Theme 8: Beyond the Visible I: radio astronomy

Until the turn of the 17th century, astronomical observations relied on the naked eye. For 250 years after this, although astronomical instrumentation made great strides, the radiation being detected was still essentially confined to visible light (Herschel discovered infrared radiation in 1800, and the advent of photography opened up the near ultraviolet, but these had little practical significance). This changed dramatically in the mid-20th century with the advent of radio astronomy.

8.1 Early work: Jansky and Reber

The atmosphere is transparent to visible light, but opaque to many other wavelengths. The only other clear "window" of transparency lies in the radio region, between 1 mm and 30 m wavelength. One might expect that the astronomical community would deliberately plan to explore this region, but in fact radio astronomy was born almost accidentally, with little if any involvement of professional astronomers.

Karl Jansky (1905–50) was a radio engineer at Bell Telephone. In 1932, while studying the cause of interference on the transatlantic radio-telephone link, he discovered that part of the interference had a periodicity of one sidereal day (23h 56m), and must therefore be coming from an extraterrestrial source. By considering the time at which the interference occurred, Jansky identified the source as the Milky Way. This interesting finding was completely ignored by professional astronomers, and was followed up only by the radio engineer and amateur astronomer **Grote Reber** (1911–2002). Reber built a modern-looking paraboloid antenna and constructed maps of the radio sky, which also failed to attract significant professional attention.

8.2 Radar and radio astronomy

Fortunately for the future of the subject, technological developments during WWII had a direct bearing on the needs of radio astronomy. Many physicists and engineers were recruited into the development and operation of radar, which uses essentially the same technology as radio astronomy. Even during the war, advances were made: **James Hey** (1909–2000) found that some "German jamming" of radar stations was actually coming from the Sun, and unidentified "sporadic" radar reflections were observed. These were later found to come from meteor trails (owing to ionisation of the atmosphere). At the end of the war, a large cadre of physicists with radar expertise returned to academic life and applied that knowledge—plus a significant amount of surplus military hardware—to the new science of radio astronomy. Radio telescopes were set up at Jodrell Bank by **Bernard Lovell** (1913–2012), at Cambridge by **Martin Ryle** (1918–84), and at Dover Heights, Sydney (a former radar station) by **Joe Pawsey** (1908–1962).

Meanwhile, **Jan Oort** (1900–92)—essentially the only professional astronomer to have reacted to the results of Jansky and Reber—realised that long wavelengths might be useful in penetrating the Milky Way's dust lane, and in 1945 his student **Hendrik van de Hulst** (1918–2000) showed that the spin flip transition of neutral hydrogen should produce a radio spectral line at 21 cm. (This is a highly forbidden transition, and is observed in astronomy only because the very low density of interstellar hydrogen suppresses the usual collisional de-excitation of this state.) The line was first observed by **Ewen** and **Purcell** at Harvard, and a little later by **Muller** and **Oort** in Leiden, in 1951. One of the first achievements of radio astronomy was the 1958 mapping of the Galaxy's spiral arms using the 21 cm line, a joint project of Oort's group at Leiden and a group at Sydney under **John Bolton**[1] (1922–1993).

The search for discrete sources of radio emission met with mixed success. Discrete sources were indeed found, first by Hey in 1945, and later by a number of groups, particularly Martin Ryle at Cambridge, but the poor resolution of single-dish radio telescopes (and even short-baseline interferometers) prevented the identification of these sources with optical counterparts. Since the radio emission was a continuum rather than a line spectrum, this made it impossible to measure the redshift of the objects, and hence their nature (small, faint and local, or large, bright and extremely distant) was entirely unclear. Accurate position measurements required interferometric techniques or special circumstances. The Crab Nebula was recognised as the radio source Taurus A by the Sydney group in 1949, using an ingenious technique whereby a receiver on a sea cliff uses the direct and reflected signals to create an interference pattern, and Cygnus A was identified as a peculiar galaxy from observations by **F. Graham Smith** (b. 1923) of the Cambridge group using more conventional interferometry in 1951. In Manchester, the combination of the big Mk I telescope and a small portable dish in the late 1950s allowed accurate source positions and sizes to be obtained by interferometry: this was the first evidence that some sources were extremely small (~1").

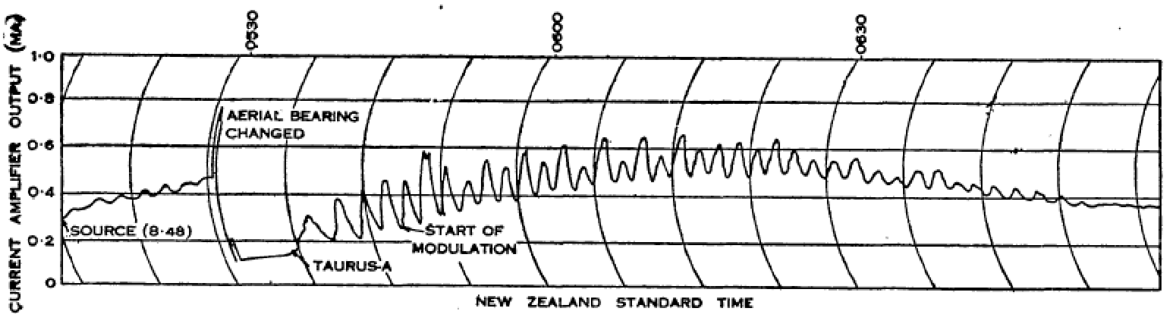

Figure 8.1: interference fringes from Taurus A, obtained by the Sydney group using a two-antenna interferometer at Leigh, New Zealand (JG Bolton and GJ Stanley, Aus. J. Sci. Res. ***2*** *(1949) 139–148).*

Even with the majority of sources still unidentified, source counts can provide useful information. If we assume that radio sources are uniformly spread throughout the Universe and all have approximately the same intrinsic brightness, then the number of sources with apparent brightness greater than some threshold b_0 should be proportional to $b_0^{-3/2}$ (because, from the inverse square law, $b \propto 1/r^2$, and the number within radius r is obviously proportional to r^3). When this exercise was carried out by the Cambridge group (Martin Ryle and colleagues) in the period 1954–1960, an excess of faint sources was found. This was claimed to be "conclusive evidence against the steady-state model" of cosmology (Ryle, *MNRAS* **122** (1961) 349), because if the sources are all intrinsically similar then an excess of faint sources implies a greater density of sources at larger distances (and hence longer look-back times). As it is axiomatic in steady-state models that the gross properties of the Universe do not change with time, this is unacceptable. However, at the time, the relatively low proportion of identified sources allowed

[1] who came originally from Sheffield and was educated at King Edward VII School, according to Wikipedia.

proponents of the steady state model to argue that the evidence was not in fact conclusive: for example, a population of faint Galactic sources plus one of bright extragalactic sources could produce the required pattern (as pointed out by Dennis Sciama in 1963 and 1964).

8.3 Quasars and radio galaxies

The early radio astronomers had a tendency to call their sources "radio stars"—perhaps because the first discrete source clearly identified, by Hey in 1942, was the Sun, or perhaps just because stars are the quintessential point sources. In 1950, Ryle, Smith and Elsmore (*MNRAS* **110** (1950) 508) noted that their 50 Northern hemisphere "radio stars" did not cluster close to the Galactic plane, and must therefore be either nearby faint objects or very luminous extra-galactic sources, and that they did not in general coincide with obvious optical objects. At that time, they rejected the latter idea because the sources were small and bright compared to what would be expected of our own Galaxy as seen from outside, and that extremely high-temperature gas would be required to attain the necessary radio flux. Also, only four weak sources coincided with known galaxies (M31, M33, M51 and M101): as these are very nearby galaxies and were *weak* sources, Ryle et al. naturally assumed that more distant galaxies would be undetectable. They therefore concluded that the majority of radio sources originated within our own Galaxy, and that they corresponded to a common but "hitherto unobserved type of stellar body, in which a very intense radio emission is associated with a very small visual intensity." This is of course special pleading, but in essence they felt that postulating a relatively small and faint class of unidentified radio emitter was preferable to the large, luminous unidentified emitters needed for the extragalactic theory.

In the 1950s, this view began to shift, as improved interferometric positions for the radio sources allowed more visual identifications. The first unequivocally extragalactic source was Cygnus A, identified by **Baade** and **Minkowski** (note that professional astronomers are now taking an interest!) as a result of a high-precision position measurement by Smith. By 1954 Baade and Minkowski felt confident enough to classify radio sources into four types: "(I) remnants of supernovae; (II) galactic nebulosities of a new type [though those listed, Cas A and Puppis A, are now also considered supernova remnants]; (III) peculiar extragalactic nebulae and (IV) normal extragalactic nebulae", and in 1958 Minkowski (*PASP* **70** (1958) 143) posed the same dilemma as Ryle, Smith and Elsmore but came to the opposite conclusion: the preferred solution was now luminous, distant extragalactic sources. However, Minkowski still comments on the disappointingly small fraction of radio sources with identified optical counterparts.

In 1959, the influential **Third Cambridge Catalogue** of radio sources (3C) was published. This was the first detailed survey of Northern hemisphere radio sources (its predecessors were restricted in declination). Several groups, particularly **Allan Sandage** (1926–2010) and colleagues in California, began systematic searches for optical counterparts, expecting to find mostly faint galaxies. However, many objects remained unidentified, and often the radio positions were not precise enough to be certain of identifications.

In 1962, **Cyril Hazard** (b. 1928), working in Sydney, observed the source 3C 273 as it was occulted by the Moon. This source had been identified with a faint galaxy, but the occultation allowed Hazard to obtain a much more accurate position which was clearly not coincident with the galaxy. Instead, it matched a thirteenth-magnitude blue star, which in Palomar 200-inch photographs seemed to be accompanied by a faint streak of nebulosity (now recognised as a jet). The occultation measurements had shown the source to be a close double: one region matched the "star", and the other the jet.

Several other 3C radio sources were also identified with blue starlike objects: indeed, 3C 48 had already been so identified by Sandage in 1960, in an unpublished conference paper. They were therefore referred to as quasi-stellar (radio) sources, subsequently abbreviated to **quasars**. Despite their starlike photographic appearance, the quasars did not have very stellar spectra: in particular, they displayed a strong ultraviolet excess, and initially unidentifiable spectral lines. The line problem was resolved in 1963, when **Maarten Schmidt** (b. 1929) realised that the spectrum of 3C 273 could be understood if its redshift was an almost unprecedented 0.158. Shortly afterwards, **Greenstein** and **Matthews** determined the redshift of 3C 48 as an even more dramatic 0.367. Initially, astronomers considered the possibility that these might be gravitational redshifts, but the masses required to produce this level of gravitational field did not seem to be physically possible. The possibility that the redshifts were Doppler shifts caused by local motion—i.e., that the quasars were small objects ejected from galaxies—was dismissed because they had no observable proper motion (i.e. the source of the ejection would have to be very close to Earth, rather than, say, the Galactic centre), and because no equivalent objects were observed being ejected from other galaxies. The other option was that the redshifts were cosmological, and the quasars were at enormous distances. By 1965, Sandage was calling quasars and their radio-quiet cousins, quasi-stellar objects, “a major new constituent of the universe” and arguing that they would “provide a crucial test of various cosmological models.”

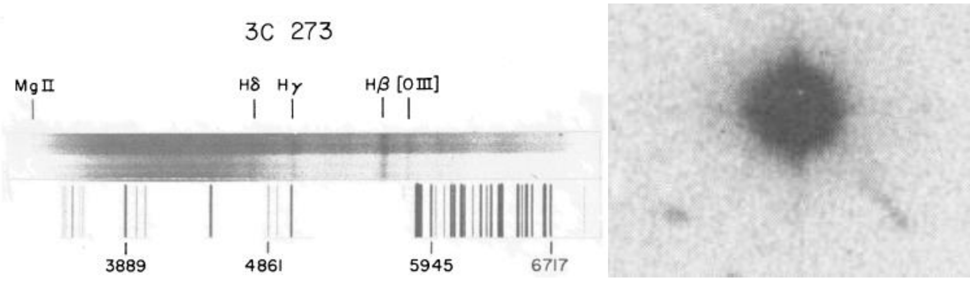

*Figure 8.2: the quasar 3C 273. Left, spectrum of the “star” taken in January 1963, with laboratory comparison spectrum. The top half of the spectrum has three times the exposure of the bottom half. Right, image taken by Allan Sandage using the Palomar 200": the jet is visible at lower right. Pictures from JL Greenstein and M Schmidt, ApJ **140** (1964) 1–34.*

A 1964 review paper by Matthews, William Morgan, and Schmidt (*ApJ* **140** (1964) 35) distinguished between “weak” radio sources, most of which were normal spirals, and “strong” sources, which included D galaxies (extended ellipticals: Morgan being a co-author, the paper of course uses Morgan’s galaxy classification scheme, which includes this class), quasars, and “N galaxies”. N galaxies have starlike nuclei, somewhat like weak quasars (I assume the “N” stands for “nuclear” or “nucleated”), but are less luminous in the radio than quasars and are surrounded by a visible galaxy; nowadays they tend to be included with the Seyfert galaxies[2]. The idea that radio emission was caused by galaxy collisions, which had been popularised by the visual appearance of Cygnus A and Centaurus A, had fallen out of favour as a result of the much more normal appearance of other radio-loud galaxies. This paper also recognises that the radio sources have disparate morphologies: among the spatially resolved objects, the authors distinguish “doubles”, “simple sources” (single Gaussian peak) and “core-halo objects” (a small, bright source superimposed on faint extended emission)—although many of the sources were still not resolved at all.

Matthews, Morgan and Schmidt regard their different classes of radio sources as genuinely distinct: although they recognise the photographic similarity of N galaxies and quasars, the radio

[2] Seyfert galaxies had actually been described earlier (in 1943) by Karl Seyfert, on the basis of their optical appearance.

luminosities seem sufficiently different to separate them. The underlying idea of "active galactic nuclei" did not appear until much later: in James Hey's 1971 textbook *The Radio Universe*, the idea of a relation between quasars and radio galaxies is still regarded as extremely speculative (and Hey is thinking of an evolutionary relation—quasars turn into radio galaxies—not a simple difference in orientation).

The extragalactic radio sources are an example of a genuinely unexpected discovery. Some categories—especially the quasars—were completely new and unprecedented, and even the "normal" D galaxies had no obvious mechanism for their intense radio emission. The visual appearance of quasars was sufficiently distinctive that, once they had been diagnosed as a distinct class of objects, it was possible to devise methods of identifying them optically, using their ultraviolet excess. Sandage then found that many objects with these optical characteristics were not intense radio sources—the **quasi-stellar objects** or radio-quiet quasars—but these would not have been discovered until very much later, if at all, without the diagnostic information provided by the identification of the radio loud quasars. The discovery of radio sources was not, however, accidental: the early radio astronomers set out deliberately to search for discrete sources of radio emission, even though there were no obvious reasons to expect any. This is certainly not an example of Popperian falsification—there are no predictions to falsify—and it isn't an obvious case of a Kuhnian paradigm either. This is not very surprising, because this type of discovery is fairly uncommon in the experimental sciences (though it does happen: for example, the discoveries of the tau lepton and high-T_c superconductors). It is, however, entirely typical of modern astronomical discoveries: surveys in newly opened technological windows have almost invariably resulted in the identification of unexpected sources.

It is appropriate that one of the first uses of the radio source catalogue was a test—however inconclusive—of a cosmological model. The radio sources opened up the high redshift universe, and hence, because of the finite speed of light, the early universe. For many years, the redshift record was always held by a radio source. The redshift distributions of quasars and of classical radio galaxies were not consistent with a steady state cosmology, although by the time this was definitively established the discovery of the cosmic microwave background had already sounded the death knell of the steady state.

8.4 Pulsars

Radio galaxies, though not predicted, were discovered as a result of systematic searches. **Pulsars**, on the other hand, were discovered entirely accidentally. The study being carried out by the Cambridge group in 1967 was intended to observe *scintillation* of known radio sources caused by ejections of gas from the solar atmosphere. The source traces showed patches of high-frequency interference, most of which was caused by artificial radio transmissions; however, **Jocelyn Bell** (b. 1943) was sufficiently alert to note that the timing of one piece of interference indicated that it was coming from a particular region of the sky, and was therefore not terrestrial. Further studies by Bell and her advisor, Anthony Hewish, showed that the signal consisted of short, extremely regularly spaced pulses with a period of about 1.3 s. This was so unusual that for a brief period the hypothesis of alien artificial origin was seriously entertained (the source was nicknamed "LGM-1", for "little green men"). However, the spacing of the pulses was sufficiently regular that the Doppler shift caused by a planet's orbit would have been detected (indeed, 30 years later such a shift was used to detect planets around a pulsar); moreover, a few months later Bell found a very similar source, with a slightly shorter period, in a different part of the sky. The chance of two alien civilisations producing essentially identical beacons was considered infinitesimal, and a natural explanation was therefore required.

Regular pulses from a natural source can be produced in two ways: by rotation or pulsation. The minimum possible rotation period of a given body can be calculated by equating the centrifugal and gravitational forces: $GM/R^2 = v^2/R = 4\pi R/P^2$, so the radius required for a given period is $R = (GMP^2/4\pi^2)^{1/3}$. Taking a typical stellar mass, say the Sun's mass, and a period of 1 s gives a maximum radius of about 1500 km, which is too small even for a white dwarf. Pulsation periods can be estimated similarly: because pulsation is effectively a sound wave, it is related to the density, since the speed of sound in a gas is given by $(dP/d\rho)^{1/2}$, leading to an estimate for the period of order $4(GM/R^3)^{-1/2}$. For a period of ~1 s, this implies $R \sim 4000$ km, again rather small even for a white dwarf[3]. Therefore it was quickly recognised that the discovered objects must be even smaller than white dwarfs. Baade's and Zwicky's hypothesised neutron stars fitted the bill, and rapidly became (and remain) the accepted explanation; the pulses are associated with the rotation period of the neutron star. The fact that young, rapidly rotating pulsars are often associated with supernova remnants, e.g. the Crab Nebula and the Vela supernova remnant, supports Baade's and Zwicky's suggestion that neutron stars are produced in supernovae.

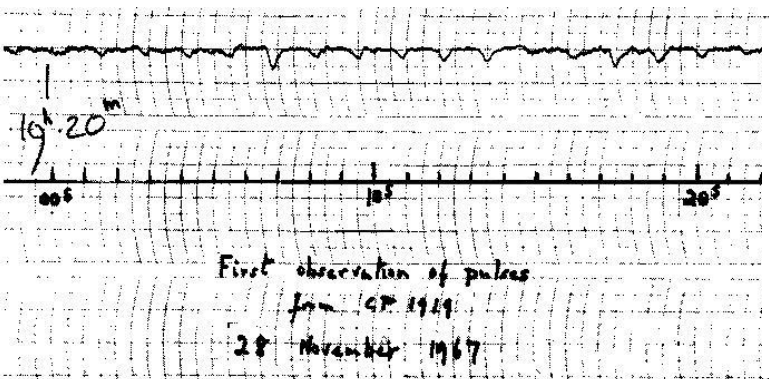

Figure 8.3: record of the first pulsar. This is the "high-speed" chart recording which showed that the signal was periodic. Note that the height of the pulses varies considerably, but the interval between pulses does not. Image from http://www.cv.nrao.edu/course/astr534/Pulsars.html

8.5 Radio spectroscopy

Since radio waves have larger wavelength and lower frequency than visible light, radio spectroscopy is not associated with the atomic transitions that produce optical spectral lines. The main sources of low-energy transitions are the rotational and vibrational transitions of molecules. Even these are often at the short end of the radio window, in the microwave and submillimetre region.

As mentioned above, the first spectral line detected in the radio region was the 21 cm line of neutral hydrogen. This was used to map the Galaxy's spiral structure, and since that time has been instrumental in determining the rotation speeds of our Galaxy and others. It remained the only detected radio spectral line for over a decade, until the 18 cm line[4] of the OH radical was observed in 1963. In 1965, some OH lines were found to be produced by masers (the microwave analogue of lasers), indicating a population inversion in the emitting region. Astronomical masers are of great practical interest because their extremely narrow line widths make it possible to detect very small Doppler shifts, and hence make precise radial velocity measurements. Coupled with high precision spatial information from very long baseline inter-

[3] There *are* pulsating white dwarfs, such as the ZZ Ceti stars: they have periods of minutes.

[4] lines, actually; there are four, at 17.4, 18.0 (two, very closely spaced), and 18.6 cm.

ferometry (VLBI), these measurements can be used to determine, for example, the distance of stars with maser-emitting gas shells near the Galactic centre, and the mass of the supermassive black hole in the nearby galaxy NGC 4258.

The advent of microwave radio astronomy in the mid-1960s, most famous for the prompt discovery of the cosmic microwave background, also greatly increased the availability of detectable spectral lines. NH_3 (ammonia) was detected in the direction of the Galactic centre by Cheung et al. in 1968 at a wavelength of about 1.25 cm. The detection of CO at 2.6 mm by Wilson, Jefferts and Penzias (yes, the microwave background Penzias and Wilson) in 1970 opened up a useful tracer for molecular gas: the obvious molecule to track, H_2, is too symmetrical to produce many spectral lines, so CO (a compound of the third and fourth most abundant elements) is generally used instead. To date, more than 130 different molecules have been identified in interstellar space, including various organic species with up to 13 atoms.

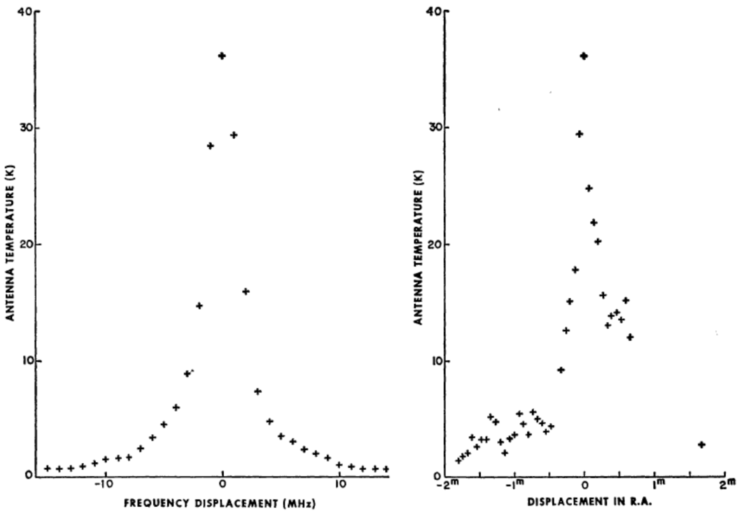

*Figure 8.4: detection of CO in the Orion Nebula by Wilson, Jefferts and Penzias, ApJ **161** (1970) L43–L44. Left, the line profile: intensity (antenna temperature) against frequency shift relative to 115267.2 MHz; right, the spatial distribution: intensity at peak frequency as a function of displacement in right ascension.*

The expected frequency for CO was 115271.2 MHz; the difference is attributed to a Doppler shift.

8.6 Summary

Radio astronomy has been and remains a technology-driven field: the pace of discovery is generally set by the availability of the necessary technology. Very few radio astronomy discoveries were made on the basis of theoretical predictions: there is one clear exception, the 21-cm spectral line, one near-exception, the cosmic microwave background, which comes into the “accidental” category because the original prediction by Alpher and Herman in 1950 had been forgotten, and the search planned by Dicke’s group in Princeton in 1965 was “scooped” by Penzias and Wilson, and one subfield that works differently: the molecular spectroscopists rely on theoretical predictions of line frequencies to tell them where to look in frequency space, and on pre-existing observational information to tell them where to look in real space (i.e. regions known to have the right conditions for molecules).

The impact of radio astronomy on astrophysics and cosmology has been much greater than the astronomical community of the 1940s would have expected. The unexpected prevalence of extragalactic sources opened up the high-redshift universe and provided the first evidence for nonstellar energy sources in galaxies. Pulsars offered information about matter in a highly exotic state (a degenerate neutron gas), and pulsars in close binary systems have since been used to make precision tests of general relativity in strong gravitational fields (work which won Hulse and Taylor the 1993 Nobel Prize). Radio astronomy also demonstrated that the “jets” observed in many active galaxies sometimes appear to move faster than light (not, of course, a real velocity, but an optical illusion caused by the angle of view; nevertheless, for the illusion to work the jets must be highly relativistic). Very Long Baseline Interferometry (VLBI) provides very high angular resolution, which, coupled with the fact that radio waves are not absorbed by interstellar dust, means that radio astronomy has often provided the highest resolution images of the Galactic centre and active galactic nuclei, though in more recent years this has been challenged by infra-red imaging using either the HST or large ground-based telescopes with adaptive optics.

In short, radio astronomy has played a key role in the development of astronomy in the past 50 years. Although many of the domains opened up by radio astronomy are now extensively studied at other wavelengths (e.g. AGN by X-rays, the Galactic Centre by X-rays and in the infra-red), the discoveries that established the field in question were made by radio astronomers and could not have been made until much later, if at all, by other techniques. It is worth noting that most of this was driven by the development of radar, which in turn was motivated by military requirements in WWII: the pre-war efforts of Jansky and Reber had met with little interest from the astronomical community (apart from Jan Oort), and it is not at all obvious that radio astronomy would have developed as a serious professional field of study in the absence of the war.