery precisely, if narrowly, in terms of new phenomena which are a long way from anywhere else in a multi-dimensional phase space of observational possibilities. This definition would probably not include a lot of things that we would regard as a discovery, but it certainly includes the "gold standard" of discoveries. Several features of his results are of interest to the discussion about ELTs. One is the discovery rate in the optical waveband. Here, as elsewhere, discoveries seem to follow the introduction of new technologies, and so follow a general exponential growth rate. We would expect this, as the subject would essentially be driven by development in the wider, growing economy. The e-folding timescale for Harwit's list of astronomical discoveries (in the optical) is about 70 years.

So far this would suggest that we would expect new discoveries with an ELT because we are employing new technology to access more of the phase space. This fixes our attention on what we think the axes of this space may be. For an ELT the extra phase space may be rather limited, by comparison that made available by other technological jumps in astronomy.

Harwit makes the extra, audacious step of asking how many discoveries there are to be made. Simply asking this question does challenge the common but unspoken view that Nature is a kind of limitless Aladdin's cave of new things to discover. Harwit is able to make this estimate because of a statistical argument, based on the rediscovery of the same phenomena in different way; we can see intuitively that this kind of duplication does suggest that the number of discoveries is finite.

In fact, Harwit estimates that there are 150 to 300 discoveries to be made. If we take this seriously, and combine the lower of these numbers with the e-folding time for discoveries in the optical, we find that all discoveries will have been made around the time we would complete a 100-m telescope.

Of course, we expect to spend a great deal longer in *understanding* discoveries than in making them – after all, planetary science did not begin and end with Kepler and Galileo. It is also not clear to me that we really know what the axes are in our space of observations. The progress of our subject combines the axes we think of as fundamental in complex ways, and some things we would all agree are discoveries – like Hubble's discovery of the expansion of the universe – do not seem to be new phenomena in Harwit's sense. However this general line of argument suggests that we should not complacently assume that building an enormous telescope will automatically result in the expansion of knowledge; it may only add detail to what we already know.

In his book Harwit foreshadowed the application of the ideas of information theory to these questions, a line of thought begun by Fellgett & Linfoot (1955). Mike Disney, of Cardiff University, developed these ideas in great detail in a paper which, regrettably, is still not published. Disney (1997) repeated Harwit’s classification of discoveries, using a much looser idea of a discovery; essentially, a discovery is what a scientist thinks is a discovery. Nonetheless, Disney arrived at similar conclusions; the number of discoveries grows exponentially, things are rediscovered; the e-folding time in the optical for Disney’s list is about 45 years.

The application of information theory is certainly one way to think of the benefits of ELTs, although it is not as different as one might think from geometrical notions of available phase space. In the following sections I will look at the effectiveness of ELTs from three points of view, one information-theoretical, one standard statistical, and one Bayesian. Of these three approaches it is the Bayesian one that gives the most encouraging conclusion, suggesting that large telescopes can open clear water between themselves and their smaller competitors, across a range of astronomical problems.

## 2. SIGNAL-TO-NOISE RATIO

Calculating signal-to-noise ratio is the standard way of building science cases for telescopes (and other experimental facilities too). However the process can make a lot of unspoken assumptions, and we know that fully general signal-to-noise calculators are cumbersome machines. Nonetheless, if we focus on fundamentals, there are some quite simple conclusions.

One way of looking at an experiment is that it is an attempt to measure a parameter. The best way of doing this (if we forget about Bayesian issues for a moment) is to estimate the parameter(s) by the method of maximum likelihood. For large amounts of data the likelihood function becomes gaussian in the parameters, no matter what statistics go in, and we find that the error in a parameter will be the signal-to-noise  $S$  on the data, scaled by the number  $n$  of observations as follows:

$$e_{\text{model}} \propto \frac{S}{\sqrt{n}}.$$

We all use this formula but not everyone knows that it is very general. The hard part is knowing how much data is “enough” for the asymptotic scaling to hold; and there are various numerical coefficients, depending on the kind of model we are fitting (the right definition of signal-to-noise also depends on this).

If we take an optimistic view and assume that with an ELT we need never worry about detector noise again, then the general expression for the signal-to-noise, in terms of aperture and exposure time  $\tau$ , follows a general form

$$S \propto D^\beta \sqrt{\tau}$$

as we will now see. From the point of view of cost-effectiveness, much depends on  $\beta$ .

For point sources, there are two well-known results; when the images are seeing-limited

$$S \propto D^1 \sqrt{\tau}$$

so the time to a fixed signal-to-noise is  $\propto D^2$ . For a diffraction-limited point source,

$$S \propto D^2 \sqrt{\tau}$$

whereas, for a uniform extended source observed at the diffraction limit

$$S \propto D^0 \sqrt{\tau}.$$

Finally, for a source limited by the Airy wings of an adjacent bright source (no speckles!),

$$S \propto D^{7/2} \sqrt{\tau}.$$

Clearly there is a wide range of scalings with aperture for various applications, and this is even more pronounced if we think of the time to achieve a fixed signal-to-noise. This time is inversely proportional to the productivity, and then we see immediately that the telescope has to be diffraction-limited to be cost-effective. For seeing-limited, background limited observations, a small telescope will be more cost-effective, given the 2.5-power of the aperture in the cost.

We can also apply this kind of reasoning to survey applications. Take a simplified case where the  $\sqrt{n}$  factor is the dominating error in a survey made with field size  $\omega$ . The number of objects to a limiting flux density  $\mathcal{S}$  will be, in a Euclidean approximation,  $\propto \omega 2\mathcal{S}^{-3/2}$ . This gives

$$e \propto \omega^{-1} D^{-3/4} \tau^{-3/8}$$

for the seeing-limited case. The integration time, to get a specified error level, scales as  $D^2$ . We might be interested in “new” objects in the survey, at fainter limits than previously reached; an approximation to the number of these would be the differential count,  $\propto \mathcal{S}^{-5/2}$ . We then get

$$e \propto \omega^{-1} D^{-5/4} \tau^{-5/8}$$

with the same  $D^2$  scaling of integration time. Finally we might consider the benefits of a GLAO system operating on a survey telescope. The corrected field depends on the availability of tip-tilt guide stars, giving roughly  $\omega \propto D^{1/3}$  and

$$e \propto D^{-83/24} \tau^{-5/8}$$

and an integration time scaling roughly as  $D^5$ . Here I have assumed we are observing new sources and made the ambitious assumption that GLAO can produce images which get sharper in inverse proportion to the aperture.

Again we can see that some observations are much more cost-effective than others. Seeing-limited surveys are better done on smaller telescopes and indeed this is the way things seem to be working out.

It might be realistic, given the enormous cost of the focal planes in survey applications, to imagine that they are limited to a certain physical size. This suggests the scaling  $\omega \propto D^{-1}$ . If we took this seriously, survey applications would show an even more negative growth in cost effectiveness with aperture.

In trading exposure time against aperture, which is implicit in arguing for cheaper and smaller telescopes, we do need to be aware of one serious experimental limitation. This is so-called  $1/f$  noise – low-frequency noise, ubiquitous, but of unknown origin. An example in astronomy might be the difficulty in maintaining a spectrophotometric calibration over a long series of exposures. Experience everywhere else in the experimental sciences suggests that after a while, the signal-to-noise does not improve like  $1/\sqrt{n}$ , and it may be that this shadowy threat turns out to be the justification of large apertures for photon-starved work.

### 3. INFORMATION THEORY

The development of information theory goes back to Shannon and was originally concerned with telephone and radio communications. A useful introduction is the Shannon & Weaver book.

Fundamental to information theory is the notion of a bit, which we can think of as the unit of information that gives the answer to a single yes or no question. Three bits, for instance, allows us to isolate one out of eight possibilities. It follows that it is the logarithm of the number of possibilities that is the appropriate measure of information. In an encoding context, if we have an alphabet of  $M$  symbols, transmitted at  $R$  symbols per second, then the information rate is

$$I = R \log_2 M.$$

The conceptual jump in applying this reasoning to telescope issues is to say that the “message” presented to us by the Universe is one of a set of equally likely alternatives, encoded into the intensity of electromagnetic radiation. There is a host of philosophical issues here and I cannot do justice to them in a short space. Disney’s remarkable paper starts from this point, and identifies the relevant alphabet as being the number of distinguishable signals. If we can detect a range of intensities  $I_{\max}$  to  $I_{\min}$  with an error level  $\delta I$ , then the number of characters available to us (and the Universe?) is

$$M = \frac{I_{\max} - I_{\min}}{\delta I}.$$

Since the  $\delta I$  term must involve the telescope aperture, we see from the presence of the logarithm that the gains with aperture of the information will be relatively slow, and this is indeed the argument that Disney develops.

However if we buy this line of argument then we have to consider the symbol rate  $R$  as well. It is clear from the previous section that this is essentially how long it takes to get one-sigma or something similar, and this scales as a power law in the aperture. In fact it makes the logarithmic factor rather irrelevant.

Disney makes a similar but I think stronger argument for the information associated with increased spatial resolution. The reasoning is somewhat subtle, but essentially depends on identifying the number of symbols with the number of independent pixels in a scene. The result, however, emerges as an information term (apart from the symbol rate) which is linear in the aperture, if diffraction limited. This may seem a slow dependence, when we think of how much of resolution space must be unexplored.

I have not done justice to Disney’s analysis, which considers many other aspects of this information theory approach. From the desiccated cost effectiveness point of view, however, I think it leads to similar conclusions as the previous section – not all types of observation are

equally cost effective. However the information theory argument does make explicit that these broad-brush arguments assign equal weight to the range of observational possibilities. While this may be reasonable *a priori* as a measure, I think it is clear that it has great deficiencies. Not all discoveries are equal, because some fit into a much better developed scientific paradigm than others.

#### 4. A BAYESIAN VIEW OF APERTURE CHOICE

Bayesian methods of inference may be able to help us to better understand those aspects of aperture that are concerned with noise, and so with the photon bucket aspect of an ELT. The advantage of the Bayesian paradigm is that it focusses attention on the final result of some experiment, in terms of the choice of one model over another. This is the role of the famous Bayes Factor, or the Evidence (see e.g. Lee 1997). This is very different to calculating a signal-to-noise on some measured parameter or collection of parameters; we may have an excellent measurement, but what we have to do in science is make a *choice*. Of course often we want a good measurement of some generally-relevant parameter (like the Hubble constant) but it is actually the application of this parameter, in some theoretical context like element abundances, where we want to make the choice...Big Bang or Other.

The Bayesian method is to define an exhaustive set of hypotheses, compute the posterior probabilities of each conditional on the available data, and then work out the odds (probability ratio) on the favoured hypothesis over the others. This method, in plain language, tells us which hypothesis to bet on.

For the ELT case I have used the Bayesian paradigm to answer the question: what is the ratio of the odds, for a large and smaller telescope, in favour of a novel hypothesis? This tells us how much we can win if we bet on the bigger telescope to make discoveries.

The actual calculation is for a simplified situation but I believe the essential features are right. Suppose we have two models, H(old) and H(new). Bayes requires us to assign prior probabilities to these; call these  $p_{\text{old}}$  and  $p_{\text{new}}$ . A new model is presumably a surprising one - surprise is a feature of a scientific discovery - so take  $p_{\text{new}} \ll p_{\text{old}}$ . This greatly simplifies the subsequent algebra, as it turns out.

For a specific aperture  $D_1$ , it is then a standard Bayes-factor style calculation to work out the posterior odds on the new model,

$$\mathcal{O}_{D_1} = \frac{p_{\text{new}}}{p_{\text{old}}} |_{D_1}$$

and this can be simplified with two more assumptions. One is that the number of data is large enough for standard asymptotic approximations for the likelihood function to apply, and the other is that the new model *is actually the right one*. The second assumption gets rid of the details of the new model, as it turns out.

Finally I find that the odds ratio for two aperture sizes, employed on the same problem, is given by

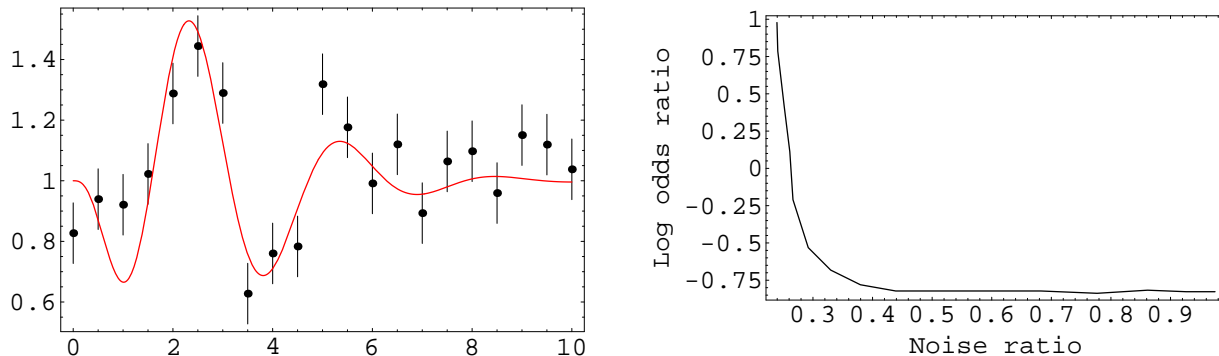
$$\log \frac{\mathcal{O}_{D_1}}{\mathcal{O}_{D_2}} = \Delta\chi^2 + \text{constants}$$

where  $\Delta\chi^2$  is the difference in the usual goodness-of-fit statistic between the fits to the *old* model, for the data obtained with the two apertures in question.

Like many Bayesian results this seems extremely obvious qualitatively; it says that the odds on a discovery are much better if you are able to refute the old model conclusively. Underlying this is the rather deep and perhaps vexed assumption that our set of models spans the space that Nature has provided.

A more detailed implication of this result arises because of the logarithm, and the inverse dependence of  $\chi^2$  on the noise level. This means that the odds ratio is a hockey-stick shaped function; the two aperture sizes have little advantage over each other until a crucial point is reached at which the bigger one accelerates away and has a decisive advantage. This too is a familiar feature; in dealing with noisy data, we find that as the data improve we reach a tipping point where we have enough signal-to-noise and the conclusions become very clear.

To illustrate this, here is a “toy” model-fitting problem, in which I am fitting a constant and an oscillatory function to some simulated data. The oscillatory function is the novelty. Because it is a simulation, the odds ratio is worked out by brute force, with none of the approximations of the calculation I sketched above. However, the result is qualitatively the same; we see that as the relative noise level of one data set improves, it rapidly (beyond some problem-dependent point) gains the advantage over the worse data set, in terms of giving much better odds on the new model.



**Figure 1.** Left, the simulated data and the new model – the oscillating function. The old model is simply a constant. Right, the log odds ratio plotted against the ratio of the noise level for the two imagined telescopes producing the data.

What does this somewhat abstract discussion mean for ELTs? I believe that it shows that a larger aperture will have a decisive advantage, in the precise sense that it will give much better odds on a new result. In other words, betting on new results from a big telescope will make you money. This may sound anodyne but the devil is in the detail; the exact point at which the odds on the large telescope take off is problem-dependent. I have not done the analysis, but it must be the case that the signal-to-noise considerations of Section XX will be relevant here. This must mean that the more “cost effective” problems are an exponentially better bet.

## 5. CONCLUSIONS

This has been a somewhat philosophical discussion, but when we are making decisions which will actually not affect the scientific lives of many of us, we do have to be careful.

A bigger telescope will always be better, in a narrow sense. In a broader sense, we might better optimize the resources of the community by dedicating ELTs to the problems where they

will clearly make a huge impact. However, we then have to have a plan for the aspirations of the rest of the community – more 8-m instruments, for instance. Unfortunately we may not be able to raise the money for this, and it may be politically impossible within our community as well. At a minimum, we should not compromise the design of ELTs in the areas where they will excel, which probably means near-infrared adaptive optics.

A huge effort in the community is going into simulations of observations, in order to build science cases. As a card-carrying Bayesian, I think these analyses would benefit from the Bayes factor approach and the need to choose between definite hypothesis. It would also be very useful, and not that much of an addition to the present amount of work, to build realistic observational schedules and expose more clearly our assumptions about how much salami-slicing of the available time we are prepared to accept to keep everyone “inside the tent”.

## References

Disney, M.J., 1998. The Comparative Discovery Potential of any Astronomical Instrument, unpublished.

Fellgett, P. B. & Linfoot, E. H., 1955. Phil. Trans. Roy. Soc. London A, **247**, 369.

Harwit, M., 1981. *Cosmic Discovery: the search, scope and heritage of astronomy* (Basic Books).

Lee, Peter M, 1997. *Bayesian Statistics*, 2nd Edition (Arnold).

Shannon, C.E. & Weaver, W., 1949. *The Mathematical Theory of Communication*, University of Illinois Press.

Weaver, B, 2003. Astrophysics and Space Science Library, **287**, 21: *The Future of Small Telescopes In The New Millennium. Volume I - Perceptions, Productivities, and Policies*, Ed. Terry D. Oswalt ( Kluwer) .