Colliding and Merging Galaxies

FRANÇOIS SCHWEIZER

Aided by advances in computer technology and observations from space, astronomers have begun to unravel the mysteries of galaxy formation and evolution. Galaxies evolve by interacting with their environment and especially with each other. During brief but often fierce galactic encounters, gravitational forces generate strong tides that survive as telltale signatures for billions of years. Because these so-called collisions dissipate orbital energy, galaxies on bound orbits may eventually merge. Collisions and mergers are responsible for a great variety of phenomena, including the triggering of widespread star formation in galaxies and the fueling of nuclear activity in quasars. Evidence is accumulating that not all galaxies formed shortly after the Big Bang. A sizable fraction of them may have formed later, and many are still experiencing significant dynamical evolution.

THE PAST DECADE HAS WITNESSED A REVOLUTION IN OUR understanding of how galaxies evolve. No longer are galaxies seen as "island universes" evolving slowly in splendid isolation, as was commonly believed until the 1970's. Rather, galaxies are now known to interact in a multitude of ways with their environment, which consists of satellites, neighbor galaxies, and often vast masses of tenuous hot gas. Among these interactions, gravitationally induced collisions and mergers of galaxies have become the best documented, and possibly also the dominant, evolutionary mechanism.

This article reviews the antecedents of this revolution, then discusses the simple physics underlying galaxy collisions and describes the observational evidence for mergers, and finally comments on such issues as "cannibalism" in clusters of galaxies, the stoking of the fires of quasars, and the delayed formation of elliptical galaxies through mergers of spirals.

In 1940, the Swedish astronomer Erik Holmberg put forward the hypothesis that galaxies may collide, experience friction, and eventually merge (*1*). He pointed out three relevant facts: (i) galaxies are likely to experience close encounters because their mean separations are relatively small, only of the order of 10 to 100 times their diameters (*2*); (ii) during an encounter, they induce tides in each other that cost them orbital energy; and (iii) as a result their orbits must shrink, and the involved galaxies may eventually even coalesce. To estimate the tidal forces and orbital energy loss, Holmberg resorted to simulations of pairs of interacting galaxies with two sets of movable lamps arranged on a large table (*3*), thus foreshadowing the important role that computer simulations would play decades later.

Despite this ground-breaking work, colliding and merging galaxies received little attention during the next 30 years, for several reasons. Most astronomers thought of encounters between galaxies as rare random events; they ignored the evidence presented by Holmberg that most galaxies do not move at random but orbit in double and multiple systems with recurring opportunities for close encounters. Also, even Zwicky's extensive observations of interacting galaxies (*4*) during the 1950's were not accompanied by similar progress in theory, partly because of a prevailing notion that gravitation could not produce the long, thin filaments of luminous material photographed by Zwicky (*5*). Finally, the discovery in 1963 of quasars and their enormous energy outputs convinced many astronomers for more than a decade that gigantic explosions were taking place in galaxies, explosions that presumably also explained the above-mentioned filaments; only during the mid-1970's was this explosion hypothesis gradually abandoned for lack of any supporting evidence.

Just as the hypothesis of continental drift lay dormant until the discovery of mid-ocean ridges and sea floor spreading, the hypothesis of colliding and merging galaxies had to await a series of major discoveries beginning around 1970 before it received widespread acceptance. In a landmark paper, Alar and Juri Toomre (*6*) demonstrated convincingly that simple gravitational models of interacting galaxies reproduce in great detail the luminous "bridges and tails" observed by Zwicky and later by Arp (*7*). Their emphasis that fierce, inelastic collisions must soon end in galaxy mergers led to further theoretical studies (*8*) and, since the late 1970's, to the identification of an increasing number of merging and merged galaxies by observers (*9*). During the past 3 years, enthusiasm for mergers has increased because quasars have been discovered to be associated intimately with colliding galaxies and because the Infrared Astronomical Satellite (IRAS) has disclosed dozens of hitherto inconspicuous galaxies that, apparently as a result of recent collisions and merging, experience vast bursts of star formation and emit great amounts of energy in the infrared.

Underlying the recent excitement and activity is an old puzzle: the dichotomy of galactic structures. Stars in galaxies often seem to be arranged in two distinct components: a spheroid and a disk. The relative proportions of spheroid and disk stars differ from galaxy to galaxy. Figure 1 shows two extreme cases, an elliptical galaxy consisting only of a spheroid of stars and a spiral galaxy consisting mainly of a disk (*10*). The traditional view (*11*) holds that both components formed within a few hundred million years after the Big Bang, that is, within the first few percent of the age of the universe. The spheroid stars supposedly formed rapidly out of free-falling protogalactic gas clouds, whence they move in random directions. The disk stars, on the other hand, condensed more slowly out of gas that had already settled into a rotating disk (*12*). According to this picture, the type of a galaxy is fixed shortly after the Big Bang, and subsequent evolution from spiral to elliptical type, or vice versa, is excluded. The Toomre brothers suggested a drastically different scenario: ellipticals may be star piles left over from spirals that collided and merged well after the Big Bang (*6, 13*). This hypothesis was first almost ignored but has steadily been gaining support from

François Schweizer is a staff member of the Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, and an adjunct staff member of the Mount Wilson and Las Campanas Observatories.

recent work on mergers, which seems to be tipping the balance of opinion in favor of a delayed formation or modification of at least a sizable fraction of galaxies.

Tidal Friction, Orbital Decay, and Violent Relaxation

The physics of stellar systems that collide and merge in essence involves only Newtonian gravitation. Compared with motions in the solar system, the main complication arises from the large numbers of stars involved, typically 10^9 to 10^{11} per galaxy. Obviously, these numbers preclude a detailed treatment of all stellar orbits. Astrophysicists have been forced to simulate interactions between galaxies with much smaller numbers of stars, typically 10^2 to 10^4 per galaxy. Even so, progress has depended on advances in computer technology. Yet the basic phenomena of the merging process were understood well before the advent of modern computers. The three most important phenomena are tidal friction, orbital decay, and violent relaxation.

Tidal friction is what makes colliding galaxies merge eventually, despite the fact that few stars, if any, will ever hit each other (*2*). This friction is a generalization of a force first described by Chandrasekhar (*14*) as "dynamical friction." For simplicity, imagine a massive object moving through a layer of uniformly distributed stars, all at rest. Much as a ship forms a wake that slows it down, the object will interact gravitationally with stars along its path, set them in motion, lose kinetic energy, and decelerate. Dynamical friction can get very strong when two galaxies overlap during a close encounter. But even at separations of a few diameters, galaxies induce tidal deformations in each other that cost them orbital energy. The surprise in the early 1970's was just how fierce these tides can be, and how strong the resulting friction (*6*).

Orbital decay is a natural consequence of tidal friction. Just as artificial satellites spiral down and return to Earth because of drag experienced in the upper atmosphere, two tidally interacting companion galaxies move on shrinking orbits. The final decay and merging occur relatively rapidly, typically during one orbital revolution ($\sim 10^8$ years), because (i) tidal forces increase approximately as the inverse cube of the remaining separation and (ii) resonant coupling builds up between the orbiting galaxies and groups of stars rotating within them. Figure 2 suggests how two interacting disk galaxies might look at different stages of their orbital decay.

In their final stages of merging, two galaxies produce a rapidly changing gravitational field that scatters and redistributes the stars into a characteristic equilibrium configuration. This redistribution process is called violent relaxation. In 1967, Lynden-Bell (*15*) calculated the form of the equilibrium configuration and showed that it resembles to a striking degree the observed light distribution of ellipticals. Although he had in mind the rapidly changing gravitational field of a collapsing protogalaxy, his results apply equally well to the rapidy fluctuating field of two merging galaxies of comparable mass. Thus one would expect that their merged remnant resembles an elliptical regardless of the structure of the progenitors. It was this reasoning that led the Toomres to suggest that ellipticals may be remnants of merged spirals.

During the past decade computer models have been used extensively to study and check these processes (*8, 16, 17*). The models have not only confirmed the efficiency of tidal friction and violent relaxation but have also yielded values for the orbital decay rates. For instance, merging occurs much faster when galaxies rotate in the same sense as they orbit than when they rotate in the opposite sense. Also, for merging to occur, the initial approach velocity must be slow lest the galaxies hurry past each other without much tidal deformation. Above all, the computations have given us a better understanding of the detailed merger physics. Mergers occur because galaxy collisions are highly inelastic. In a single collision, up to 50 percent of the orbital energy may be converted into internal energy of the galaxies. Finally, the stars escaping in tidal tails carry much angular momentum with them, making it possible for two galaxies to sink toward their common center of mass and merge.

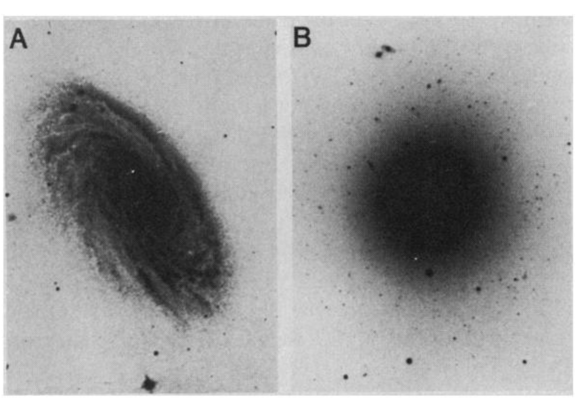


Fig. 1. Two fundamentally different types of galaxies: (A) disk galaxy with spiral arms and (B) elliptical galaxy. There is increasing evidence for the hypothesis that ellipticals may form through mergers of disk galaxies.

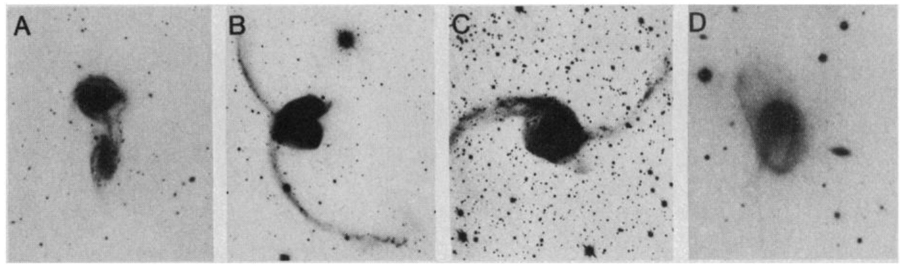
Observational Evidence for Mergers

Given this theoretical understanding of the merger process, what do observations say about merging galaxies, and what are the properties of remnants? Among the approximately 5000 bright galaxies listed in the century-old *New General Catalog* (NGC), a few dozen have so far been identified as possible merger remnants (*9*). Yet only about six have been observed in sufficient detail to make a reasonably compelling case for their being recent mergers. Figure 3 shows photographs of three of these.

The first, NGC 1316, has an elliptical-like body embedded in an extensive envelope about 200,000 parsecs in diameter (roughly seven times the size of our own Milky Way; 1 parsec = 3.26 light-years $\approx 3 \times 10^{13}$ km). From this envelope protrude two major filaments, which have been interpreted as the tidal tails of small disk galaxies that fell in about 10^9 years ago (*18*). The masked print (Fig. 3) reveals that the inner body is marked by "ripples," arclike features that consist of moderately old stars and are a frequent signature of infallen disk galaxies, as will be discussed below. Spectroscopic observations also show that buried deep within the galaxy lies a disk of ionized gas that rotates much faster than the stars (about 400 km sec^{-1} compared to the stars' 140 km sec^{-1}) and at nearly a right angle to them. Because of its high rotation velocity and tilt, the gas cannot have been left over from early days or been shed by evolving stars. Rather, it is likely to have been freshly supplied by the merging galaxies.

The second galaxy of Fig. 3, NGC 5128, has an active nucleus that emits radiation across the whole electromagnetic spectrum from γ- and x-rays to radio wavelengths. First believed to be a system of two colliding galaxies during the 1950's, then a single galaxy in the aftermath of a major explosion during the 1960's, NGC 5128 has recently been determined to be a merger remnant (*20*). Like NGC 1316, it is an elliptical with a disk of gas and dust rotating at nearly right angles to the stars and faster than they (Fig. 3B). And, again like NGC 1316, NGC 5128 shows signatures of an infallen disk

Fig. 2. Interacting disk galaxies at different stages of orbital decay: (A) NGC 5426/5427, a pair of weakly interacting disks; (B) NGC 4038/4039 (nicknamed "The Antennae"), two colliding disk galaxies each with a long tidal tail; (C) NGC 3256, a pair of colliding disks that appear already partially merged; and (D) NGC 3921, the presumed remnant of two merged disk galaxies now identifiable only by their tidal tails. [Panels A and D are reproduced from photographs by V. C. Rubin and H. C. Arp, respectively]



galaxy in the form of arc-shaped ripples that pervade its body. Numerical simulations of the merger place its beginning at about 1 to 2×10^9 years in the past (*21*).

The third merger of Fig. 3, NGC 7252, is perhaps the best understood. Two long tidal tails (projected lengths, 80,000 and 130,000 parsecs) show unambiguously that two disk galaxies of about equal mass collided. Yet despite the indication of two participants, there is now only one body. The most direct evidence that there were two progenitor galaxies stems from velocities measured in the ionized gas of this body: they clearly reveal two separate, surviving motion systems (*22*). Apparently, two large spirals, each perhaps the size of our own Milky Way, began merging about 0.5×10^9 to 2×10^9 years ago, the time being calculated from the tail lengths divided by their velocities. After coalescing they formed a single, relatively symmetric, although messy-looking remnant with a radial light distribution characteristic of an elliptical galaxy rather than of the original spirals. This suggests that mixing and violent relaxation have been efficient in rearranging the stars. Therefore, NGC 7252 provides a small but solid piece of evidence in favor of the delayed formation of ellipticals through mergers of disk galaxies.

Ripples in Ellipticals

Elliptical galaxies used to be thought to possess extremely smooth light distributions. Yet in 1979, while experimenting on astronomical plates with high-contrast copying and unsharp masking techniques, Malin discovered faint arcs and plumes around otherwise normal-looking ellipticals (*23*). Well over 100 galaxies with such structures have been found by now, thanks to advances in photographic emulsions and processing (*9, 24*). Some of them were already known to Arp, who included them in his *Atlas of Peculiar Galaxies* (*7*). These faint arcs and plumes seem to be closely related to ripples and tails in galaxies believed to be merger remnants (Fig. 3). Hence, their discovery in ellipticals opens up the exciting prospect of learning about accretion and merger events there.

Ripples occur from near the centers to the outermost regions, forming incomplete arcs that encircle typically less than half the elliptical and rarely number more than ten. On prints produced by unsharp masking (*25*), they show sharp outer edges and look like shells of luminous material. In some especially striking cases, they appear aligned on opposite sides of the galaxy and interleaved in radius; that is, ripples of increasing radius lie on alternate sides (see cover). Their physical properties are difficult to measure because of the low contrast. On brightness profiles of ellipticals, ripples appear as minor steps of about 1 to 10 percent in the local light distribution. Optical spectroscopy shows that most of the visible light does not stem from gas (*18, 26*). Photoelectrically measured colors, on the other hand, suggest that ripples consist of stars. The colors are on average slightly bluer than the colors of the underlying ellipticals and agree best with the colors of old stars in disks of spiral galaxies (*26*).

The frequent association of ripples with tidal tails, their similarities to features observed in merger galaxies, and their colors all suggest that ripples may be formed when material from a disk galaxy, or even a whole galaxy itself, falls into an elliptical during a close passage (*18*). This hypothesis has recently gained theoretical support from numerical simulations of such events made by Quinn (*27*); his models suggest that ripples are indeed the distorted remains of the former disk material. The top two rows of Fig. 4, from a simulation by A. Toomre, illustrate this accretion process

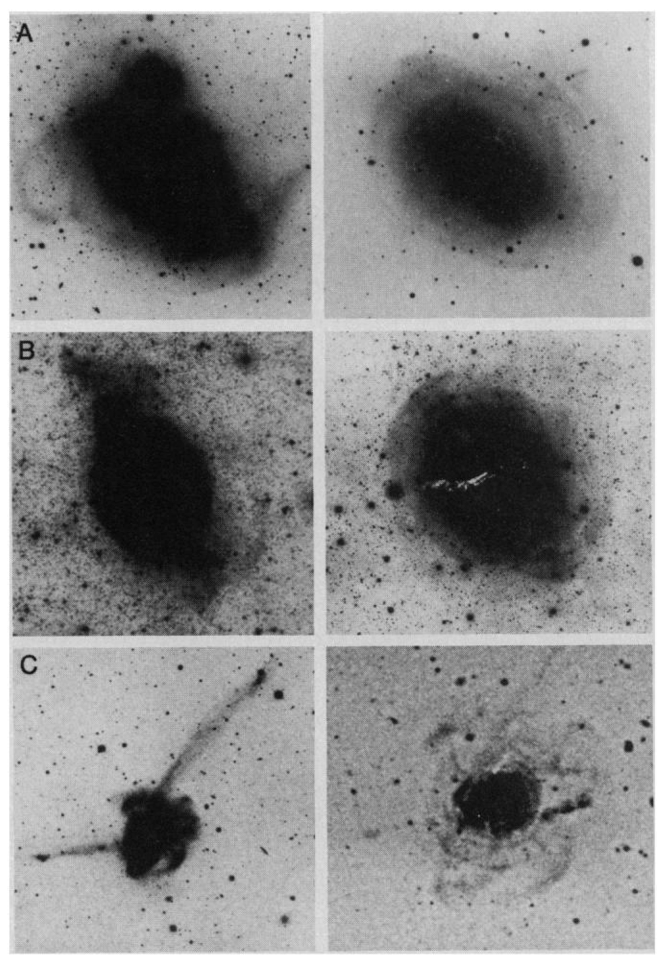


Fig. 3. Three remnants of recent mergers: (A) NGC 1316, (B) NGC 5128, and (C) NGC 7252. Each galaxy is shown twice side by side: in a high-contrast photograph emphasizing faint outer tails and wisps (left) and in an enlarged and masked print (*25*) showing inner ripples (right). [Photographs of NGC 5128 are reproduced from Cannon (*19*) and Malin *et al.* (*20*); the masked print of NGC 7252 was processed digitally as described (*50*)]

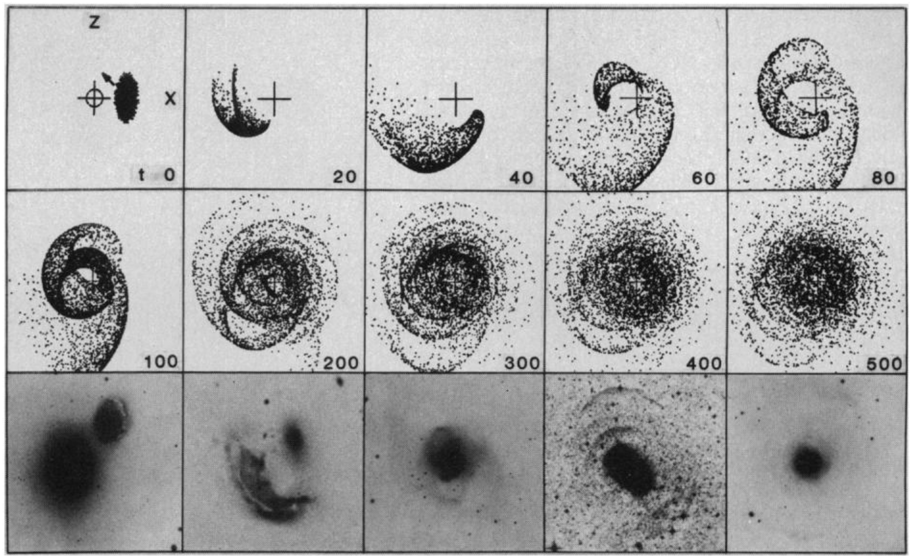


Fig. 4. Accretion of a disk galaxy by an elliptical. The top two rows show a model simulation of the event. A disk of test particles (top left, time zero) is "dropped" into the elliptical (marked here only by a cross) in the direction of the arrow. The orbiting disk is bent and stretched by differential gravity and wraps itself around the center of the elliptical (t = 20 to 80). As the wrapping progresses, more and more ripples appear wherever the distorted surface of the former disk is viewed edge on. For comparison, the bottom row shows five systems observed in nature (left to right): Arp 116, Arp 142, Arp 222, NGC 3923, and Arp 227. [The top two rows are by A. Toomre. The Arp galaxies are reproduced from Arp (*7*), and NGC 3923 is from Malin and Carter (*23*)]

especially clearly. A disk of test particles representing a galaxy falls into an elliptical and gets increasingly stretched and distorted by gravitational forces. As it wraps itself an increasing number of times around the center of the elliptical, the disk material forms a series of ripples that become more numerous but also weaker. After about ten revolutions (Fig. 4, last frame), the ripples become too weak for detection; this number of revolutions corresponds to about 1×10^9 to 2×10^9 years in a typical giant elliptical.

Even when the disk galaxy plunges directly into the center of the elliptical on a radial orbit, the phenomenon of ripple formation persists. Quinn's calculations show that such rare bull's-eye collisions may be responsible for the striking geometrical patterns of ripples interleaving on opposite sides (*27*). As stars of the disk intruder oscillate through the center of the elliptical with different periods, they reach different heights depending on their initial energy. At the turning points of their orbits, they crowd together with stars of similar energy and form ripples. The most distant ripples, or "shells" in Quinn's terminology, contain stars with the highest orbital energy, and successive ripples on opposite sides contain stars that differ in phase by half an orbit. If we had 10^9 years to watch the aftereffects of such a spiral-elliptical collision, we would see individual stars wandering back and forth between ripples on opposite sides and the ripples themselves propagating outward as density waves, a truly majestic form of music of the spheres.

This emerging picture of collisions between spirals and ellipticals still needs observational checking and theoretical refinement. For instance, the models predict ripple velocities that can barely be measured by present techniques (*27, 28*). Theoretical work suggesting that even a small elliptical falling into a large elliptical may form ripples needs to be followed up (*29*). More attention needs also to be paid to the possibility that ripples may form through mass transfer from a spiral to an elliptical during a close encounter without subsequent merger, as suggested by the observation that many ellipticals with ripples have spiral companions (such as Arp 227, 228, and 229 and NGC 596, 4382, and 5018).

An important issue remaining to be settled concerns the total amount of matter accreted by ellipticals up to the present date: does the ripple material represent only icing on a preexisting cake, or does it signify that most ellipticals acquired the bulk of their luminous matter in a succession of accretion processes of this sort? A study of ripple frequencies in a sample of relatively isolated ellipticals suggests that, over the age of the universe, typical bright ellipticals have probably accreted material from four to ten disk galaxies (*9*). However, we cannot tell from observations whether more than a few percent of the mass arrived in each event. Yet there is at least a hint from merger dynamics that fairly rapid accretion may occur only for mass contributions of 10 percent or more. If this is true, then ellipticals would seem to be built substantially from matter acquired in installments. This random accretion would then be the likely cause of the unexpectedly slow rotation of bright ellipticals.

S0 Galaxies with Polar Rings

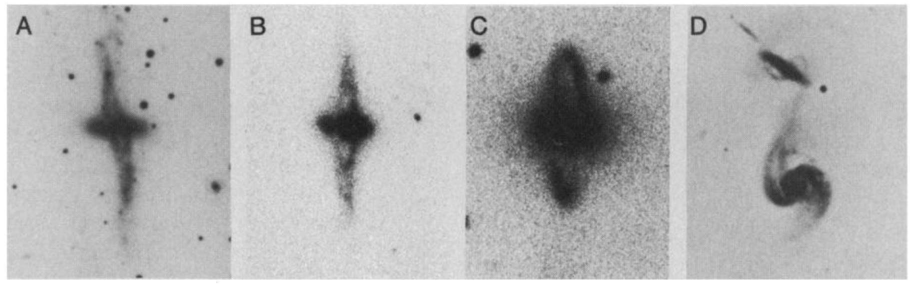
In recent years, astronomers have discovered still another category of galaxies whose peculiar structures seem to be due to collisions and mergers. The first three panels of Fig. 5 show a sample of these galaxies, which have been described variously as "spindle," "cigar through ring," and "Saturn-like." They are rare, with barely two dozen specimens known; these specimens are, on average, distant and faint (*30*). The main reason why they have attracted attention is the apparent spindle shape of their central bodies.

For some time now, astronomers have suspected that ellipticals may have triaxial shapes and that at least some of them ought to be prolate—that is, cigar-shaped—as opposed to oblate or flattened at the poles like spinning planets; yet it has been unexpectedly difficult to find supporting evidence. Therefore, apparent spindle galaxies such as those in Fig. 5 raised new hopes for finding prolate ellipticals.

However, all three galaxies studied in detail so far have turned out to be disks seen nearly edge on, and there are reasons to believe that most remaining examples are also disks and of Hubble type S0 (this morphological type designates disk galaxies that are nearly devoid of gas and lack spiral structure). Virtual proof of the disk nature of the three well-studied galaxies is their fast rotation, which exceeds the internal random motions, whereas the opposite is true in ellipticals (*30, 31*); hence prolate ellipticals continue to elude us. Yet the discovery that the apparent spindles are inclined S0 disks is no less exciting: as Fig. 5 shows, the surrounding rings pass nearly over the poles of the S0 disks, and measurements demonstrate that the two components rotate approximately at right angles (unlike the body and rings of Saturn, which share the same rotation axis). How did these S0 galaxies acquire polar rings made of gas, dust, and young stars?

It seems unlikely that the collapse of any single protogalactic gas cloud could have led to the formation of two disks rotating at right

Fig. 5. Three S0 galaxies with polar rings and one related object: (A) NGC 4650A, (B) A0136-0801, (C) AM2020-5050, and (D) NGC 3608. The apparent spindles are S0 disks of old stars seen nearly edge on and surrounded by polar rings of gas, dust, and young stars believed to be debris from other galaxies. Panel D illustrates the mass transfer that often takes place when two galaxies collide and that may lead to the formation of a polar ring around the smaller galaxy of the pair. [NGC 3608 is reproduced from Arp (*7*)]



angles. Rather, the formation history of these galaxies must have included some second event. Two hypotheses are (i) that a small gas-rich companion fell in over the poles of an S0 disk and disintegrated (*13*) and (ii) that a mass transfer during a close passage between a spiral and an S0 galaxy placed some material into orbit around the latter (*30*). How this might occur is shown in the last panel of Fig. 5. Probably both processes occur, so that some S0 galaxies with polar rings may owe their formation to close collisions and others to nearly complete mergers. In this picture, the near-polar orientation of rings results quite naturally from a selection process. Accretion of material occurs presumably at random angles, but only rings formed within about 20° from the poles survive for periods comparable to the age of the universe. Rings at larger angles precess too rapidly in the gravitational field of the S0 disk and settle within a few billion years into its equatorial plane. Although many details about the formation and evolution of these objects remain to be worked out, the hypothesis that collisions and mergers play a crucial role seems increasingly secure. In at least two cases so far, deep images reveal that the systems are embedded in telltale messes of luminous wisps and ripples (*32*).

Cannibalism in Clusters of Galaxies?

Clusters of galaxies contain typically from a few dozen to several thousand members. Near the centers of some clusters lie unusually large and bright galaxies called supergiant D (cD) galaxies. These galaxies have elliptical-like centers surrounded by faint, extensive envelopes that often permeate a sizable fraction of the cluster. They have received a great deal of attention since Ostriker and Tremaine proposed in 1975 that they grow and develop their special form by consuming other cluster members (*33*). Here again, the cause would be the dynamical friction experienced by cluster galaxies orbiting in the extended envelope. This friction would cause the cluster members to spiral inward and finally merge with the cD "cannibal."

Despite much observational work, solid evidence for the occurrence of this process has been hard to come by. One reason is the remoteness of cD galaxies, which renders detailed observations difficult. Probably the most widely accepted evidence in support of cannibalism has been the presence of "multiple nuclei" in about one-quarter to one-half of all cD's (*34*). These galaxy-like condensations within a projected distance of about 10,000 parsecs from cD centers have been interpreted as remains of galaxies being cannibalized. Yet unambiguous morphological signatures of gravitational interaction, such as tidal tails or ripples, have generally not been observed.

In the past 2 years, new theoretical work and observations have raised various doubts about this cannibalism picture for cD galaxies. New calculations show that merger rates in cD's are lower than previously believed, and mergers are likely to have added only a few percent to the present-day mass of these galaxies since the formation of the clusters (*35*). Photometric observations of bright cluster members suggest that there may be a continuous transition in the luminosity distribution from non-cD to cD types, the latter simply being the brightest galaxies; hence their extended envelopes may no longer require a special formation mechanism (*36*). Also, high-precision measurements of colors of cD's have not detected the blueing predicted by cannibalism models; cD galaxies appear as red as normal giant ellipticals (*37*). Finally, a sizable fraction of multiple nuclei are now known to have velocities too high relative to the cD's to be regarded as slowed-down cluster members spiraling in; rather, they seem to be cluster members appearing superposed by chance or plunging for a brief period to the center of the cluster (*38*). Either way, they are unlikely to be cannibalized soon.

In response to these difficulties, various alternate hypotheses for cD formation are being considered (or reconsidered), among them being the theory that cD envelopes may consist of tidal debris accumulating from the whole cluster and the hypothesis that cD's may grow continuously through condensation of infalling cluster gas. It is also possible that the cD's arose from mergers of lesser galaxies, but mostly during the violent phases of cluster collapse rather than continuously since then. Given the complexity of phenomena at the centers of galaxy clusters, it may be some time before the relative importance of the various competing mechanisms, including mergers, can be established.

Stoking the Furnace, Starbursts, and IRAS

During the past 2 years, the focus of research on colliding and merging galaxies has shifted from processes involving only stars to processes also involving gas. It has long been known, of course, that spiral and irregular galaxies contain typically 5 to 20 percent of their mass in the form of gas, but the complexity of phenomena in supersonically colliding gas clouds discouraged theoretical work for many years. Nevertheless, there were good reasons to believe that collisions and mergers may produce far more spectacular effects in the gas than in the stars. The Toomres had already asked in 1972 (*6*): "Would not the violent mechanical agitation of a close tidal encounter—let alone an actual merger—tend to bring ***deep*** into a galaxy a fairly ***sudden*** supply of fresh fuel in the form of interstellar material, either from its own outlying disk or by accretion from its partner? And in a previously gas-poor system or nucleus, would not the relatively mundane process of prolific star formation thereupon mimic much of the activity that is observed?"

Even earlier, Zwicky (*4*) and Arp (*39*) had noticed the frequent presence of extremely bright, star-forming regions near the centers of interacting galaxies. Studies done since then at wavelengths from x-rays to the radio range have confirmed that interacting galaxies exhibit signs of central activity unusually often (*40*). In some cases this activity is dominated by a powerful central source of energy, which occupies a region less than 1 parsec in diameter and is now widely believed to be an accretion disk around a massive black hole. In many other cases, however, the central activity extends over much larger regions, 10^2 to 10^3 parsecs in diameter, and seems to be

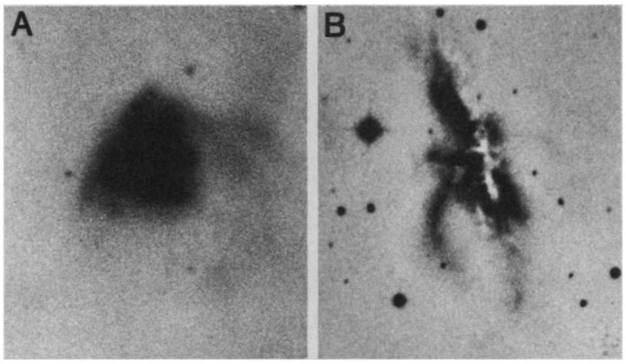


Fig. 6. Two ongoing mergers with strong central bursts of star formation and extraordinarily high infrared luminosities detected by IRAS: (A) Arp 220 and (B) NGC 6240. [Arp 220 is reproduced from Arp (*7*) and NGC 6240 from Fosbury and Wall (*42*)]

caused by exceptionally large numbers of hot, luminous young stars still embedded in their parent gas. Galaxies with such hot spots are called starburst galaxies because of the bursts of star formation that they have experienced in the recent past.

Although starburst galaxies have been studied optically for several years, IRAS detected new examples with extraordinarily high rates of star formation. IRAS was launched in 1983 and scanned the sky for 10 months at wavelengths of 12 to 100 μm, a range that was previously unobservable from the ground (*41*). At these infrared wavelengths, its detectors sensed objects with temperatures of 30 to 300 K; in external galaxies these objects tend to be large clouds of gas and dust particles heated by starlight. The big surprise from the IRAS observations was that some colliding and merging galaxies emit up to 99 percent of their total luminosity in this part of the spectrum, whereas normal galaxies emit typically less than half their luminosity there. In these interacting galaxies, supersonic collisions of gas clouds seem to have triggered gigantic starbursts. Millions of newly born stars emit ultraviolet and visible light as usual, but the surrounding dust is in the way, absorbs most of this light, and finally reemits the absorbed energy in an avalanche of infrared photons. Figure 6 shows Arp 220 and NGC 6240, two candidate mergers that have recently attracted attention because of their enormous starbursts and infrared luminosities (*42, 43*).

A series of optical and near infrared follow-up studies of IRAS discoveries is beginning to yield a glimpse of fierce, although transient, phases of galaxy evolution in colliding and merging galaxies. The star-formation bursts are generally concentrated within a few kiloparsecs from the centers. Most of the luminosity originates in very massive stars (*44*). As the merger progresses, the luminosity rises and seems to peak relatively early, with some observed values exceeding 10^{12} solar luminosities (4×10^{45} ergs sec^{-1}). Gas and dust are converted to stars at rates of up to several hundred solar masses per year ($\sim 10^{28}$ g sec^{-1}), or several tens of times faster than normal. At these rates, the gas contents of average spirals would be depleted in 10^7 to 10^8 years, a natural upper limit on the duration of peak activity. As gas clouds collide with relative velocities of several hundred kilometers per second, they produce massive quantities of molecular gas, including shock-heated H_2 that glows in the near infrared (*43*). Merger signatures such as tidal tails and ripples take typically 10^8 to 10^9 years to develop, which is longer than starbursts last. Hence it is possible that some ongoing mergers such as Arp 220 (Fig. 6) and NGC 6052 do not yet show obvious morphological signatures, although they may develop them in the future. By the time these signatures themselves are fading, the fireworks of star formation may have long subsided. This probably explains the relatively modest infrared excesses of mergers such as NGC 3921 (Fig. 2) and NGC 7252 (Fig. 3).

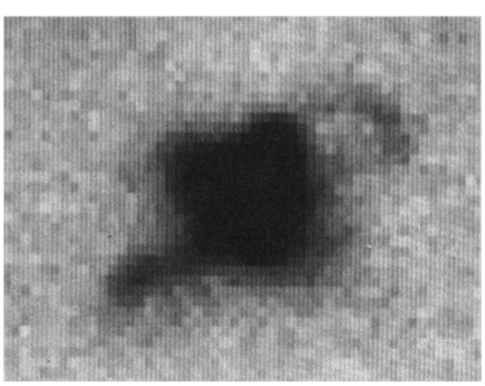


Fig. 7. Quasar 3CR249.1 with two tails of ionized gas lit by the central energy source. The tails suggest that a collision of two disk galaxies, and perhaps even a merger, is taking place. The image was obtained by Stockton and MacKenty (*47*) with a photoelectric array detector placed behind an interference filter that passes mainly a bright emission line of doubly ionized oxygen ([O III]λ5007) at the redshift of the quasar.

Even the activity in quasars may be seriously indebted to galaxy collisions. After Stockton's (*45*) discovery of unusually compact companions within a few kiloparsecs from the centers of three quasars, surveys of nearby quasars have shown that about one-third of them appear to be interacting with neighbor galaxies (*46*). Among isolated quasars, two have been found to possess double tails whose gas is lit by the intense ultraviolet light from the central energy source; they are suspected to be mergers similar to NGC 7252 (*47*) (Fig. 7). Therefore, the same mechanisms of gas transport that trigger starbursts may occasionally fuel even more energetic nuclear activity in colliding and merging galaxies, turning some of them into quasars.

Formation of Ellipticals

Let us return briefly to the hypothesis that ellipticals may be remnants of merged disk systems. As was discussed earlier, this hypothesis was made mainly to explain the puzzling dichotomy of galaxy structures. When confronted with other observational facts, the new hypothesis led naturally to a number of worries, some of which seemed serious 10 years ago. Interestingly, of the half-dozen or so main worries, most if not all have diminished since then.

For instance, for many years there appeared to be reasons for fearing that (i) the mixing occurring in mergers would have erased the chemical-abundance gradients of the progenitor spirals, whereas similar gradients are also observed in ellipticals, and (ii) the well-established positive correlation between the luminosity of ellipticals and the heavy-element content of their stars could not have arisen in mergers. Both worries were based on the simplified concept of mergers involving only stars, and both were thought to favor the early formation of ellipticals from gaseous protogalaxies. Yet the recent IRAS results on starburst galaxies have shown beyond doubt that furnaces are being stoked with gas even in present-day collisions and mergers. What more feasible way could there be for building abundance gradients in galaxies than to have starbursts enrich the central gas with heavy elements ejected by massive stars and supernovae? And is the mean heavy-element content of remnants not bound to increase with each successive merger and luminosity increment? It is gradually becoming clear that the stellar bulk of the mass in spirals, far from being a hindrance, may actually act as some sort of a piston that forces gas clouds to collide with a vehemence otherwise unknown, resulting in super-efficient star formation and heavy-element enrichment.

Other worries have diminished similarly because of new insights. Many ellipticals are known to be more centrally concentrated than spirals, and simple mass and energy conservation arguments used to suggest that mergers would lower rather than raise the degree of central concentration. Yet recent numerical simulations have shown that the central density tends to increase on average and can more than triple when gas clouds are included in the models (*17*). It also used to be thought that giant ellipticals and dwarf ellipticals form a continuous sequence, but dwarf ellipticals are smaller than spirals and could not have arisen from mergers of them. Yet new observations have shown convincingly that in the majority of their measurable properties giant and dwarf ellipticals form two disjointed sequences; the dwarf ellipticals now appear to be related more closely to disk systems that have been stripped of their gas than to giant ellipticals (*48*). Finally, former worries about the extremely low gas content of ellipticals when compared to the presumed progenitor spirals have now nearly turned into an asset: observations of starburst galaxies identified by IRAS suggest that the energy transferred to the gas by supernovae alone far exceeds the minimum energy necessary to set up a hot galactic wind and blow all gas out of the remnant (*49*).

Whereas arguments against a merger origin for ellipticals have waned, arguments in favor are waxing. An extensive set of numerical simulations of disk mergers by Negroponte and White (*17*) has yielded remnants in remarkably good accord with observed ellipticals: their shape is always ellipsoidal and can be oblate, triaxial, or prolate depending on the orientation and orbital configuration of the progenitor disks. The apparent flattening of model remnants spans the full observed range, and the random and rotational motions are also roughly as observed. On the observational side, the elliptical-like light distribution of the merger remnant NGC 7252 (Fig. 3) has been mentioned already. The discovery of a considerable variety of fine structures in ellipticals (*23, 50*) is showing that these galaxies are not the mathematically simple equilibrium configurations they were once thought to be, and telltale signatures are increasingly pointing to a multiple-component origin. For instance, one elliptical has recently been found to possess an equatorial gas disk that rotates in the opposite direction from the stars (*51*).

Yet the most powerful argument in favor of a merger formation of ellipticals remains the one advanced already by the Toomres (*6, 13*). When extrapolated into the past, the statistics of ongoing disk mergers shows that about one-tenth of all galaxies in the *New General Catalog* should be remnants of disk mergers. If these remnants are not ellipticals, which make up just about this fraction of the catalog, where else have they gone?

Outlook

As with any promising new theory that is catching on, there is now danger of going overboard with the concept of galaxy mergers. For instance, one should beware of calling every messy object a merger. Although there can be little doubt that at least some ellipticals have formed through disk mergers and many more have accreted luminous matter, it is far from established that other ellipticals may not have formed from gaseous protogalaxies. Therefore, the debate about the formation and evolution of galaxies, dubbed by some as the debate of nature versus nurture, promises to continue for some time.

Yet a new picture of galaxy formation and evolution has begun to emerge despite many unanswered questions. The formation of galaxies seems to have been not nearly as rapid and complete as was thought a decade ago: even at the present epoch, the death of merging spirals can lead to the birth of new ellipticals. The delayed transformation of galaxy types, once an anathema, is now an observed reality. Perhaps even some well-established spirals continue to accrete satellites and build bigger bulges from time to time. Collisions and mergers are no longer of importance only for gas clouds in collapsing protogalaxies. Rather, they are now observed in action throughout the aging universe, igniting some distant young quasars as well as reshaping some nearby old galaxies. Through them we have learned to recognize the importance of galactic processes such as mass loss and mass transfer. Galaxy evolution was once believed to be detectable only in the faint smudges of remote galaxies at high redshifts, yet we have begun to unravel it faster and better from bright nearby galaxies. Detailed studies of their collisions and mergers are helping us to reconstruct the glory of galaxy evolution in the past, when our universe was denser and more interactive.

REFERENCES AND NOTES

1. E. Holmberg, *Astrophys. J.* **92**, 200 (1940).
2. In contrast, most stars will never collide because their mean separations are typically 10^8 times their diameters. Stellar and galactic velocities being roughly equal (20 to 200 km sec^{-1}), a star is about 10^{12} times less likely to collide with another star than a galaxy with another galaxy. Even when two galaxies, each with 10^9 to 10^{11} stars, collide, virtually all their stars intermingle without suffering collisions themselves.
3. E. Holmberg, *Astrophys. J.* **94**, 385 (1941).
4. F. Zwicky, *Ergeb. Exakten Naturwiss.* **29**, 344 (1956); *Handb. Phys.* **53**, 373 (1959); ——— and M. L. Humason, *Astrophys. J.* **139**, 269 (1964).
5. Noteworthy exceptions to this mistaken view were the early numerical simulations of interacting disk galaxies by J. Pfleiderer [*Z. Astrophys.* **58**, 12 (1963)] and the first analytical estimates of energy exchange in colliding galaxies by S. M. Alladin [*Astrophys. J.* **141**, 768 (1965)].
6. A. Toomre and J. Toomre, *Astrophys. J.* **178**, 623 (1972).
7. H. C. Arp, *Atlas of Peculiar Galaxies* (California Institute of Technology, Pasadena, 1966).
8. For excellent reviews, see S. Tremaine, in *The Structure and Evolution of Normal Galaxies* (Cambridge Univ. Press, Cambridge, 1981), pp. 67–84; S. D. M. White, in *Internal Kinematics and Dynamics of Galaxies* (Reidel, Dordrecht, 1983), pp. 337–345.
9. For a review, see F. Schweizer, in *Internal Kinematics and Dynamics of Galaxies* (Reidel, Dordrecht, 1983), pp. 319–329.
10. For a detailed description of the type classification of galaxies, see A. Sandage, *The Hubble Atlas of Galaxies* (Carnegie Institution of Washington, Washington, DC, 1961).
11. O. J. Eggen, D. Lynden-Bell, A. R. Sandage, *Astrophys. J.* **136**, 748 (1962); A. Sandage, K. C. Freeman, N. R. Stokes, *ibid.* **160**, 831 (1970).
12. Why stars in ellipticals should have formed so much more rapidly than stars in disk galaxies remains an unsolved riddle of this traditional view.
13. A. Toomre, in *The Evolution of Galaxies and Stellar Populations* (Yale University Observatory, New Haven, CT 1977), pp. 401–426.
14. S. Chandrasekhar, *Astrophys. J.* **97**, 255 (1943).
15. D. Lynden-Bell, *Mon. Not. R. Astron. Soc.* **136**, 101 (1967).
16. O. E. Gerhard, *ibid.* **197**, 179 (1981); R. T. Farouki and S. L. Shapiro, *Astrophys. J.* **259**, 103 (1982); K. Borne, *ibid.* **287**, 503 (1984).
17. J. Negroponte and S. D. M. White, *Mon. Not. R. Astron. Soc.* **205**, 1009 (1983).
18. F. Schweizer, *Astrophys. J.* **237**, 303 (1980).
19. R. D. Cannon, in *Optical Jets in Galaxies* (SP-162, European Space Agency, Nordwijk, 1981), pp. 45–47; see also R. F. Haynes, R. D. Cannon, R. D. Ekers, *Proc. Astron. Soc. Australia* **5**, 241 (1983).
20. J. A. Graham, *Astrophys. J.* **232**, 60 (1979); D. F. Malin, P. J. Quinn, J. A. Graham, *Astrophys. J. Lett.* **272**, L5 (1983).
21. A. D. Tubbs, *Astrophys. J.* **241**, 969 (1980); G. F. Simonson, thesis, Yale University (1982).
22. F. Schweizer, *Astrophys. J.* **252**, 455 (1982).
23. D. F. Malin, *Nature (London)* **277**, 279 (1979); ——— and D. Carter, *ibid.* **285**, 643 (1980).
24. D. F. Malin and D. Carter, *Astrophys. J.* **274**, 534 (1983).
25. The technique of unsharp masking aims at diminishing large-scale density variations in an image while preserving small-scale variations. Photographically, this is achieved by superposing onto the original, negative plate of a galaxy a positive, out-of-focus copy and making a print from this sandwich. See, for example, C. E. K. Mees and F. H. James, *The Theory of the Photographic Process* (Macmillan, New York, ed. 3, 1966), p. 495; D. F. Malin, *Photo. Bull. Am. Astron.* **16**, 10 (1977). Unsharp masking can also be achieved digitally with images stored in a computer (*50*).
26. D. Carter, D. A. Allen, D. F. Malin, *Nature (London)* **295**, 126 (1982); D. Carter, in *Proceedings of the IAU Regional European Meeting 8* (Springer, Heidelberg, in press).
27. P. J. Quinn, thesis, Australian National University, Sydney (1982); *Astrophys. J.* **279**, 596 (1984).
28. A. Bosma, R. M. Smith, K. J. Wellington, *Mon. Not. R. Astron. Soc.* **212**, 301 (1985).
29. C. Dupraz, thesis, Université de Paris VII (1984).

30. F. Schweizer, B. C. Whitmore, V. C. Rubin, *Astron. J.* **88**, 909 (1983).
31. P. L. Schechter and J. E. Gunn, *ibid.* **83**, 1360 (1978); P. L. Schechter, M.-H. Ulrich, A. Boksenberg, *Astrophys. J.* **277**, 526 (1984).
32. P. L. Schechter, *Carnegie Inst. Washington Yearb.* **83**, 51 (1984); V. C. Rubin, private communication.
33. J. P. Ostriker and S. P. Tremaine, *Astrophys. J. Lett.* **202**, L113 (1975); M. A. Hausman and J. P. Ostriker, *Astrophys. J.* **224**, 320 (1978).
34. D. P. Schneider, J. E. Gunn, J. G. Hoessel, *Astrophys. J.* **268**, 476 (1983).
35. D. Merritt, *ibid.* **289**, 18 (1985).
36. P. M. Lugger, *ibid.* **286**, 106 (1984); for a dissenting view, see J. Schombert, thesis, Yale University (1984).
37. P. M. Lugger, *Astrophys. J.* **278**, 51 (1984).
38. J. L. Tonry, *ibid.* **279**, 13 (1984); *ibid.* **291**, 45 (1985); J. G. Hoessel, K. D. Borne, D. P. Schneider, *ibid.*, in press.
39. H. Arp, *Astron. Astrophys.* **3**, 418 (1969).
40. B. Balick and T. M. Heckman, *Annu. Rev. Astron. Astrophys.* **20**, 431 (1982); R. D. Joseph *et al.*, *Mon. Not. R. Astron. Soc.* **209**, 111 (1984).
41. G. Neugebauer *et al.*, *Science* **224**, 14 (1984).
42. R. A. E. Fosbury and J. V. Wall, *Mon. Not. R. Astron. Soc.* **189**, 79 (1979).
43. R. D. Joseph *et al.*, *Nature (London)* **311**, 132 (1984); B. T. Soifer *et al.*, *Astrophys. J. Lett.* **283**, L1 (1984); G. H. Rieke *et al.*, *Astrophys. J.* **290**, 116 (1985).
44. C. J. Lonsdale, S. E. Persson, K. Matthews, *Astrophys. J.* **287**, 95 (1984).
45. A. Stockton, *ibid.* **257**, 33 (1982).
46. J. B. Hutchings, *Publ. Astron. Soc. Pacific* **95**, 799 (1983); ———, D. Crampton, B. Campbell, *Astrophys. J.* **280**, 41 (1984).
47. A. Stockton and J. W. MacKenty, *Nature (London)* **305**, 678 (1983).
48. A. Wirth and J. S. Gallagher, *Astrophys. J.* **282**, 85 (1984); J. Kormendy, *ibid.*, in press.
49. J. R. Graham *et al.*, *Nature (London)* **310**, 213 (1984).
50. F. Schweizer and W. Kent Ford, Jr., in *Proceedings of the IAU Regional European Meeting 8* (Springer, Berlin, in press).
51. N. Caldwell, R. P. Kirshner, D. O. Richstone, *Bull. Am. Astron. Soc.* **16**, 455 (1984).
52. I thank H. C. Arp, R. D. Cannon, R. A. E. Fosbury, J. A. Graham, D. F. Malin, V. C. Rubin, A. N. Stockton, and A. Toomre for generously permitting me to reproduce their figures and photographs. The photograph by R. D. Cannon (Fig. 3B, left) was taken with the UK Schmidt telescope in Australia and processed further at the Royal Observatory in Edinburgh, Scotland. Photographs not credited in the captions are by the author and were obtained mostly with the Cerro Tololo Interamerican Observatory 4-m telescope. I thank my collaborator W. Kent Ford, Jr., for his efforts in developing an astronomical picture-processing system at the Department of Terrestrial Magnetism. Supported by NSF grants AST 82-16979 and 83-18845.

Defense Strategies Against Hypoxia and Hypothermia

P. W. Hochachka

Because aerobic metabolic rates decrease in hypoxia-sensitive cells under oxygen-limiting conditions, the demand for glucose or glycogen for anaerobic glycolysis may rise drastically as a means of making up for the energetic shortfall. However, ion and electrical potentials typically cannot be sustained because of energy insufficiency and high membrane permeabilities; therefore metabolic and membrane functions in effect become decoupled. In hypoxia-tolerant animals, these problems are resolved through a number of biochemical and physiological mechanisms; of these (i) metabolic arrest and (ii) stabilized membrane functions are the most effective strategies for extending tolerance to hypoxia. Metabolic arrest is achieved by means of a reversed or negative Pasteur effect (reduced or unchanging glycolytic flux at reduced O_2 availability); and coupling of metabolic and membrane function is achievable, in spite of the lower energy turnover rates, by maintaining membranes of low permeability (probably via reduced densities of ion-specific channels). The possibility of combining metabolic arrest with channel arrest has been recognized as an intervention strategy. To date, the success of this strategy has been minimal, mainly because depression of metabolism through cold is the usual arrest mechanism used, and hypothermia in itself perturbs controlled cell function in most endotherms.

Although some degree of hypoxia and hypothermia can be sustained by all animals, both conditions ultimately are incompatible with survival of most mammalian tissues. Some ectothermic animals, however, are capable of surviving for long periods without O_2, and some mammals (notably hibernators) can tolerate—in fact, can take advantage of—hypothermia. When we investigated the means by which these abilities are achieved at the cellular level, we found that the most serious perturbations of hypoxia and of hypothermia arise from an imbalance between (i) the extent of depression of adenosine triphosphate (ATP) synthesis rates and (ii) the depression of processes requiring membrane-based ATP. When metabolic and membrane functions are decoupled, the cells (tissues or organisms) necessarily become sensitive to hypoxia, to hypothermia, or to both conditions. When, however, the two rate processes are matched despite O_2 limitations or low temperature, then an impressive tolerance to hypoxia, to hypothermia, or to both conditions is achievable. The situation of simultaneous resistance to lack of O_2 and to low temperature arises because some of the fundamental mechanisms used by hypoxia-tolerant animals to protect tissues and organs against hypoxia are the same as, or at least remarkably similar to, those used by cold-tolerant organisms such as hibernators to sustain prolonged hypothermia. In this article, similarities and differences in mechanisms of adaptation to hypoxia and hypothermia are analyzed, with emphasis on opportunities, wherever possible, for novel intervention strategies.

How Animals Survive Oxygen Lack

Careful analysis of animals profoundly resistant to hypoxia indicate two broad adaptive categories. In one category, typified by species adapted to high altitude and by patients suffering from chronic hypoxia (*1*), metabolic mechanisms are directed toward sustained oxidative function despite potentially long-term O_2 limita-

P. W. Hochachka is in the Department of Zoology, University of British Columbia, Vancouver, Canada V6T 2A9.