

# SELECTION EFFECTS AND BIASES

M. DISNEY

*Department of Physics and Astronomy  
University of Wales College of Cardiff*

## 1. Introduction

A respected astronomer recently suggested, because the whole spectrum has been explored, that the heroic days of astronomy are now over. I will try, by referring to the narrow optical and 21-cm windows with which I am familiar, to argue that he was quite wrong. We still haven't covered the whole spectrum, and in any case prejudices, biases and selection effects could still veil much of the truth from sight.

## 2. Human Bias

Human bias can grossly affect the universe we perceive. For instance: (a) Our life-spans are very limited. We are therefore prone to believe, because we hope it is true, that we are rapidly converging upon some true picture of the universe. The notion that millennia may have to pass, that observations lasting for centuries will have to be done, before the picture emerges, are not ones we readily entertain. (b) We are mostly educated as, and therefore tend to think like, physicists. But astronomy is an altogether different subject in which half our major discoveries are still serendipitous. An instance of this is "Mad Big-Telescope Disease," a mind-set borrowed from the Super-Collider school, which believes that only gargantuan machines are capable of new discoveries. In rebuttal simply think of: Expansion of the Universe (24"); Clusters (18"); QSOAL's (84"); Lenses (84"); Large-scale flows (60"); Extra Solar planets (74"); MACHO's (50") ... (c) As a species we are far better equipped to exchange information with each other than fathom it out for ourselves. Virtually all our knowledge is thus second-hand. Any misconception in the network will be difficult to identify, viz. the Flat Earth, and flat Space. (d) As astronomers we have special problems with Occam's razor. It urges us to null hypotheses such as: "If

you can't detect it then assume it's not there" or "If you don't see it vary, assume it's constant." Yes, but ...

### 3. Economic Bias

Economic bias can slant one's ideas dramatically. For instance in  $F(\nu)$  units the quasar 3C 273 appears a strong radio source, but in energy terms (*i.e.*,  $\nu F(\nu)$  units) its radio output is trifling compared to the UV. The reason we think of quasars, and other phenomena, predominantly as radio sources may have to do with an economic bias—the cost per unit area of telescopes at different wavelengths, which are, in (1996) \$US  $m^{-2}$

Radio	mm	Optical	NIR	All other wavelengths
$10^{3-4}$	$10^5$	$10^6$	$10^6$	$10^9$ ( <i>i.e.</i> , Space)

### 4. Wavelength Bias

Have we carried out a first reconnaissance of the whole spectrum, and, if not, how far have we still to go? I now argue that we are only half way there.

Figure 1 shows much of the Cosmos plotted in  $\nu F(\nu)$  units. Some sources are broad-band—*e.g.*, (d) Crab Nebula; (f) Cyg-A; (g) BL-Lac; (h) 3C-273—while the rest are narrow-band or thermal. (a) is a cool bright star, (b) a nearby Elliptical and (c) a nearby Spiral. A crude sampling of the whole spectrum would suffice to find the broad-band sources but the thermal sources could easily elude such a crude inspection. Until we are sure that no class of luminous thermal sources is still evading our census we are still open to surprises. The question then becomes "How many surveys are required to turn up the brighter thermal sources at any temperature?" I claim that the answer is about 19.

To see why, consider the Planck spectrum, which has the same identical shape in Figure 1 at whichever wavelength range you choose to locate it—see the 2.7° CMB spectrum at lower left. How many such Planck spectra will fit across the entire 15 dex frequency range from radio to  $\gamma$ -rays? The answer will depend on how deep you take each spectrum to be from crest to base. One could argue as follows. A preliminary survey of the sky at a new wavelength should detect a few thousand sources, that being the maximum number you could hope to follow up in other windows. But isotropic sources tend to follow the relation  $N(>S) \sim S^{-3/2}$  where  $S$  is flux. For an  $N$  of thousands you therefore want  $S_{\min}/S_{\max} \sim 1/300 = 10^{-2.5}$ . This sets the depth of a Planck function in Figure 1 at 2.5 dex. One can read off the width of the function at this depth; it is about 1.5 dex. In other words

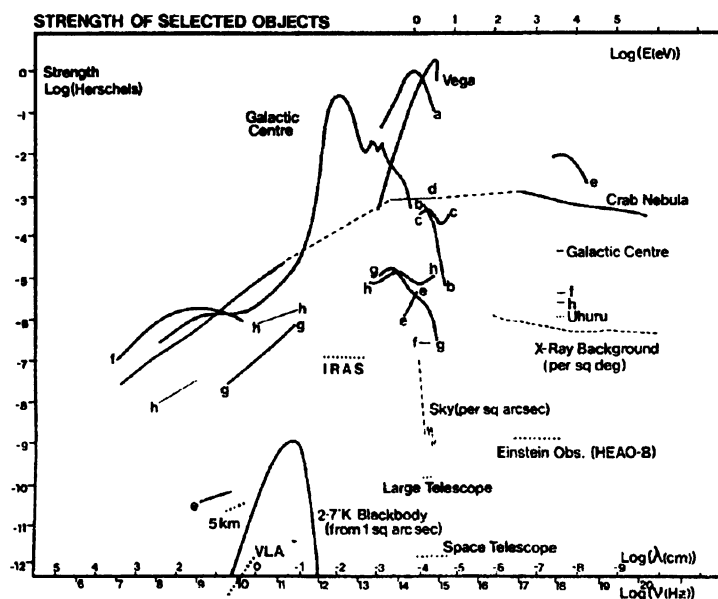


Figure 1. Prominent cosmic objects in  $\nu F(\nu)$  units—in which only the relative ordinate scale is important—see text for labels (Disney and Sparks 1982).

neighbouring surveys for thermal sources, each containing several thousand sources, only overlap in the frequency domain when they are less than 1.5 dex in frequency apart. To be certain no new kinds of sources are missed at the margins we probably need 50 per cent overlap in the frequency domain. Thus neighbouring surveys for a few thousand sources each, may miss new classes of thermal objects if the surveys are situated more than 0.75 dex in frequency apart.

Dividing the total spectral range of 15 dex by 0.75 dex yields a requirement for 20 surveys in all, less one for the UV curtain. Thus 19 (The argument is not strictly applicable longward of the CMB—where thermal sources must be hidden below sky).

The nature of the 19 windows must be such that

$$10^{0.75} = 5.6 = \frac{\nu_{n+1}}{\nu_n} = \frac{\lambda_n}{\lambda_{n+1}} = \frac{E_{n+1}}{E_n} = \frac{T_{n+1}}{T_n}$$

There should be 5 surveys longward of the CMB, 5 more between the CMB and the UV curtain ( $0.1\mu$ ) and 9 high energy surveys between 0.1 keV and 1GeV. For instance, between the CMB and UV curtain we need surveys at roughly:

$\lambda(\mu)$	300 $\mu$	60 $\mu$	12 $\mu$	2.5 $\mu$	0.5 $\mu$	0.1 $\mu$
$\sim T^\circ(\text{K})$	12 $^\circ$	60 $^\circ$	300 $^\circ$	1500 $^\circ$	6500 $^\circ$	36,000 $^\circ$
Survey	None	IRAS	IRAS	Caltech	POSS	None

Counting the ones that have been done is not easy, particularly at higher energy, because the spectral purities are not always so clear and the positions not adequate for identification. My rough estimate is: Optical (1), IRAS (2), Radio (3–4); X-ray (1–3) and  $\gamma$ -Ray (1–3). We have between 11 and 6 more all-sky surveys to do before we can say we've had a good first look at the cosmos. There are obvious gaps at 300 $\mu$  (12 $^\circ$ K), 0.1 $\mu$  (36,000 $^\circ$ K), 5mm, and 100 keV.

## 5. Luminosity Bias

Knowing what classes of object inhabit the universe at a given wavelength is not the same thing as knowing which are significant. For that purpose you need, as a first step, to know the relative numbers of intrinsically luminous and less luminous sources—the so-called “Luminosity Function” (LF). And that is far harder to come by for it requires distances to large numbers of objects. The problem is that the Visibility  $V(L)$  of a source of luminosity  $L$ , that is to say the maximum Volume in which it could lie and still be detectable in a flux-limited survey, rises as  $L^{3/2}$ . Hence luminous sources may be grossly over-represented in such surveys; viz naked-eye stars.

LF's are frequently represented as a power-law of the form  $N(L).dL \sim L^{-\alpha}.dL$  where the index  $\alpha$  may be a slowly varying function of  $L$ . At high  $L$   $\alpha$  must be  $> 2$  for luminosity convergence; at low  $L$   $\alpha < 1$  for number-convergence. Somewhere between  $\alpha \sim 1.5$ , and one generally finds that sources in the LF close to the point where  $\alpha \sim 1.5$  are much the most conspicuous in a flux-limited survey (galaxy astronomers call them “L\*” galaxies); the more luminous ones are too rare and the less luminous too faint to figure prominently. LF surveys containing a few thousand sources invariably reveal a slope  $\sim 1.5$ , which tells one very little about the source population, everything about lack of dynamic-range in such surveys. If  $O(L) \equiv$  number of sources observed per logarithmic interval in luminosity  $L$ , which is what you want in order to measure the LF, then it is easy to show that  $d \log O(L) / d \log L = 5/2 - \alpha$ . One can then ask how far down the LF you can measure if you observe  $N$  distances in total. If you want to go down to  $L/L^* = 10^{-y}$  and you want  $10^x$  objects in the faintest bin—for statistical accuracy, then you must find distances to  $N > 10^x 10^{y(5/2-\alpha)}$  sources. For 20% accuracy in the lowest logL bin you need to observe the following numbers of sources to go  $10^y$  below L\*:

	y=1	y=1.6	y=2	y=2.5	y=3	y=4
$\alpha=1.0$	1,000	4,000	10,000	30,000	$10^5$	$10^6$
$\alpha=1.5$	250	1,000	2,500	7,000	25,000	250,000
$\alpha=2.0$	100	400	1,000	3,000	10,000	100,000

Such numbers warn us that making a census of the universe, even at a single wavelength, will be a hard slog, and that claims to know the LF, based on a thousand or so sources, are naive. Quite apart from statistics, intrinsically sub-luminous objects can only be found close by. But if our neighbourhood is not typical, we will be left with an incurably biased view.

## 6. Surface Brightness Bias

Looking out into the night from a lighted room one sees only other similarly illuminated areas. The darkened buildings and mountains will be hidden beneath the local glare. But we do live in such a lighted room, close to a bright star in a spiral arm of a giant galaxy. From here most structures in the universe may be invisible, or at best very difficult to see, at any wavelength.

Take optical galaxies. It has long been known (Freeman 1970; Disney 1976) that the vast majority of catalogued galaxies have surface brightnesses (SB's) almost identical to the terrestrial sky. Is this an extraordinarily fortuitous coincidence? Or does it mean that the majority of galaxies, even in our neighbourhood, still remains to be found? Which explanation is less incredible? The same SB constancy is true for ellipticals, for dwarfs as well as giants, for blue galaxies as well as red (Disney and Phillips 1985, even though the SB is measured in B!) and, almost incredibly (since  $SB \propto (1+z)^4$ ), for galaxies in the Hubble Deep Field at a median redshift of one (Jones and Disney 1996). What is going on?

It is possible to calculate the Visibility of a galaxy as a function of its SB, and the result is dramatic. Galaxies will be included in a catalogue only if their apparent angular sizes  $\Theta$  and apparent luminosities  $l$ , measured at some limiting isophote, exceed some minimum values  $\Theta_{min}$  and  $l_{min}$ . For a galaxy of given luminosity these isophotal values will be sensitive functions of its SB  $\Sigma$ . If  $\Sigma$  is too high the galaxy will be both physically and apparently small; if too low most of its light will fall below the limiting isophote, implying a faint apparent  $l$ . If one calculates the "Visibility"  $V(L, \Sigma)$  of a galaxy, that is to say the maximum volume within which it can lie, yet still exceed the catalogue limits, then  $V(L, \Sigma) = L^{3/2} \Lambda(\Sigma)$  where  $\Lambda(\Sigma)$  is a uniquely pointed function of  $\Sigma$ , centred at a value defined by  $\Sigma_{CAT} \equiv l_{min} / \pi \Theta_{min}^2$ .  $\Lambda(\Sigma)$  (Disney and Phillipps 1983) is so peaked because it is the lower envelope of two plunging curves, one defined by  $\Theta = \Theta_{min}$

and the other by  $l = l_{min}$ , which intersect at  $\Sigma_{CAT}$ . The fact that most catalogued galaxies have SB's very close to this preferred value could then be a selection effect. Credence is lent to this by the observations of Davies *et al.* (1994) of 918 spirals in the ESO catalogue with redshifts. They find an uncanny fit between the numbers in each SB bin, and the median volumes from which they are drawn. What else could they be observing but a dramatic selection effect which is hiding most of the galaxies, even in our neighbourhood, below the glare of the sky?

When first suggested this conjectured world of "Iceberg Galaxies" was quickly ruled out by HI observers (*e.g.*, Shostack 1977) who claimed they would have picked up their 21-cm signatures in "off-beams" while observing optical galaxies. But their claim was based on the unreasonable implied assumption that galaxies of low surface density would nevertheless have high enough HI columns for them to be observed. But it is easy to show (*e.g.*, Disney and Banks 1996) that 21-cm observations also have a SB, or in this case column-density, limit:

$$N_{HI} \geq 10^{18} T_{sys} \sqrt{\Delta V (kmsec^{-1}) / t_{obs}(sec)} \text{ cm}^{-2}$$

(where  $\Delta V$  is the profile velocity-width) which is independent of telescope size—because small telescopes have bigger beams. And 21-cm observers have traditionally used integration times (Briggs 1990) far too short to reach down to the columns ( $10^{18} \text{ cm}^{-2}$ ) expected of Icebergs with normal  $M_{HI}/L_B$ 's.

Imagine a Shadow World consisting of the same galaxies, at the same distances, as we see in the UGC, but where each Shadow Galaxy is 10 times larger in diameter. The shadow M31 would be almost 40 degrees across. The mean SB of these shadows would be 100 below sky, and their  $N_{HI}$ 's about 100 below current survey limits. The  $10^4$  shadows, each more than  $15'$  across, would cover 1% of the sky. To be sure of finding about 25 of them in a clustered universe one would need to survey  $\sim 1\%$  of the sky (*i.e.*, 12 Schmidt plates, 20,000 CCD frames, 10,000 big radio-telescope beams] to extremely deep limits of  $28B\mu$ ,  $N_{HI} = 10^{18} \text{ cm}^{-2}$ ). We certainly haven't searched for such a possible world seriously so far, and it is unlikely we will find it serendipitously except as QSOALs (where we may have seen it already—Phillipps *et al.* 1991) and as the very occasional "Crouching Giant"—that is to say a giant Iceberg whose higher SB bulge has already been misclassified as a dwarf. Predicted in 1980 (Disney), Bothun *et al.* (1987) found the first Crouching Giant serendipitously in 1987 and nicknamed it Malin 1. That no more have turned up may simply reflect that Malin 1 has an anomalously high  $M_{HI}/L_B$  ( $\sim 5 \times$  solar), and therefore  $N_{HI}$ . Most such Crouching Giants with normal  $M_{HI}/L_B$ 's ( $\sim 0.3$ ) would be far harder to find.

More puzzling is the measured deficiency of high SB galaxies. I can only surmise that in the optical they may be largely opaque—in which case they will turn up in surveys at  $300\mu$  and beyond (Jura 1980, Disney *et al.* 1989, Davies and Burstein 1994).

The future of this field probably lies with deep blind 21-cm surveys made possible by recent developments in technology (Staveley-Smith *et al.* 1996). We are embarking on such a survey with multi-beams at Parkes and Jodrell in 1997. The Doppler effect prevents foreground HI from blinding us to the extra-galactic variety and should enable us to reach down to  $10^{18}$  cm<sup>-2</sup> in limited areas, and therefore to corresponding SB's of 29 B $\mu$  or dimmer.

I have tried to show how dramatic SB selection effects could be in two well-studied wavelengths, and how reluctant we have been to recognise them. Observers at all wavelengths should be alert to similar pernicious effects.

## 7. In Conclusion

I have tried to demonstrate, from one very parochial point of view, that we have the majority of the great astronomical discoveries still to make (see Harwit 1981 for a much deeper discussion of this point). Even in optical astronomy our deepest all-sky survey has been made with 1-metre telescopes with 1 hour exposures and 1% efficient detectors. Is a 36 second glance at the Universe all we are going to need?

## References

- Bothun G.D., Impey C., Malin D. and Mould J. 1987, *Astron.J.*, 94, 23  
 Briggs F.H. 1990, *Astron.J.*, 100, 999  
 Davies J.I. *et al.* 1994, *Mon.Not.R.astron.Soc.*, 268, 984  
 Davies J.I. and Burstein D. 1994, "The Opacity of Spiral Discs," Kluwer  
 Disney M.J. 1976, *Nature*, 263, 573  
 Disney M.J. 1980, "Dwarf Galaxies," ESO Conf Proc, ed Tarenghi and Kjar, 151  
 Disney, M.J. and Sparks, W.B. 1982, *The Observatory*, 102, 231  
 Disney M.J. and Phillipps S. 1983, *Mon.Not.R.astron.Soc.*, 205, 1253  
 Disney M.J. and Phillipps S. 1985, *Mon.Not.R.astron.Soc.*, 216, 53  
 Disney M.J. and Sparks W.B. 1982, *Observatory*, 102, 231  
 Disney M.J., Davies J.I. and Phillipps S. 1989, *Mon.Not.R.astron.Soc.*, 239, 939  
 Disney M.J. and Banks G.D. 1996, *Procs.Astr.Soc.Australia*, submitted.  
 Freeman K.C. 1970, *Astrophys.J.*, 160, 811.  
 Harwit M. 1981, "Cosmic Discovery," Harvester Press.  
 Jones J.B. and Disney M.J. 1997, "HST & the Deep Universe," *Procs Conf* (submitted)  
 Jura M. 1980, *ApJ*, 238, 337  
 Phillipps S., Disney M.J. and Davies J.I. 1991, *Mon.Not.R.astron.Soc.*, 242, 235  
 Shostack G.S. 1977, *Astron.Astrophys.*, 54, 919  
 Staveley-Smith L. *et al.* 1996, *Procs.Astr.Soc.Australia*, submitted.