

No al «primo uso» di armi nucleari

La sicurezza mondiale aumenterebbe se, in risposta a un attacco non nucleare, gli Stati Uniti e i loro alleati adottassero una strategia militare che escluda il ricorso alle armi nucleari

di Kurt Gottfried, Henry W. Kendall e John M. Lee

Dalla fine degli anni quaranta gli Stati Uniti e i loro alleati della North Atlantic Treaty Organization (NATO) hanno fondato la loro comune difesa sulla strategia, apertamente dichiarata, di dare il via a una guerra nucleare se le loro forze militari convenzionali, ossia non nucleari, fossero minacciate di sconfitta in Europa o altrove. La dottrina del «first use» («primo uso») di armi nucleari fu formulata in un momento in cui la NATO si trovava di fronte a due alternative per contrastare il grande esercito permanente schierato dall'Unione Sovietica: o chiamare sotto le armi ed equipaggiare un gran numero di uomini (una mossa ovviamente impopolare dopo sei anni di guerra), oppure fare affidamento su un'arma relativamente poco costosa ed enormemente distruttiva di cui i loro avversari non disponevano. Non si perse molto sonno per operare la scelta o per ripensarci fin tanto che gli Stati Uniti godevano di un notevole margine di vantaggio sull'Unione Sovietica in fatto di armi nucleari.

Ora però l'era dell'indiscussa superiorità nucleare americana si è conclusa da tempo e, con la sua fine, si è erosa la fiducia nella politica della NATO del «primo uso». I dubbi sono sorti in parte perché si è dimostrato impossibile escogitare piani che garantissero di acquisire qualsiasi vantaggio militare con un impiego limitato di armi nucleari. Inoltre non è stato prospettato nessun sistema capace di interrompere, una volta iniziata, una

guerra nucleare. Per molti, di conseguenza, è ormai evidente che anche l'uso più limitato delle armi nucleari potrebbe sfociare benissimo nella catastrofe definitiva di una guerra nucleare totale. Ciononostante, nell'ambito dell'attuale politica del primo uso, le armi nucleari fanno parte integrante delle fasi di addestramento, progettazione e armamento di tutte le operazioni militari della NATO. Nel caso di sconfitta in un conflitto di un certo rilievo le pressioni tendenti a ricorrere a esse sarebbero quasi irresistibili.

Queste preoccupazioni hanno indotto alcuni osservatori di primo piano ad auspicare che la NATO adotti una nuova politica in merito all'avvio di una guerra nucleare. In particolare, nella primavera del 1982 quattro ex alti funzionari del governo americano (McGeorge Bundy, George F. Kennan, Robert S. McNamara e Gerard C. Smith) pubblicarono su «Foreign Affairs» un articolo nel quale si auspicava che gli Stati Uniti e i loro alleati della NATO prendessero in seria considerazione una politica che escludesse il primo uso. Nel medesimo tempo la Union of Concerned Scientists esaminò pubblicamente una serie di misure intese ad aumentare la sicurezza internazionale e a ridurre il rischio di una guerra nucleare. Una delle raccomandazioni che emersero da questo esame era che la NATO dovesse orientarsi verso una politica che escludesse il primo uso di armi nucleari in Europa e che gli Stati Uniti, indipendentemente dai loro alleati europei, facessero

lo stesso in ogni altra parte del mondo.

Questo gruppo di scienziati, di cui fanno parte i tre estensori dell'articolo, patrocinò poi uno studio più circostanziato sull'opzione che escludeva il primo uso con particolare riferimento alla concomitante necessità di rafforzare le forze convenzionali della NATO. A tale studio, diretto da uno di noi (Lee), hanno collaborato (oltre agli altri due di noi) Gerald Steinberg e Peter Trubowitz. Molti funzionari militari di grado più elevato attualmente a riposo e molti ex funzionari della difesa civile di qua e di là dall'Atlantico presero parte al progetto e sottoscrissero la relazione finale. Questo articolo si basa in gran parte sui risultati di quello sforzo collettivo.

Qualsiasi analisi della proposta che esclude il primo uso deve prendere in considerazione innanzitutto la situazione militare sul fronte centroeuropeo, dove i paesi della NATO e del Patto di Varsavia si fronteggiano con le due forze militari più potenti che siano mai state schierate in tempo di pace. In genere si è concordi sul fatto che la NATO farebbe ricorso alle armi nucleari soltanto se un'offensiva condotta dal Patto di Varsavia sembrasse a un passo dal successo. Di conseguenza le perdite in campo militare sarebbero già dell'ordine delle migliaia di unità e il fronte in rapido spostamento si troverebbe ad attraversare a un certo punto le zone densamente popolate della Germania Occidentale. In circostanze del

genere l'impiego di armi nucleari tattiche, o a breve raggio, potrebbe dare un'efficace sostegno alle truppe impegnate sul campo di battaglia soltanto se i comandanti fossero in grado di prendere rapide decisioni sulla base di precise informazioni raccolte dai servizi segreti e se fosse possibile disporre immediatamente delle armi per l'uso. Anche se le operazioni nucleari iniziali della NATO avessero scopi diversi dall'appoggio diretto al campo di battaglia (se cioè, poniamo, scegliessero come obiettivo i centri di comando e controllo del Patto di Varsavia o le sue forze di riserva dislocate nell'estrema retrovia) vi sarebbe la stessa necessità di coordinarle con una situazione in rapido mutamento sulla base delle migliori informazioni disponibili.

A queste operazioni della NATO si opporrebbero ostacoli formidabili. L'intricato sistema sul quale la NATO si basa per il comando, il controllo, le comunicazioni e le informazioni raccolte dai servizi segreti (C³I), condiviso dalle forze convenzionali e nucleari, è estremamente vulnerabile a un attacco e presumibilmente sarebbe uno dei principali obiettivi delle forze del Patto di Varsavia sin dall'inizio delle ostilità. Attualmente gli ufficiali superiori della NATO non possono varcare la «soglia nucleare» di propria iniziativa. Si parte dal presupposto che le massime autorità politiche dei paesi membri dell'alleanza atlantica concordino sulla scelta del momento, sull'ordine di grandezza e sull'ubicazione di qualsiasi attacco nucleare, e per questa ragione la NATO ha creato un'elaborata procedura per raggiungere tali decisioni.

Di conseguenza le esigenze politiche e militari dell'alleanza atlantica sono in diretto contrasto fra di loro. Dal punto di vista militare è essenziale agire con immediatezza finché il sistema C³I è ancora in grado di fornire dati sicuri e prima che le truppe della NATO, il cui addestramento si è basato sulla dipendenza dalle armi nucleari, incomincino a disgregarsi. Dal punto di vista politico la decisione richiederebbe un consenso capace di resistere alle tensioni senza precedenti intrinseche all'avvio e al proseguimento di una guerra nucleare. Un consenso del genere si potrebbe ottenere soltanto con il tempo, mentre la situazione militare potrebbe deteriorarsi fino al punto di rottura se l'autorità politica ritardasse - o rifiutasse di concedere - l'uso di armi nucleari.

I giochi di guerra e altre esercitazioni militari condotti dalle istituzioni americane e inglesi addette alla difesa gettano ulteriori dubbi sull'utilità militare dell'attuale politica del primo uso. Da essi risulta che se la NATO introducesse armi nucleari tattiche nel conflitto, lo scambio nucleare che quasi certamente ne conseguirebbe infliggerebbe più danni alla NATO stessa che al Patto di Varsavia. Il ripiegamento delle forze convenzionali dell'alleanza atlantica verrebbe a costituire un bersaglio ideale per un attacco nucleare: una concentrazione di truppe entro la propria popolazione. Pertanto le forze del Patto di Varsavia potrebbero

permettersi il lusso di essere molto meno discriminatrici nell'uso di armi nucleari di quanto sarebbero costrette a essere le forze della NATO. La struttura politica estremamente centralizzata del Patto di Varsavia, dominato dall'Unione Sovietica, gli permetterebbe di sfruttare rapidamente ogni occasione in uno scontro nucleare tattico qualsiasi.

Naturalmente la distruzione causata a entrambe le parti da uno scambio nucleare di questo genere potrebbe essere immensa. Delle circa 9000 armi nucleari tattiche installate nel teatro delle operazioni europeo, migliaia verrebbero fatte esplodere presumibilmente nella Germania Orientale e nella Germania Occidentale,

esponendo direttamente molte aree urbane al fuoco e all'onda di pressione. Il conseguente fallout radioattivo potrebbe contaminare decine di migliaia di chilometri quadrati da una parte e dall'altra del fronte centrale (si veda l'illustrazione a pagina 16).

Perché una guerra nucleare rimanga limitata, le due parti dovrebbero negoziare limiti assolutamente precisi e verificabili in fatto di armi e di obiettivi. Questa impresa diplomatica senza precedenti dovrebbe essere condotta in uno stato di crisi e d'incertezza che non ha riscontro nella storia e mentre ognuna delle due parti tiene di riserva un arsenale



Su questa cartina sono indicati i paesi membri delle due alleanze militari che si contrappongono in Europa. Le nazioni appartenenti alla NATO sono in grigio. Le nazioni del Patto di Varsavia sono in colore. La Francia è membro della NATO, ma non partecipa alla sua struttura di comando militare. Al di là dell'Atlantico i paesi membri della NATO sono Stati Uniti e Canada.

	NATO	PATTO DI VARSAVIA
POPOLAZIONE	575 MILIONI	375 MILIONI
FORZE ARMATE	4,90 MILIONI	4,76 MILIONI
PRODOTTO NAZIONALE LORDO (DOLLARI)	5,77 BILIONI	2,02 BILIONI
SPESE MILITARI ANNUE (DOLLARI)	260 MILIARDI	127 MILIARDI

Confronto fra i dati aggregati dei paesi della NATO e del Patto di Varsavia. I totali NATO comprendono le cifre relative alla Francia, ma non alla Spagna, entrata nell'alleanza nel 1982.

nucleare che potrebbe portare il conflitto praticamente a qualsiasi livello di distruzione. Sarebbe avventato presumere che in circostanze del genere sia possibile evitare una *escalation* che si estenda molto oltre il campo di battaglia. Non desta quasi sorpresa il fatto che da molto tempo a questa parte molti ufficiali di grado elevato e molti funzionari della difesa civile sostengano che gli esplosivi nucleari non hanno la benché minima utilità come armi da campo.

Il riconoscimento del fatto che la decisione di ricorrere all'impiego delle armi nucleari è troppo grave per essere legata agli imperativi tattici del momento ha indotto alcuni strateghi ad abbracciare un'altra opzione nucleare per evitare una sconfitta convenzionale. In questo caso il primo uso assumerebbe la forma di un attacco nucleare a uno o più bersagli militari, lontani dal fronte, che abbiano un valore sufficientemente elevato «a scopo dimostrativo», ma non tanto elevato da provocare una controrisposta molto mag-

giore. Il fatto che bersagli di questo genere esistano o non esistano è tutto da dimostrare. In ogni caso, qualunque sia l'efficacia di questa strategia, essa dipenderebbe dalla minaccia intrinseca di *escalation*, e quindi essa non offre alcuna possibilità di sfuggire al dilemma presentato dalla strategia del primo uso di armi nucleari sul campo di battaglia.

In breve, in un conflitto fra le due superpotenze non vi è alcuna sceneggiatura plausibile per l'impiego di armi nucleari che non porti con sé il pericolo di una catastrofica *escalation*. Per taluni questa conclusione non fa altro che mettere in evidenza il valore, come deterrente, della dottrina del primo uso. Altri sono arrivati al punto di credere che la NATO non possa giocarsi il futuro su una strategia che è «o un bluff o un patto suicida», per citare il feldmaresciallo Lord Carver, ex capo del British Defence Staff. Questi ultimi sostengono che una politica che escluda il primo uso costituirebbe un fondamento più sicuro per la difesa. Il resto

di questo articolo è dedicato a un esame di questa affermazione.

Una politica che escluda il primo uso relegherebbe le armi nucleari a una sola funzione: quella di rappresaglia in caso di attacco nucleare. Una politica del genere sarebbe molto più di una dichiarazione d'intenti; essa richiederebbe mutamenti sostanziali nell'addestramento, nella pianificazione e nell'armamento delle forze convenzionali della NATO, in modo da eliminare la dipendenza fondamentale dalle armi nucleari per le operazioni militari. Una volta che questa politica venisse adottata, nessun comandante di un'unità tattica potrebbe pensare di valersi, come appoggio, delle armi nucleari, così come oggi non si pensa al primo uso di gas nervino o di armi biologiche.

L'errata quanto diffusa percezione secondo la quale l'Unione Sovietica ha attualmente una schiacciante superiorità in fatto di armi convenzionali ha indotto alcuni critici a scartare come utopistica l'opzione che esclude il primo uso. Così come viene rappresentato di solito, l'equilibrio convenzionale delle forze ignora il fatto che la funzione della NATO è di difendere, non di attaccare. Anche se è vero che in fatto di armamenti convenzionali il Patto di Varsavia è più forte della NATO, va tenuto presente che, in un conflitto fra eserciti di capacità più o meno pari, chi attacca deve disporre di un vantaggio considerevole per essere sicuro del successo. La disparità tra le forze del Patto di Varsavia e quelle della NATO non è di tali proporzioni. Siamo convinti inoltre che, al livello attuale o quasi delle forze, la NATO potrebbe disporre di un deterrente convenzionale ancora più formidabile con l'adozione di un programma di miglioramenti a un costo ridotto rispetto alle spese attuali. Un programma del genere potrebbe essere finanziato in gran parte con la riduzione delle spese in armi nucleari.

Nell'ambito di una politica che escluda il primo uso, le forze convenzionali della NATO dovrebbero scoraggiare - e se necessario bloccare - un attacco convenzionale da parte del Patto di Varsavia senza ricorrere alle armi nucleari. A tal fine si renderebbe necessario modificare la composizione e la struttura delle forze attuali della NATO, cambiare i piani operativi e disporre di una scorta molto più grande di approvvigionamenti.

I servizi d'informazione del Patto di Varsavia sarebbero pienamente consapevoli di questi sviluppi. Pur non essendo da escludere che non si fiderebbero mai al cento per cento di una dichiarazione della NATO che escluda il primo uso, i responsabili del Patto di Varsavia probabilmente non sarebbero alieni dal convincersi che la nuova dottrina sia realmente in vigore. Analogamente, dal controllo, fra gli altri indicatori, delle esercitazioni e delle manovre militari del Patto di Varsavia, i servizi d'informazione della NATO saprebbero se la recente dichiarazione del Patto di Varsavia che esclude il primo uso trovi effettivamente riscontro nell'adde-

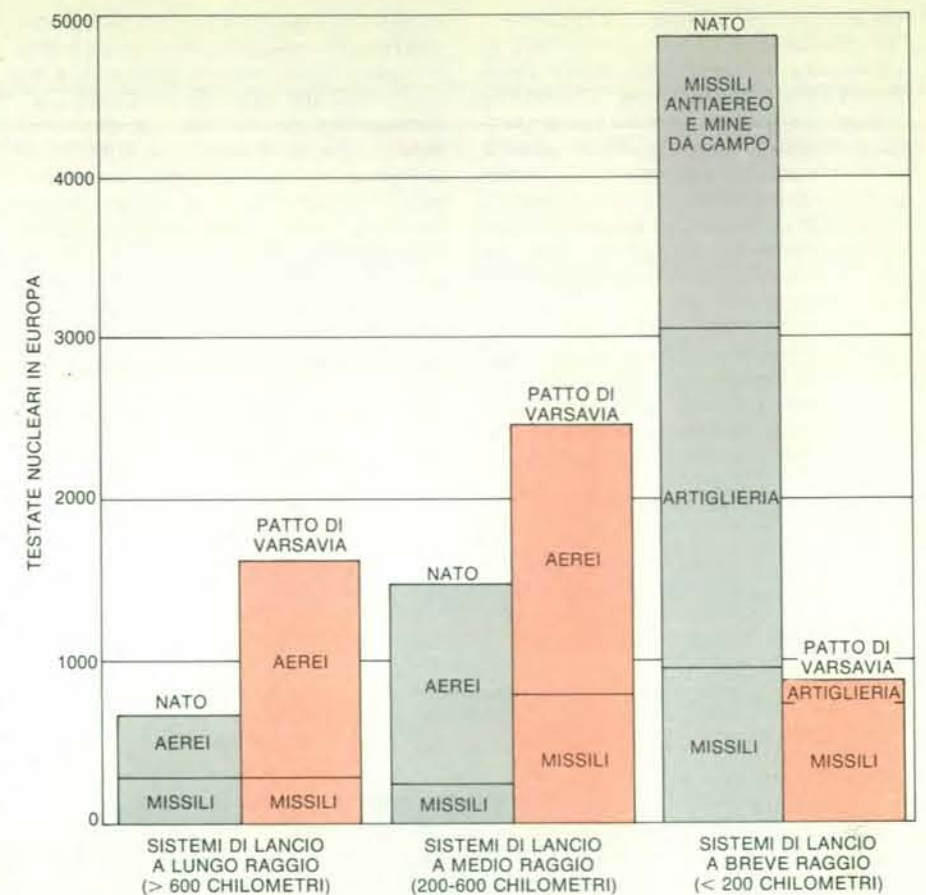
stramento e nella pianificazione militare.

Nell'ambito della politica del primo uso, oggi prevalente, sono molti nell'Europa occidentale coloro i quali hanno paura più dei loro stessi sistemi di difesa che di un potenziale attacco. Questo timore minaccia di paralizzare politica l'intera alleanza e ha creato negli ambienti militari la preoccupazione che la NATO possa avere già *de facto* una politica che escluda il primo uso senza alcuni dei necessari preparativi. Una politica che escluda il primo uso, la quale sia deliberatamente preparata e concordata in tempo di pace, porterebbe in un momento di crisi a decisioni più sicure e più coerenti, e questo fatto aumenterebbe di per sé la deterrenza.

La dichiarazione di una politica che escluda il primo uso dovrebbe essere accompagnata anche da cambiamenti di grande rilievo nell'arsenale nucleare tattico. In realtà è opinione generale, condivisa da molti fautori dell'attuale politica del primo uso, che la grande quantità di armi nucleari tattiche della NATO installate vicino al fronte costituisca una minaccia maggiore per la NATO che non per il Patto di Varsavia. Il problema è che, di fronte a un'offensiva del Patto di Varsavia, vi sarebbero enormi pressioni all'impiego di queste armi piuttosto che lasciarle sovrappiù e catturare. In una politica che escluda il primo uso, la maggior parte di questi sistemi perderebbe la ragion d'essere che ha ora.

È stato detto che una dichiarazione che escluda il primo uso equivarrebbe ad abbandonare l'Europa, e che l'impegno nei confronti dell'attuale politica del primo uso è indispensabile per unire il destino degli Stati Uniti a quello dei loro alleati della NATO. Questa affermazione dà per scontato che esporre immediatamente al combattimento circa 300 000 soldati americani costituisca un legame più debole della promessa non molto plausibile di esporre gli Stati Uniti alla devastazione totale nel caso di un'incursione del Patto di Varsavia nell'Europa occidentale. Una politica che escluda il primo uso richiederebbe un maggiore coinvolgimento degli Stati Uniti in una difesa convenzionale realistica e credibile dei loro alleati europei e dovrebbe quindi far nascere una maggiore fiducia nel «legame» transatlantico. Gli Stati Uniti continuerebbero ancora a offrire ai loro alleati una garanzia nucleare di natura assolutamente completa, ossia la loro capacità di rispondere a un attacco nucleare, che li protegga dalla minaccia di distruzione nucleare.

Ci si può fare un'idea più concreta dell'opzione che esclude il primo uso esaminando innanzitutto parecchie sceneggiature possibili. Si consideri, per cominciare, un'invasione generale di prevenzione dell'Europa occidentale da parte del Patto di Varsavia in una situazione di crisi. L'esistenza stessa di vasti arsenali nucleari dà una misura di «deterrenza estesa» rispetto al primo uso di armi nucleari. Sebbene l'*escalation* nucleare non abbia alcun senso né dal punto di vista militare né da quello politico, nessuna delle due



Le testate nucleari in Europa sono qui classificate secondo i loro sistemi di lancio e il raggio d'azione approssimativo. L'istogramma non comprende le numerose armi nucleari strategiche, o a raggio intercontinentale, installate da entrambe le parti sulla terraferma, su sommergibili e su aerei.

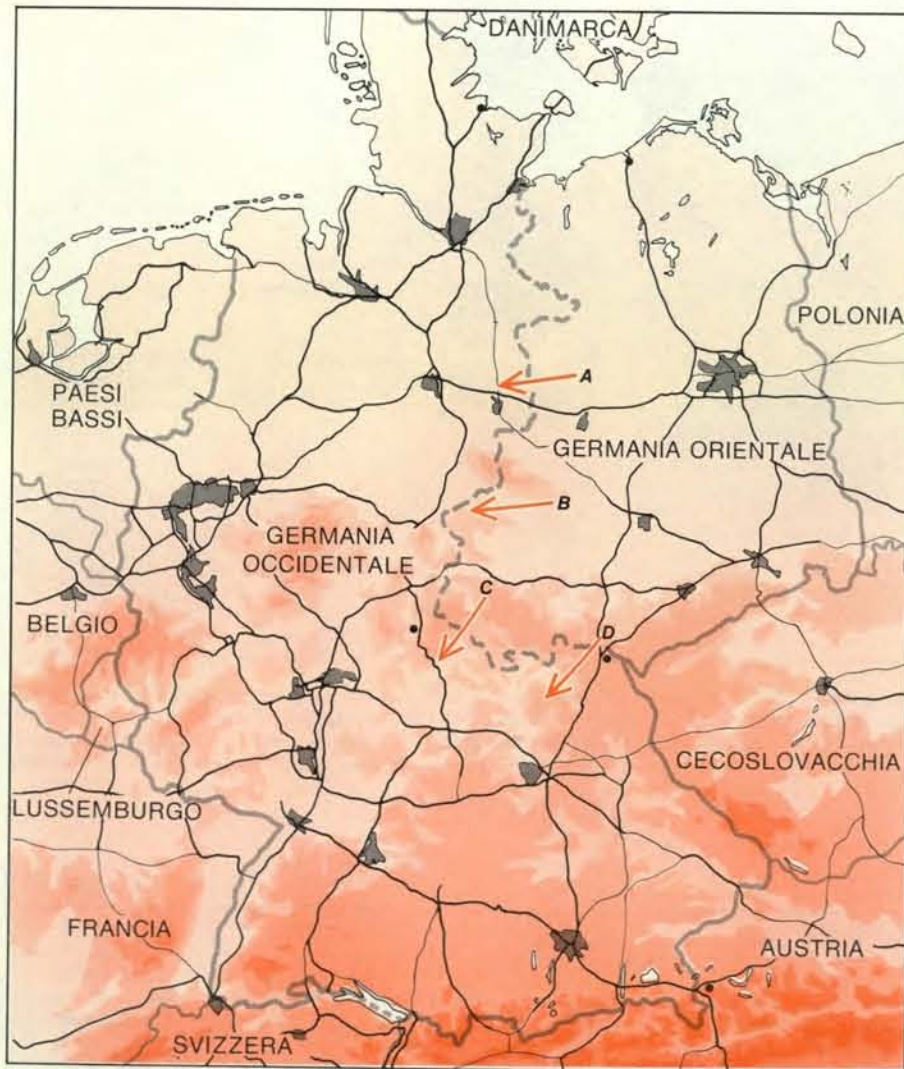
parti potrebbe essere sicura che sarebbe possibile evitare una guerra nucleare su vasta scala, quali che fossero le intenzioni delle parti in causa prima dello scoppio delle ostilità. Un attacco del genere comporterebbe quindi che il Patto di Varsavia accettasse il rischio che alla fine il conflitto diventi nucleare.

In secondo luogo, si consideri un'offensiva limitata dell'Unione Sovietica mirante alla conquista di uno o più obiettivi resi più allettanti dal ritiro del deterrente nucleare nei confronti di un attacco convenzionale. Uno di questi obiettivi potrebbe essere Amburgo, il cui possesso garantirebbe ai sovietici un porto sull'Atlantico di grande valore strategico. Rientrando nel raggio d'azione dell'artiglieria di frontiera, Amburgo potrebbe essere soggetta a una conquista improvvisa anche con un migliorato sistema difensivo della NATO e con un rischio relativamente ridotto di *escalation*, ammesso che la NATO tenesse fede alla propria dichiarazione di esclusione del primo uso. Un attacco di questo genere che lasciasse la NATO fondamentalmente indenne sarebbe un'impresa estremamente rischiosa. Un rischio così grande da parte dei capi del Patto di Varsavia avrebbe senso soltanto se l'Unione Sovietica potesse contare sulla successiva disgregazione di una NATO demoralizzata. Un esito almeno altrettanto probabile

sarebbe un'era di pericolosa tensione, la cessazione di ogni scambio commerciale tra Est e Ovest e la ripresa del conflitto dopo che l'Occidente si fosse riarmato fino ai denti.

Una terza sceneggiatura è costituita da una serie continua di errori di calcolo e di percezioni errate stile 1914 che porti alla guerra contro la volontà e gli interessi dei partecipanti. Questo evento potrebbe verificarsi se la NATO si sentisse costretta a mobilitarsi a causa di un massiccio afflusso di truppe sovietiche per soffocare un diffuso senso di inquietudine nei paesi dell'Europa orientale. Secondo molti funzionari e responsabili politici questa sequenza di eventi è una delle strade che hanno maggiori probabilità di portare a un conflitto in Europa. Una variante della sceneggiatura comporta l'espansione nel teatro europeo di altre ostilità tra Stati Uniti e Unione Sovietica (a incominciare, per esempio, dal Medio Oriente). Nell'uno e nell'altro caso la dichiarazione che escluda il primo uso darebbe il tempo e la relativa tranquillità per permettere l'avvio di negoziati prima che il conflitto raggiunga proporzioni paragonabili a un'invasione totale.

Per riassumere, una strategia che escluda il primo uso avrebbe qualche valore nel caso di un'invasione di prevenzione su larga scala da parte del Patto di Varsavia.



Secondo gli strateghi della NATO il fronte centro-europeo è il punto più probabile di un attacco da parte del Patto di Varsavia. Quattro possibili direttrici di un'invasione sono la pianura della Germania settentrionale (A), il corridoio di Gottinga (B), la gola del Fulda (C) e il corridoio di Hof (D).

via; essa potrebbe aumentare le probabilità di considerate avventure come la conquista di sorpresa di un obiettivo prezioso e vulnerabile della NATO e sarebbe chiaramente benefica in caso di una guerra involontaria o dovuta all'allargarsi di un altro conflitto. Fintanto che le superpotenze disporranno di vasti arsenali di missili a lungo raggio, un massiccio attacco convenzionale - inaspettato, per così dire - sembra essere del tutto improbabile e diventerà sempre più improbabile perché le tecniche di sorveglianza con base nello spazio o su aerei forniranno in definitiva informazioni in tempo reale, notte e giorno e in qualsiasi condizione atmosferica su tutte le zone relative al fronte. Sebbene non si possa escludere a priori,

un attacco limitato a un obiettivo come Amburgo non sarebbe decisivo dal punto di vista militare e per di più sarebbe un azzardo ad alto valore di rischio da qualsiasi altro punto di vista. Un'azione del genere non si conforma al modello di grande cautela che ha sempre caratterizzato la politica estera dell'Unione Sovietica. Le cause di gran lunga più probabili sono gli errori grossolani e le false percezioni. Dato il pericolo sempre presente di una disastrosa escalation, una politica che escluda il primo uso aumenta significativamente la stabilità militare.

Pur essendo al di là dei fini di questo articolo, si può analogamente dimostrare la fondatezza delle ragioni che

spingerebbero ad adottare queste stesse misure in qualsiasi altro posto del mondo. Riducendo il rischio di guerre nucleari locali, tali misure potrebbero rendere meno probabile l'escalation verso una guerra totale e aumenterebbe quindi la sicurezza di tutte le nazioni.

Sulle prime si potrebbe avere l'impressione che sarebbe possibile ottenere tutti i vantaggi di una politica dichiarata che escluda il primo uso tenendo segreta questa politica o, alternativamente, dichiarando una politica che escluda il primo uso all'inizio, una politica cioè che escluda il ricorso alle armi nucleari nelle fasi iniziali di un conflitto. Quest'ultima politica è indistinguibile da quella attuale della NATO e, come abbiamo sostenuto,

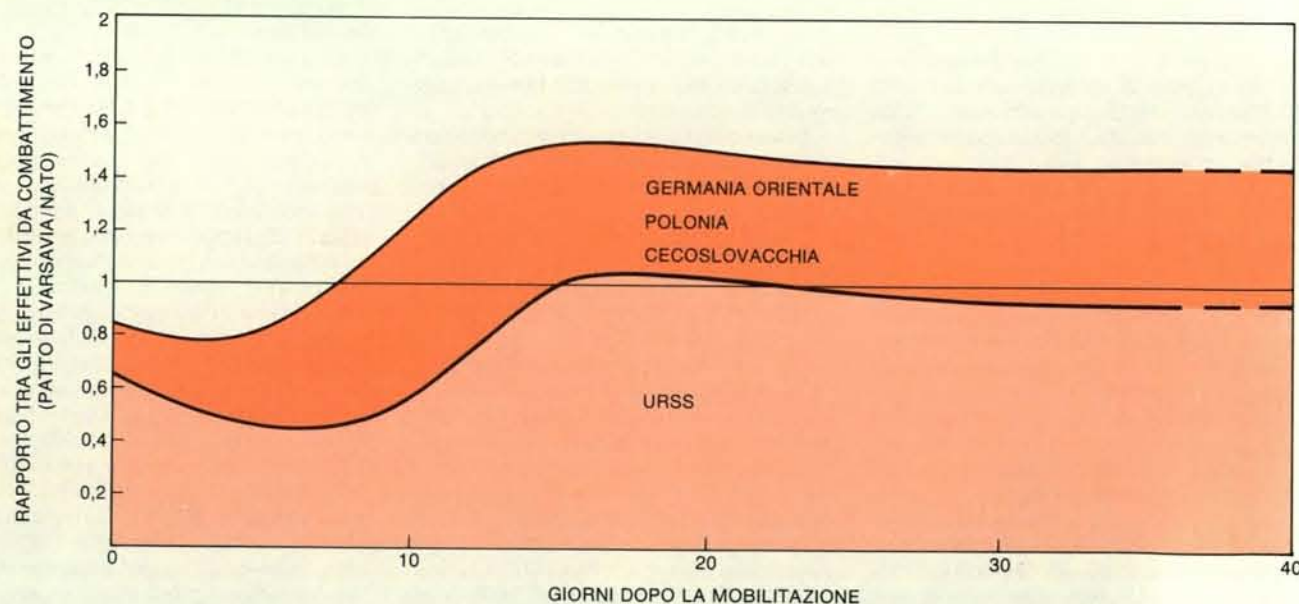
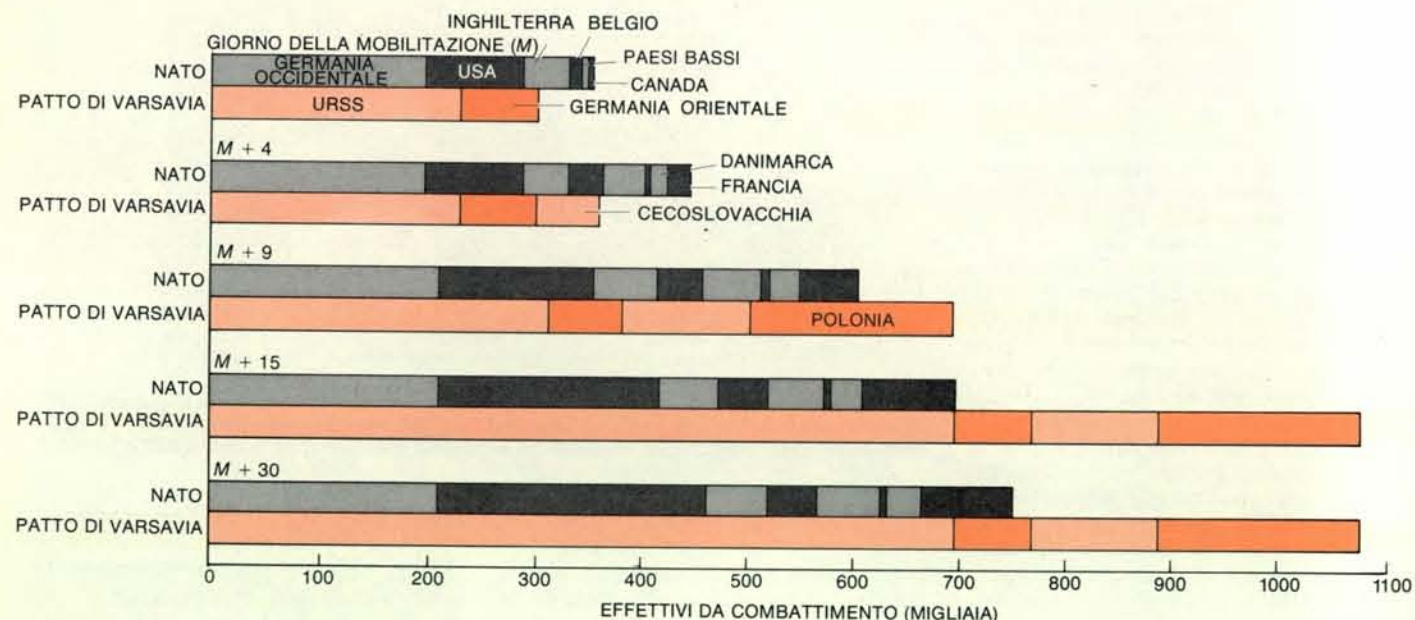
non dà molte promesse di produrre quelle misure militari concrete che libererebbero le truppe dell'alleanza atlantica dalla loro dipendenza dalle armi nucleari. Se si vogliono raccogliere i benefici politici e militari di una politica che escluda il primo uso, è necessario che tale politica venga accettata pubblicamente dalle autorità e dalle popolazioni dei paesi membri della NATO, altrimenti non si arriverà mai sia a realizzare una migliore preparazione a far fronte a un attacco sia ad aumentare il valore deterrente della nuova politica.

Nell'Europa centrale l'Unione Sovietica e i suoi alleati sono numericamente superiori alla NATO tanto dal punto di vista delle truppe quanto da quello degli

armamenti. Le forze del Patto di Varsavia sono in grado di sferrare una formidabile offensiva dalle posizioni che occupano in tempo di pace o attacchi sempre più intensi in fasi successive della mobilitazione e dello spiegamento. Un'offensiva in grande stile del Patto di Varsavia che volesse avere carattere decisivo muoverebbe probabilmente lungo uno o più assi importanti est-ovest; la pianura della Germania settentrionale, il corridoio di Gottinga a sud del massiccio dell'Harz, la gola del Fulda o (meno probabilmente) il corridoio di Hof (si veda l'illustrazione a pagina 12). Questi previsti corridoi di attacco sono definiti dalla topografia e dal sistema stradale della Germania Occidentale e dagli obiettivi più probabili della

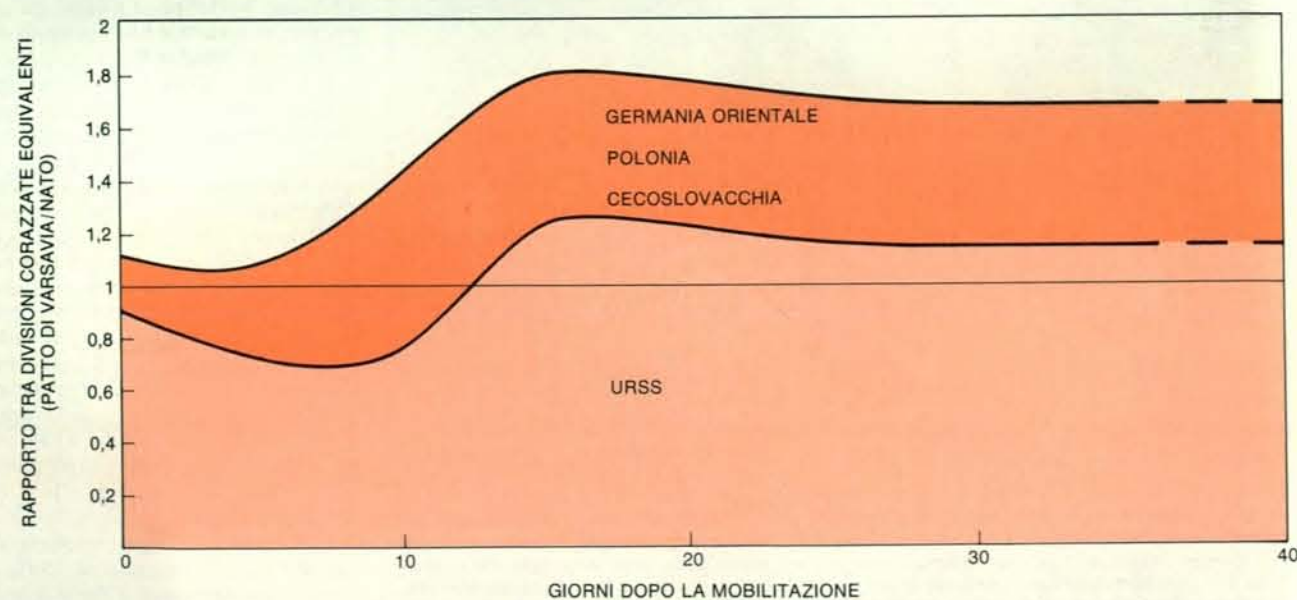
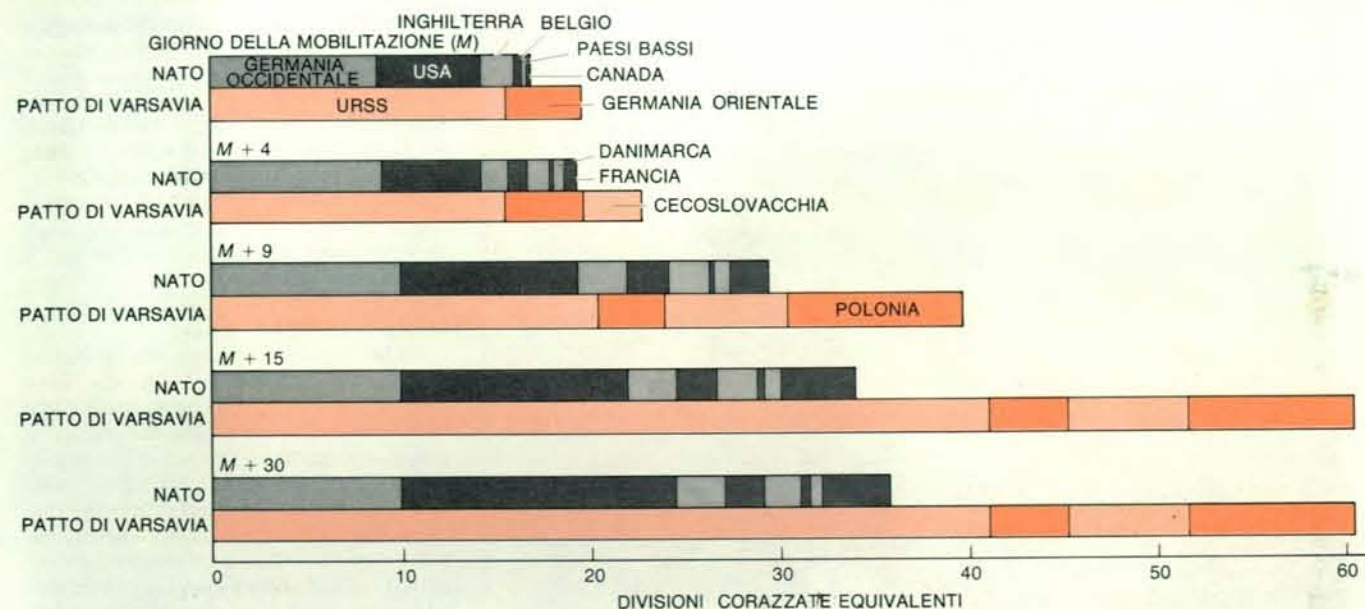
campagna condotta dalle forze del Patto di Varsavia.

L'offensiva sarebbe progettata in modo da raggiungere rapidamente gli obiettivi iniziali, quali la linea del Reno. L'attacco potrebbe assumere la forma o di un guerra lampo o di un'ampia offensiva frontale. In una guerra lampo chi attacca concentra forze corazzate in uno o più punti, estesi forse, inizialmente, da cinque a otto chilometri, e di colpo si spinge in avanti in modo da aprirsi un varco. Quindi le forze attaccanti si riversano nella breccia e penetrano a fondo nelle linee avversarie fino alle retrovie, evitando di dare battaglia, ma distruggendo il sistema delle comunicazioni e impadronendosi di punti chiave vitali per la difesa.



Il rapporto tra le forze di terra pronte al combattimento della NATO e del Patto di Varsavia sul fronte centro-europeo oscillerebbe un poco nei trenta giorni successivi alla mobilitazione, ma in nessun momento si avvicinerebbe al vantaggio critico di due a uno ritenuto necessario per il successo di un attacco. L'equilibrio delle forze è presentato qui con istogrammi (in alto) e con grafici (in basso). La parte sinistra dell'illu-

strazione dà l'equilibrio in termini di effettivi da combattimento; le cifre non sono che approssimazioni, poiché in tempo di pace il numero dei soldati delle unità tattiche di maggior rilievo varia. La parte a destra dà l'equilibrio in termini di «divisioni corazzate equivalenti», una misura standard della forza militare creata dall'US Army che dà risalto più all'efficacia delle armi che al numero degli effettivi. Le differenze fra i



due sistemi di misurazione indicano l'incertezza delle stime. L'URSS ha circa 20 divisioni di non rapido approntamento la cui mobilitazione richiederebbe da tre a quattro mesi. Gli USA potrebbero fare un'analoga aggiunta alle forze della NATO se vi includessero le unità ora tenute da parte per le Rapid Deployment Forces o per la National Guard. Perciò il rapporto di forza che si venisse a creare trenta giorni

dopo la mobilitazione (M + 30) persisterebbe per altri due mesi (fino a M + 90), dopo di che potrebbe diventare poco più o poco meno favorevole alla NATO. Le forze del Patto di Varsavia comprenderebbero numerosi soldati della Germania Orientale, della Cecoslovacchia e della Polonia. L'incerta affidabilità di queste truppe, non riflessa in queste valutazioni numeriche, potrebbe costituire un problema per l'URSS.

Un esteso attacco frontale sfrutta la potenza di fuoco (per esempio artiglieria concentrata) per logorare e sconfiggere le forze avversarie. In un caso o nell'altro, le forze del Patto di Varsavia applicherebbero presumibilmente il loro sistema di «attacco a scaglioni»; il che significa che, via via che un'unità di primo piano si logora, essa passa nelle retrovie per ricomporre le file e viene sostituita nella sua totalità da un reparto fresco.

La NATO farebbe fronte a questi attacchi costringendo il nemico ad attraversare una fascia di territorio occupata da piccole unità anticarro che si sostengono a vicenda e che sono capaci sia di mantene-

re le proprie posizioni su un terreno difficile, sia di ostacolare le forze nemiche nei più importanti corridoi di attacco. L'attacco qui sarebbe affrontato dalle principali forze corazzate e meccanizzate della NATO, le quali approfitterebbero sia di posizioni predisposte per l'attacco, sia della propria intrinseca mobilità.

Molti fattori di natura militare e geopolitica hanno indotto l'alleanza atlantica ad adottare una politica di difesa avanzata. Tanto per cominciare, il lato orientale della Germania Occidentale è, ove si eccettui il Reno, il migliore terreno difensivo. In secondo luogo, le linee laterali di comunicazione sono ottime vicino al lato

occidentale, ma scarse su quello orientale. Infine, non vi è molto terreno da cedere senza incorrere in perdite di importanza fondamentale.

Le forze che la NATO e il Patto di Varsavia sono potenzialmente in grado di mettere in azione in questi scontri si equivalgono in realtà sul piano numerico, dato che la NATO sarebbe sulla difensiva. L'Unione Sovietica e i suoi alleati sono superiori in fatto di numero di uomini nelle unità di combattimento, di numero di divisioni (anche se le divisioni del Patto di Varsavia sono un po' più piccole di quelle della NATO) e di «divisioni corazzate equivalenti», una misura della forza di un'unità tattica basata sul numero e sulle prestazioni delle armi di cui è dotata. Questo vantaggio non solo è valido per gli schieramenti in tempo di pace, ma conserverebbe la sua validità anche durante e dopo una mobilitazione. Ciononostante, le forze che si difendono hanno tradizionalmente un vantaggio su quelle che attaccano. Data una sollecita mobilitazione da parte della NATO, il rapporto delle forze nella zona di combattimento rimarrebbe in ogni fase al di sotto del valore di due a uno generalmente accettato dagli strateghi militari come necessario per il successo di un attacco portato a un difensore abile attestato in posizioni pronte a farvi fronte.

In termini del rapporto fra le truppe pronte all'azione la NATO ha una lieve superiorità in tempo di pace, ma il Patto di Varsavia, con i vantaggi dell'iniziativa e degli spiegamenti più ridotti, si porterebbe in testa e raggiungerebbe un valore massimo al di sotto del rapporto critico di due a uno un paio di settimane dopo l'inizio della mobilitazione. In seguito i rinforzi della NATO prevarrebbero per parecchi mesi fino all'arrivo dall'Unione Sovietica di alcune divisioni di riserva il cui approntamento richiede tempi piuttosto lunghi.

Il Patto di Varsavia concentrerebbe le proprie forze lungo i principali assi di attacco, mirando a una schiacciante superiorità nei punti di sfondamento. Gli effettivi di queste forze sono in grado di raggiungere in tali punti rapporti prossimi a 10 a uno circa due settimane dopo la mobilitazione, se la NATO non provvedesse a rinforzare le divisioni sottoposte all'attacco. Attingendo alle riserve disponibili o alle divisioni adiacenti, la NATO può peraltro mantenere il rapporto di tre o quattro a uno dopo due settimane e a livelli inferiori in seguito; così il vantaggio di chi attacca rimarrebbe al di sotto di quel valore di quattro o cinque a uno ritenuto necessario localmente per operare uno sfondamento.

Conclusioni analoghe risultano da una valutazione dell'equilibrio delle forze aeree. Oggi come oggi il Patto di Varsavia supera numericamente la NATO in fatto di aerei da combattimento con un margine che rimarrebbe più o meno costante dopo la mobilitazione. Gli aerei della NATO però hanno tipicamente una maggiore autonomia di volo e un carico di

bombe superiore, e molte cose stanno a indicare che i piloti americani e quelli dell'Europa occidentale siano meglio addestrati. Inoltre quasi tutti gli aerei del Patto di Varsavia sono progettati per un unico scopo, mentre molti aerei dell'alleanza atlantica sono polivalenti e possono essere impiegati per gli scontri aria-aria, per bombardamenti e per il lancio di missili aria-terra.

Nel valutare l'equilibrio generale delle forze bisogna considerare anche parecchi fattori politici di natura incerta. C'è da chiedersi per esempio se nel caso di un attacco i paesi membri della NATO sarebbero o non sarebbero in grado di agire rapidamente e all'unisono. Un'altra incertezza riguarda il ruolo della Francia, la quale, pur essendo formalmente un paese membro dell'alleanza, non partecipa direttamente alla struttura di comando militare dell'organizzazione. La Francia mantiene nella Germania Occidentale tre divisioni attive, ma non si sa con quale rapidità queste forze potrebbero essere integrate nel comando della NATO. Si ritiene comunque molto improbabile che la Francia rimanga neutrale nel caso di un attacco portato dal Patto di Varsavia.

La problematica affidabilità delle forze non sovietiche del Patto di Varsavia è per l'Unione Sovietica un problema grave che non può essere valutato con nessun criterio di entità numerica. I soldati polacchi e cecoslovacchi parteciperebbero al cento per cento a un combattimento non collegato direttamente con la propria difesa? Potrebbero essere impiegati soldati della Germania Orientale per un'invasione della Germania Occidentale? In realtà è possibile che in una guerra contro la NATO l'Unione Sovietica si veda costretta a valersi delle proprie truppe per assicurarsi la lealtà dei propri alleati. L'Unione Sovietica inoltre dovrebbe prospettare la possibilità che la Cina approfitti delle ostilità in Europa, il che limiterebbe la capacità sovietica di ricorrere alle proprie riserve strategiche per il teatro europeo.

Tutti i rapporti di forza sopra accennati e la loro importanza militare dipendono da molti presupposti: la rapidità e l'efficacia dello spiegamento delle forze degli uni e degli altri, la celerità della NATO nel prendere decisioni, l'estrema sollecitudine delle unità del Patto di Varsavia a rispondere alla mobilitazione, l'elevata capacità di raccogliere informazioni segrete e così via. Né questi né altri fattori che si potrebbero citare dimostrano quale sarebbe l'esito di un conflitto su vasta scala. Essi stabiliscono però che, lungi dall'avere una superiorità schiacciante, le forze del Patto di Varsavia dovrebbero essere considerate appena appena adatte al massimo per portare un'offensiva nell'Europa centrale.

Nelle difese convenzionali dell'alleanza atlantica vi sono molti punti deboli e molte carenze a cui sarebbe bene ovviare se si vuole che la NATO affronti con fiducia il Patto di Varsavia senza l'opzione del primo uso di armi nucleari. Le misure da prendere rientrano in quattro categorie. In primo luogo, la NATO deve essere in

	NATO	PATTO DI VARSAVIA	RAPPORTO
CARRI ARMATI PESANTI	17 629	46 500	1 : 2,64
ARTIGLIERIA E LANCIARAZZI	9829	20 300	1 : 2,07
MISSILI GUIDATI TERRA-ARIA	1662	6293	1 : 3,79
MISSILI GUIDATI ANTICARRO	4644	1822	2,55 : 1
ELICOTTERI ARMATI	795	1406	1 : 1,77
CANNONI ANTIAEREO	5207	6486	1 : 1,24
CANNONI ANTICARRO	996	3724	1 : 3,74

Viene qui dato l'elenco dell'equipaggiamento militare a disposizione di entrambe le parti per un conflitto convenzionale, ossia non nucleare, di un certo rilievo in Europa. Il numero relativamente ridotto di quasi tutti i tipi di armi non nucleari della NATO rispecchia la sua dipendenza da una politica di «first use» («primo uso») per prevenire una sconfitta in una guerra convenzionale.

	NATO	PATTO DI VARSAVIA	RAPPORTO
AEREI TATTICI (GIORNO M)	2100	2700	1 : 1,2
AEREI TATTICI (M + 15)	3700	4900	1 : 1,3

Attualmente le forze aeree tattiche a disposizione della NATO e del Patto di Varsavia si equivalgono e si pensa che la situazione rimarrà inalterata anche dopo la mobilitazione. Il vantaggio numerico delle forze del Patto di Varsavia è controbilanciato a quanto pare dal miglior grado di addestramento dei piloti e dalla maggiore versatilità degli aerei appartenenti alle forze della NATO.

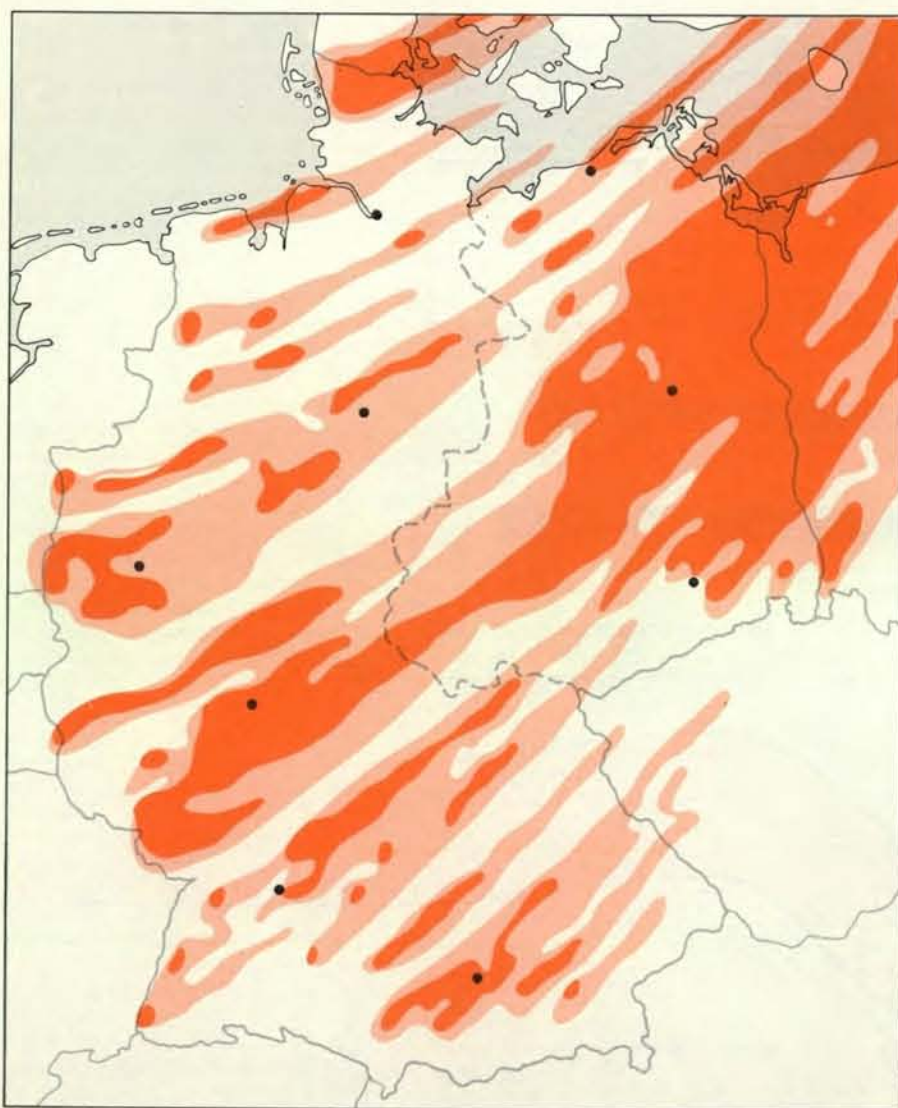
grado di prendere decisioni politiche più rapide e di mettere in campo sollecitamente le forze e le unità di rincalzo provenienti dall'Europa e da oltremare. In secondo luogo, si dovrebbero apportare molti miglioramenti operativi, fra cui la costruzione di ostacoli per i carri armati e di fortificazioni da campo, e si dovrebbero organizzare in modo migliore le forze di riserva per far fronte alla minaccia di una guerra che si prolunghi nel tempo. In terzo luogo, la NATO dovrebbe accrescere il proprio attuale vantaggio in fatto di tecnologia avanzata. Infine, bisognerebbe rafforzare la capacità della NATO di sostenere operazioni militari.

Tradizionalmente l'alleanza atlantica ha preso le proprie decisioni all'unanimità. Di conseguenza la decisione di passare a uno stato elevato di preparazione, di schierare le forze in Europa e di mobilitare rinforzi potrebbe non essere tempestiva. L'indecisione su questi punti è una debolezza fondamentale a cui in un momento di crisi non è possibile ovviare con l'improvvisazione. La NATO dovrebbe migliorare la propria capacità di rispondere a un attacco da parte del Patto di Varsavia dopo un breve periodo di mobilitazione, anche se questa non è una sceneggiatura probabile per lo scoppio di una guerra. Se la NATO scopre i preparativi del Patto di Varsavia e si mobilita nel giro di tre o quattro giorni, l'equilibrio delle forze dovrebbe garantire il successo della difesa. Se però la decisione della NATO viene rinviata e la sua mobilitazione risulta in ritardo di una settimana o più rispetto a quella del Patto di Varsavia, la superiorità di quest'ultimo in fatto di forze schierate supera il livello di pericolo e rimane elevata.

Anche la capacità dell'alleanza atlantica di rispondere prontamente e con effi-

cacia verrebbe migliorata in maniera significativa se la Francia chiarisse le sue intenzioni nel caso di una mobilitazione o di un attacco da parte del Patto di Varsavia. Per fare un esempio, la Francia potrebbe prendere misure concrete che diano al comando della NATO maggiori garanzie sulla tempestiva disponibilità delle attrezzature francesi e potrebbe associarsi a una decisione pubblicamente manifesta di creare nel proprio paese una linea secondaria di comunicazione in appoggio alla linea primaria della NATO.

La NATO dovrebbe anche migliorare la propria capacità di spostare immediatamente le forze in Europa dalle località loro assegnate in tempo di pace al fronte principale. Le guarnigioni delle forze della NATO in Europa in tempo di pace non sono dislocate in modo adatto a un rapido spiegamento. Una soluzione pratica si baserebbe sulla fanteria leggera, che potrebbe essere schierata rapidamente lungo il fronte principale. Dotate di moderne armi anticarro, queste forze potrebbero essere messe di guarnigione nelle vicinanze o all'interno dei settori loro assegnati, e potrebbero essere integrate rapidamente da riserve locali che conoscono bene la zona. Sfruttando il valore difensivo della località, queste truppe ritarderebbero, logorerebbero e incanalerebbero l'attacco, mascherando al tempo stesso il contrattacco delle divisioni della NATO dotate di armamento più pesante. Inequivocabilmente difensiva, la fanteria leggera presenterebbe l'ulteriore vantaggio di poter essere portata, in un periodo di crisi, al punto di intervenire senza esacerbare una situazione politica già tesa. Unità di fanteria leggera esistono attualmente in molti eserciti alleati e molte altre potrebbero essere formate con uomini di riserva che sia possibile mobilitare con una certa sollecitudine.



Il fenomeno di fallout radioattivo dovuto a un'ipotetica guerra nucleare «limitata» che non vada al di là di alcuni obiettivi della Germania Orientale e di quella Occidentale è rappresentato qui tenendo conto delle condizioni di vento e atmosferiche prevalenti in una tipica giornata di giugno nell'Europa centrale. Lo schema si basa su una sceneggiatura nella quale si ipotizza che ognuna delle due parti porti un attacco nucleare di prevenzione contro obiettivi militari (basi missilistiche, basi aeree, depositi di armi nucleari e simili) della parte avversa. Gli attacchi producono un totale di 171 esplosioni a livello del suolo di ordigni nucleari da 200 chiloton. Le aree in colore corrispondono a dosi assorbite di radiazioni misurate in rad (forma abbreviata di radiation absorbed dose); le aree in colore chiaro corrispondono a dosi variabili da 200 a 600 rad, quelle in colore intenso a più di 600 rad. Le morti da effetti di radiazioni incomincerebbero a verificarsi a dosi inferiori a 200 rad. L'illustrazione si basa su uno studio di William Arkin, Frank von Hippel e Barbara G. Levi uscito su «Ambio», una pubblicazione dell'Accademia reale svedese delle scienze.

	COSTO (MILIARDI DI DOLLARI)
OSTACOLI ANTICARRO, FORTIFICAZIONI DA CAMPO, RIFUGI NELLE BASI AEREE	1
AUMENTO DI SCORTE (PER 29 DIVISIONI)	40
MIGLIORAMENTO DEL PROGRAMMA POMCUS	3
MIGLIORAMENTO DELLA CAPACITÀ USA DI PONTI AEREI	4
ALTRE CINQUE DIVISIONI USA DI RISERVA	28
ALTRE TRE DIVISIONI ALLEATE DI RISERVA	15
TOTALE	97

Il costo stimato del miglioramento della capacità delle forze convenzionali della NATO di resistere a invasioni dell'Europa occidentale senza ricorrere alle armi nucleari assommerebbe a circa 100 miliardi di dollari in sei anni. Nell'ambito del «POMCUS» (Pre-Positioned Material Configured to Unit Sets) l'armamento di molte divisioni e unità di appoggio di stanza negli Stati Uniti viene preallestito in Europa in modo che gli effettivi vi possano essere portati in volo direttamente.

Anche la capacità di intervenire rapidamente potrebbe essere migliorata in misura notevole completando il programma americano POMCUS. Il POMCUS (Pre-Positioned Material Configured to Unit Sets) tiene pronti nell'Europa centrale gli armamenti per molte divisioni e unità d'appoggio di stanza negli Stati Uniti in modo che gli effettivi vi possano essere portati in volo direttamente in breve tempo. Per il momento l'obiettivo è limitato a sei gruppi di divisioni (più le unità di appoggio) in Europa. Quattro di questi gruppi sono in servizio attivo, mentre gli altri due devono ancora essere autorizzati. L'intero programma fornirebbe una riserva mobile immediata e potente in ogni settore di gruppi di armate e la cosa migliorerebbe in misura notevole l'equilibrio delle forze convenzionali nel momento più critico del processo di concentrazione, ossia due settimane dopo l'inizio della mobilitazione da parte del Patto di Varsavia.

È degno di nota il fatto che la NATO non abbia fortificazioni su nessuno dei suoi fronti. Ciò è contrario a ogni consuetudine militare precedente e non è giustificato dalla moderna tecnologia o dalle esperienze della recente storia militare. In realtà le fortificazioni svolgono un ruolo di primo piano nella prassi militare sovietica attuale. Eppure, per ragioni di natura politica, nella Germania Occidentale non sono stati costruiti né ostacoli né fortificazioni da campo.

Un sistema di ostacoli anticarro unito a fortificazioni da campo sarebbe di estrema importanza. Ostacoli come fossati, campi minati e strutture di calcestruzzo sono relativamente poco costosi e facili da approntare. Le fortificazioni fisse comprendono postazioni attrezzate per i carri armati e per l'artiglieria, posti di osservazione per controllare il tiro dell'artiglieria a lungo raggio, concentrazioni di cannoni anticarro, munizioni guidate con precisione e telecontrollate e tanti piccoli bunker di calcestruzzo destinati a fornire riparo e sostegno alla fanteria leggera. L'efficacia attuale del fuoco di artiglieria, che si basa su tecniche avanzate di rilevamento e controllo, ha reso obsolete le fortificazioni permanenti di grandi dimensioni. Ciononostante, potrebbero ancora essere

efficaci installazioni piccole, corazzate, che si possono mimetizzare. In alcune circostanze è possibile creare gli ostacoli, necessari, ricorrendo a campi di mine o a candelotti esplosivi sepolti.

L'alleanza atlantica deve continuare anche i propri sforzi per far fronte alla minaccia di un conflitto prolungato. A questo scopo non sono necessari costosi aumenti degli effettivi militari o un ritorno alla coscrizione in paesi che attualmente non l'hanno; basterebbe aumentare il numero delle divisioni di riserva americane e alleate. Le forze di riserva possono costituire un potente complemento alle forze in servizio attivo a un costo unitario pari più o meno alla metà.

Il vantaggio di cui gode attualmente la NATO sul Patto di Varsavia in molti settori cruciali della tecnologia militare dovrebbe essere conservato e sfruttato. La guerra arabo-israeliana del 1973 e i combattimenti più recenti tra le forze siriane e israeliane in Libano dimostrano l'importanza della tecnologia militare innovativa. Le «armi intelligenti» sono un esempio straordinario del potenziale della tecnologia militare. Un'arma intelligente è un'arma intesa a colpire e a distruggere il proprio bersaglio con una probabilità del 50 per cento in condizioni di battaglia. L'efficacia delle armi intelligenti è stata dimostrata nelle operazioni delle isole Falkland, nel corso delle quali entrambe le parti inflissero gravi danni al rispettivo avversario per mezzo di munizioni guidate con precisione lanciate dall'aria (si veda l'articolo *Le armi «intelligenti» e il futuro della guerra navale* di Paul F. Walker, in «Le Scienze», n. 179, luglio 1983).

Le armi intelligenti attualmente in fase di sviluppo porteranno testate multiple in grado di ricercare e scegliere il bersaglio. Un'unica salva di questo genere distruggerebbe carri armati, pezzi di artiglieria e veicoli da combattimento della fanteria. Unite agli ostacoli anticarro e alle fortificazioni, le armi intelligenti possono costituire un potente antidoto contro i carri armati. Il raggio d'azione e la flessibilità di queste armi e dei loro sistemi di controllo sono tali che il loro fuoco potrebbe essere usato in qualunque momento fosse necessario in qualsiasi punto del fronte in cui fosse possibile disporre di informazioni tempestive.

Le armi intelligenti sarebbero preziose anche nell'attaccare bersagli situati molto all'interno del territorio nemico, dove il bombardamento aereo è limitato dall'alta concentrazione e dall'efficacia dei mezzi difensivi aerei e terrestri. Missili da crociera lanciati da terra, dall'aria o dal mare e dotati di testate esplosive convenzionali, potrebbero provocare gravi danni a obiettivi situati molto all'interno con una minore esposizione e con minori perdite di quanto si possa fare con gli aerei guidati da un pilota.

Ovviamente sarebbe un errore dare eccessivo rilievo alle soluzioni tecnologiche o fare affidamento esclusivamente su uno qualsiasi di questi sistemi. Nel caso delle munizioni guidate con precisione, per esempio, la loro efficacia generale è sensibile allo sviluppo di contromisure. È importante tuttavia riconoscere che la tecnologia avanzata continuerà a svolgere un ruolo di primo piano nei sistemi di difesa.

Ostilità su vasta scala tra la NATO e il Patto di Varsavia consumerebbero gli approvvigionamenti bellici a un tasso senza precedenti. La fornitura giornaliera di combustibile, pezzi di ricambio, rimpiazzi e munizioni è molto più grande e costosa di quanto sia mai stata in passato. Gli investimenti nelle scorte hanno avuto finora una bassa priorità per via del presupposto secondo il quale un conflitto su vasta scala tra la NATO e il Patto di Varsavia si intensificherebbe rapidamente fino a sfociare in una guerra nucleare; di conseguenza le scorte di materiale bellico nell'area dell'Europa centrale sono piuttosto ridotte. Nelle simulazioni di guerra alcune unità della NATO esauriscono le loro scorte in appena due settimane di operazioni intense. Nel caso di una politica che escluda il primo uso sarebbe necessario aumentare queste scorte provvisoriamente per una durata di 30 giorni e, nel lungo termine, per non meno di 45 giorni. Come primo passo, la NATO dovrebbe aumentare le scorte delle proprie divisioni permanenti in modo da coprire altre due settimane di combattimenti.

Molti provvedimenti relativamente poco costosi possono essere presi per aumentare la capacità di sopravvivenza delle basi aeree della NATO, i centri di comando e i depositi di approvvigionamento, che potrebbero essere soggetti a pesanti attacchi aerei da parte del Patto di Varsavia. La costruzione di rifugi antiaerei nelle basi aeree alleate dell'Europa centrale ne è un esempio cruciale. Non si è provveduto a questi impianti per i numerosissimi rinforzi aeronautici che arriverebbero nei primi giorni di concentrazione delle forze. La mancata costruzione di queste opere difensive estremamente efficaci e relativamente poco costose per i suddetti rinforzi, vulnerabili, costosi e vitali, potrebbe essere la più notevole delle false economie del programma di difesa della NATO.

Per i principali miglioramenti suggeriti nelle righe precedenti sono state fatte stime relative ai costi, i quali, tutto con-

siderato, ammonterebbero a meno di 100 miliardi di dollari per un periodo di sei anni. Non è stato calcolato da parte nostra il costo di certe misure supplementari, come la creazione di una forza di riserva di fanteria leggera lungo il fronte centrale. È ragionevolmente chiaro per altro che queste misure aggiuntive non farebbero superare al costo globale dei miglioramenti richiesti da una politica che escluda il primo uso i livelli di spesa ritenuti politicamente ed economicamente convenienti.

I 100 miliardi di dollari da noi previsti per i miglioramenti a più alta priorità comporterebbero un aumento annuo del 2 per cento circa in termini reali delle spese annue totali della NATO per la difesa, che sono dell'ordine di 300 miliardi di dollari. Questo aumento sarebbe al di sotto dell'attuale livello del 3 per cento concordato dai membri della NATO nel 1978 e notevolmente al di sotto del 4 per cento richiesto di recente per il periodo 1983-1988 dal generale Bernard W. Rogers, comandante supremo alleato in Europa.

I fondi per mettere in atto questi miglioramenti convenzionali potrebbero venire da una ristrutturazione delle priorità militari della NATO. In particolare i fondi stanziati per l'acquisizione di armi nucleari potrebbero essere trasferiti in modo da coprire il costo di molti dei programmi convenzionali di più alta priorità. Per fare un solo esempio, l'annullamento del programma del missile MX potrebbe portare a un risparmio di oltre 20 miliardi di dollari.

La nostra tesi è che l'adozione di una politica che escluda il primo uso andrebbe a vantaggio militare e politico della NATO, indipendentemente dalle azioni dell'Unione Sovietica e dei suoi alleati, e che pertanto questa decisione non è condizionata da trattative. Esistono tuttavia altre due questioni critiche che di diritto sono componenti fondamentali del confronto europeo e pertanto possono essere affrontate soltanto al tavolo delle trattative. Se fosse possibile risolverle, esse aumenterebbero in misura significativa la sicurezza europea e ridurrebbero gli ostacoli di natura politica che si frapporterebbero a una politica che escluda il primo uso. Ci riferiamo qui ai negoziati, vecchi di dieci anni, per la riduzione reciproca e bilanciata delle forze (MBFR, da Mutual and Balanced Force Reduction), i quali cercano di creare in Europa un equilibrio convenzionale stabile, e alle proposte per una zona denuclearizzata a cavallo del confine Est-Ovest.

Così come si pensa attualmente, la caratteristica principale di un accordo MBFR sarebbe la riduzione degli effettivi delle forze di terra in servizio attivo nell'Europa centrale, per le quali verrebbe fissato, per ognuna delle due parti, un tetto di 700 000 uomini, che diventerebbero 900 000 sommando le forze di terra e dell'aria. L'accordo proibirebbe anche l'introduzione di altre unità militari nelle regioni interessate, in modo da evitare

che si raggiungano i livelli di forza necessari per il successo di un attacco. Le proposte per verificare la conformità alle norme dell'accordo e per controllare le deviazioni dai modelli normali dell'attività militare sono oggetto di trattative.

Un trattato MBFR ridurrebbe i timori che il rafforzamento delle forze convenzionali della NATO non possa fare altro che dare il via a una corsa agli armamenti convenzionali. Esso garantirebbe anche in misura notevole che nessuna delle due parti si stia preparando per un conflitto. Un'eventuale violazione di un trattato del genere sarebbe per le autorità politiche della NATO una chiara dimostrazione di intenzioni ostili e giustificerebbe quindi l'intensificazione degli sforzi difensivi, ivi compresa un'immediata mobilitazione.

La proposta di una zona denuclearizzata in Europa che ha riscosso il più vasto consenso internazionale ai livelli più autorevoli è quella avanzata dalla Independent Commission on Disarmament and Security Issues, presieduta da Olaf Palme, attuale primo ministro svedese. La commissione Palme pubblicò nel giugno 1982 una relazione nella quale si chiedeva l'istituzione da una parte e dall'altra del fronte centrale di una zona denuclearizzata, che inizialmente sarebbe arrivata quasi al centro della Germania Orientale e della Germania Occidentale e che a poco a poco si sarebbe estesa fino a includere l'intero fronte europeo fra le due alleanze (si veda *Una zona denuclearizzata in Europa* di Barry M. Blechman e Mark R. Moore, in «Le Scienze», n. 178, giugno 1983). Tenuto conto dei nostri precedenti commenti sulla minaccia presentata dai sistemi nucleari con base avanzata, è chiaro che la proposta della commissione Palme aumenterebbe la sicurezza della NATO e costituirebbe il complemento naturale di una politica che escluda il primo uso.

Il requisito fondamentale per il completamento dei negoziati MBFR è l'appoggio continuo dei massimi funzionari governativi di entrambe le parti. Una forte guida politica sia sul fronte interno sia su quello internazionale sarebbe necessaria per trasformare in realtà una dichiarazione che escluda il primo uso, un trattato MBFR e le raccomandazioni della commissione Palme. Se fosse possibile raggiungerla, una meta del genere costituirebbe una notevole diminuzione del rischio di guerra in Europa.

L'adozione di una politica che escluda il primo uso avrebbe profonde ripercussioni. A differenza delle misure relative al controllo degli armamenti, che raramente limitano l'uso effettivo delle armi, quella politica trasformerebbe il fondamento concettuale sul quale poggiano la pianificazione e la strategia militari. Così come viene definito in questo articolo, un regime che escluda il primo uso rafforzerebbe la coesione dell'alleanza NATO e renderebbe i rapporti con l'Unione Sovietica meno rischiosi in momenti di crisi, riducendo così in misura notevole il pericolo di una guerra nucleare.

SOCIETÀ EDITORIALE FARMACEUTICA

s.r.l.
Via Ausonio, 12 - 20123 Milano

D. Valente - L. Prencipe
R. Pozzoli - F. Chiodo
Laboratorio di Biochimica
e Microbiologia
dell'Ospedale Maggiore di Milano

LE ANALISI CLINICHE

edizione 1983

- Modalità di prelievo
- Valori normali
- Significato delle analisi
Volume di pagine 288 - L. 22.000

R. Pozzoli - F. Chiodo
Laboratorio di Biochimica
e Microbiologia
dell'Ospedale Maggiore di Milano

IL LABORATORIO CLINICO DI ANALISI MICROBIOLOGICHE

- Nozioni di carattere generale
- Cenni di sistematica batterica
- Esami batteriologici
- Sierologia
- Esami sierologici
- Cenni di parassitologia
- Cenni di virologia

Volume di pagine 336 - 63 figure - L. 36.000

CORSO DI AGGIORNAMENTO SU ARGOMENTI DI FARMACOLOGIA

Atti del corso, organizzato dall'Università di Milano, dall'Ordine dei Farmacisti della Provincia di Milano, diretto dal Prof. R. Paoletti, sui più attuali temi della moderna farmacologia. Utile per l'aggiornamento e per i concorsi alle sedi farmaceutiche.
Volume di pagine 85 - 15 figure - L. 22.000

I BATTERI ANAEROBI

Questo volume ha il vanto di essere l'unico testo sulla materia in lingua italiana. Tratta di patologia, isolamento, identificazione e sensibilità agli antibiotici e capitoli specifici sono dedicati ai generi Borellia, Campylobacter e Treponema. Il libro è d'interesse per medici, clinici, laboratoristi e studenti.
Volume di pagine 79 - 12 tabelle - 6 tavole - L. 16.500

Gradirei ricevere:

N. copie di LE ANALISI CLINICHE
N. copie di IL LABORATORIO CLINICO DI ANALISI MICROBIOLOGICHE
N. copie di CORSO DI AGGIORNAMENTO SU ARGOMENTI DI FARMACOLOGIA
N. copie di I BATTERI ANAEROBI

Pagherò c/assegno al postino l'importo di cui sopra (+ L. 2.000 di spese)

Signor

Via N.

Cap. Città.

La dinamica degli abissi

Le correnti fredde che scorrono verso l'equatore a grandi profondità sono spesso perturbate da violente ed estese tempeste che trasportano enormi masse di sedimenti da un punto all'altro del fondo oceanico

di Charles D. Hollister, Arthur R. M. Nowell e Peter A. Jumars

«Era una notte buia e tempestosa». Questa frase iniziale d'un romanzo dell'epoca vittoriana (tanto citata a scopo parodistico) riesce a descrivere con esattezza le condizioni esistenti a 4800 metri di profondità marina, alla base della risalita continentale, a 450 miglia al largo della costa del New England, a metà settembre 1979. Naturalmente a tutte le profondità superiori ai 1000 metri è sempre buio. Gli strumenti che avevamo ormeggiato alcune settimane prima, alla base del rialzo della Nuova Scozia, a profondità assai superiore, rilevarono la presenza d'una «tempesta» che imperversò per circa una settimana. I nostri strumenti indicavano che una massiccia corrente abissale si spostava verso sud-est a una velocità superiore a mezzo metro al secondo (circa un nodo), sconvolgendo al suo passaggio il fondo oceanico. Le forze che agivano sugli strumenti ormeggiati erano pari a quelle che al di sopra del livello del mare sarebbero state generate da un vento di burrasca di velocità compresa tra i 34 e i 40 nodi (ossia tra 63 e 74 chilometri all'ora).

La scoperta di un fenomeno che coinvolgeva una così notevole quantità di energia a 4800 metri di profondità stupì soprattutto coloro che erano convinti che gli abissi fossero altrettanto calmi che bui. Studi recenti hanno ora dimostrato che sul lato occidentale dei grandi bacini oceanici, al piede della risalita continentale, tempeste periodiche trasportano enormi masse di fini sedimenti che modificano radicalmente il fondo marino, erodendolo in alcune aree e depositando grandi volumi di silt e di argilla in altre. La maggior parte delle prove dell'esistenza di queste correnti abissali sono state raccolte da un programma patrocinato dall'Office of Naval Research e noto con la sigla HEBBLE (da High-Energy Benthic Boundary-Layer Experiment).

Il quadro generale emerso è che i fenomeni dinamici presenti nelle profondità oceaniche imitano grosso modo la circolazione generale dell'atmosfera, la quale dà origine ai sistemi meteorologici più attivi su scala planetaria. I vortici ciclonici

delle medie latitudini, spesso associati a forti venti e a intense precipitazioni, hanno la loro controparte nei potenti vortici di profondità, che durano da alcuni giorni a due settimane e che spostano una «tempesta» di silt e di argilla da un luogo all'altro. Tempeste del genere non sono tuttavia ubiquitarie e non interessano neppure tutto il lato occidentale dei bacini oceanici. Per esempio, a sole 50 miglia circa a ovest del sito HEBBLE, a una profondità di 2500 metri, dove giace il relitto del sommergibile nucleare *Thresher*, le fotografie non mostrano alcun segno di sconvolgimento del fondo marino attorno allo scafo e inoltre sulla superficie dell'imbarcazione si è depositata solo una quantità molto modesta di silt. Sembra che le tempeste seguano alcuni percorsi ben definiti, ma la loro frequenza in ogni dato sito è ancora poco nota. Secondo stime preliminari, sembra che le tempeste che provocano erosione capitino nella regione più studiata, circa una volta ogni due mesi.

L'erosione periodica del fondo oceanico e la deposizione saltuaria di spessi strati di materiale fine producono una stratigrafia complessa, molto diversa da quella che si svilupperebbe per effetto di lievi piogge di sedimenti a cui spesso fanno riferimento i geologi marini. Il trasporto episodico di sedimenti crea sequenze stratificate che rassomigliano molto a quelle tipiche dei mari poco profondi, dove forti tempeste di vento creano quelle strutture chiamate tempestiti, nelle quali la normale semplice stratificazione è sovrachiata dall'accavallamento dei letti sedimentari.

Le nostre nuove scoperte potrebbero far rivedere ai geologi impegnati in ricerche petrolifere le idee tradizionali sulle località più opportune per la scoperta di depositi non ancora sfruttati. La scoperta di forti correnti e di tempeste abissali ha anche implicazioni importanti per quanto riguarda le operazioni militari condotte con i sommergibili e la strategia antisommergibile, la sistemazione e la progettazione di strutture destinate al fondo marino e la realizzabilità di depositi di rifiuti nelle profondità oceaniche.

L'idea della tranquillità degli abissi è rimasta così a lungo indiscussa che molti ricercatori sono stati inizialmente riluttanti ad accettare i dati relativi alle correnti e alle tempeste violente presenti nelle profondità oceaniche. Le prime considerazioni a favore dell'esistenza di tali correnti sono state di natura teorica. L'acqua fredda è più densa dell'acqua calda e i modelli della circolazione oceanica dimostravano che la discesa di acqua fredda in vicinanza dei poli avrebbe dovuto generare correnti forti, profonde e costanti in direzione dell'equatore. Osservazioni successive non solo hanno confermato la presenza di correnti profonde, ma hanno svelato altresì l'esistenza, sul lato occidentale dei bacini oceanici, di vortici che possono avere un'energia circa 300 volte superiore a quella della corrente media. Le fotografie del fondo oceanico situato sotto le correnti profonde hanno evidenziato estesi letti di depositi i quali indicano un trasporto attivo di sedimenti. Le prove decisive a dimostrazione dell'attività dinamica presente a grandi profondità sono state fornite da misurazioni dirette delle correnti e dei sedimenti condotte nell'Atlantico settentrionale nell'ambito del programma HEBBLE.

Prima di descrivere nei particolari i risultati del programma HEBBLE ci soffermiamo brevemente sull'origine e sulla deposizione dei sedimenti di mare profondo e sulle forze che attivano la circolazione oceanica globale. I sedimenti che finiscono sul fondo oceanico sono di due tipi principali. Uno è costituito dai detriti provenienti dall'alterazione da parte degli agenti atmosferici delle rocce continentali e insulari. Questi detriti, assieme a sostanze vegetali in decomposizione provenienti da piante terrestri, vengono portati dai fiumi fino al bordo dei continenti e poi sulla piattaforma continentale, dove vengono prelevati dalle correnti marine. Non appena questo materiale raggiunge il margine della piattaforma, precipita verso la base della risalita continentale per effetto del proprio peso. Grandi quantità di materiale terrigeno vengono anche portate fino al mare nelle regioni subtro-

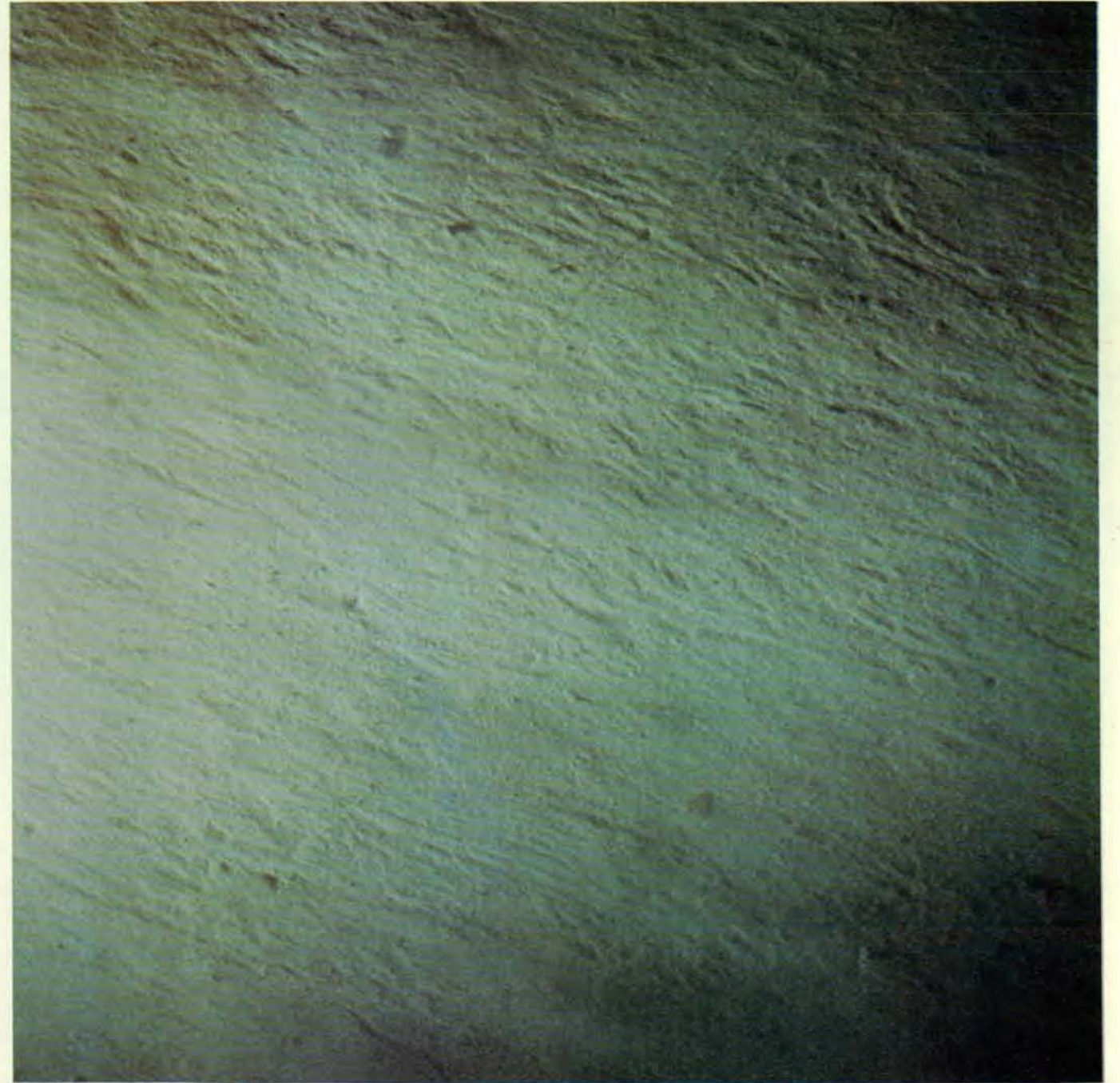
picali dai forti venti del deserto. Ogni anno circa 15 miliardi di tonnellate di materiale continentale raggiungono la foce di torrenti e di fiumi; di queste, la maggior parte si ferma alla foce o sulle piattaforme continentali, mentre solo pochi miliardi di tonnellate raggiungono le profondità oceaniche.

Il secondo importante tipo di sedimenti che giunge sul fondo oceanico è costituito dai gusci e dagli scheletri di organismi microscopici che proliferano e muoiono nei primi 100 metri superficiali di tutti gli oceani illuminati dal Sole. Questo mate-

riale biologico contribuisce all'accumulo totale di sedimenti del fondo per circa tre miliardi di tonnellate all'anno. I tassi di accumulo sono regolati dai tassi di produttività biologica, i quali sono controllati (in parte) a loro volta dalle correnti superficiali. Quando le correnti superficiali si incontrano si dice che convergono; quando si allontanano, si dice che divergono. Le zone di divergenza delle principali masse idriche consentono anche alle acque più profonde, ricche di nutrienti, di «affiorare» nella zona illuminata, dove avviene la fotosintesi, con la conseguente

fissazione di carbonio. Queste fasce di alta produttività e di accumulo elevato si trovano normalmente attorno ai principali fronti oceanici (come la regione attorno all'Antartide) e lungo i bordi delle correnti principali (come la Corrente del Golfo al largo del New England e la Kuroshio al largo del Giappone). Le acque ricche di nutrienti affiorano anche in una zona situata lungo l'equatore, dove esiste una divergenza tra due grandi «celle» di circolazione le quali sono legate all'azione del vento.

La velocità di sedimentazione degli or-



Questa zona di fondo oceanico, che appare solcata ed erosa, è stata fotografata a una profondità di 4850 metri alla base della risalita continentale a circa 450 miglia al largo della costa del New England. Questo sito viene studiato nell'ambito del programma HEBBLE (High-Energy Ben-

thic Boundary-Layer Experiment), a cui gli autori partecipano. Il progetto è finanziato dall'Office of Naval Research. L'assenza sul fondo di caratteristiche biologiche conferma i dati ottenuti dai correntometri ormeggiati: il fondo è stato eroso da poco da una tempesta abissale.

ganismi marini viene influenzata fortemente anche dalla profondità dell'acqua. La possibilità che un guscio resista alla dissoluzione dopo la discesa verso il fondo oceanico è influenzata notevolmente dal tempo di permanenza sul fondo prima di essere coperto dai sedimenti e quindi protetto dall'azione corrosiva dell'acqua di mare profon-

do, che è sottosatura di carbonati di calcio. Nelle acque più profonde la deposizione è in genere più lenta, i gusci hanno un percorso di discesa più lungo e attraversano acque progressivamente più corrosive. Se le acque superiori fossero uniformemente produttive e se non vi fossero correnti abissali, la coltre di materiale

sedimentario che si depositerebbe sopra il fondo oceanico, originariamente vulcanico e accidentato, sarebbe uniforme. Ma ciò non avviene. Dando uno sguardo al mappamondo si può vedere facilmente che i più grandi fiumi dell'America Settentrionale e dell'America Meridionale sfociano nell'Atlantico il quale riceve

quindi una quantità molto maggiore di materiale sedimentario fluviale rispetto a quella che riceve il Pacifico. Inoltre le fosse oceaniche profonde distribuite attorno al Pacifico inglobano buona parte del materiale che raggiunge il bordo occidentale di questo oceano.

L'Atlantico è inoltre più piccolo e

meno profondo del Pacifico, cosicché i residui d'origine biologica vengono sepolti più rapidamente e si conservano più a lungo rispetto a quelli del Pacifico. Perciò in media il fondo dell'Atlantico riceve più sedimento per unità di superficie (circa un centimetro ogni 1000 anni) del fondo del Pacifico. Inoltre l'andamento globale del-

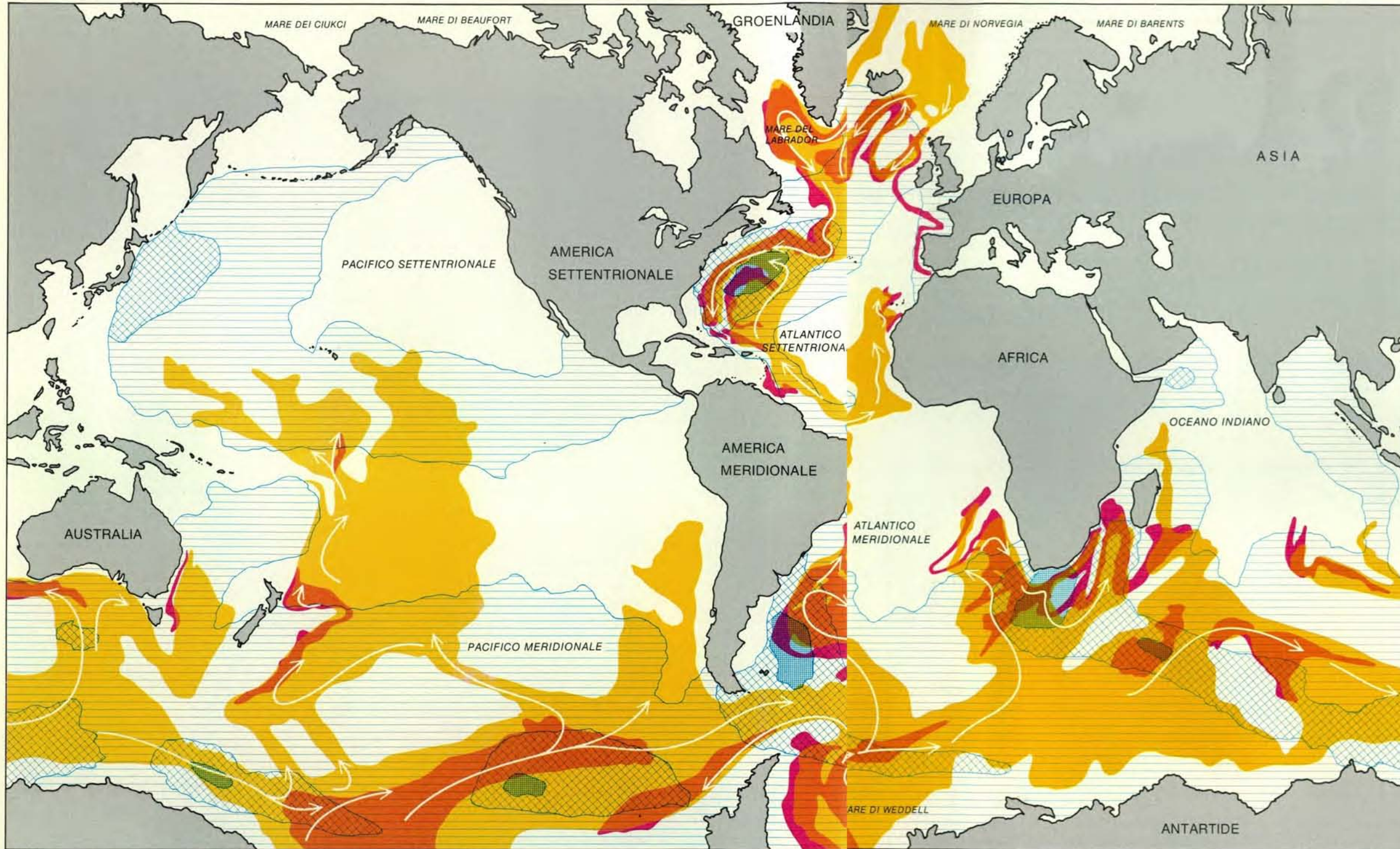
le forti correnti abissali crea una possibilità maggiore di redistribuzione nell'Atlantico rispetto al Pacifico. Per tutte queste ragioni l'Atlantico è l'oceano ideale in cui studiare gli effetti geologici della intensa circolazione abissale.

Consideriamo ora la natura delle correnti abissali. Nelle regioni polari l'acqua superficiale è più densa rispetto all'acqua di latitudine inferiore, perché è più fredda e più salata. (La salinità dell'acqua superficiale che fluisce verso i poli aumenta a causa dell'evaporazione; inoltre aumenta ulteriormente per la cessione di sale da parte dei ghiacci in formazione.) Man mano che aumenta di densità, l'acqua scende e nel contempo si diffonde verso l'equatore. Secondo le prime teorie, questo flusso termalino avrebbe assunto la forma di correnti profonde che sarebbero state deviate verso ovest nel fluire verso l'equatore. Questa deviazione è una semplice conseguenza della rotazione verso est della Terra che genera una forza palese, quella di Coriolis. Il percorso preciso seguito dalla circolazione dell'acqua è influenzato anche dalla distribuzione delle masse continentali e dalla topografia del fondo oceanico.

Tutt'e tre gli oceani principali, Atlantico, Pacifico e Indiano, sono direttamente connessi con l'Antartide, dove sono alimentati da correnti fredde e profonde. Poiché queste correnti fluiscono verso l'equatore, vengono deviate verso ovest dalla forza di Coriolis. Così una massa d'acqua che si sposta verso nord nell'emisfero australe è deviata verso sinistra se si guarda nel senso della corrente. Le correnti profonde sono spinte contro il margine occidentale di tutt'e tre i bacini oceanici e delle loro dorsali. Questa pressione contro i continenti fa acquistare velocità alle correnti, proprio come un corso d'acqua scorre più velocemente a mano a mano che il suo alveo si restringe. L'oceano Indiano è tuttavia diverso dagli altri due in quanto non è collegato con la regione polare artica e quindi dispone di una sola fonte d'acqua profonda fredda.

Sia il Pacifico sia l'Atlantico sono entrambi collegati al Mare Artico, ma nel Pacifico il flusso di acqua artica profonda è trascurabile in quanto il Pacifico settentrionale è efficacemente bloccato a nord dal poco profondo Stretto di Bering, che separa l'Alaska dall'Unione Sovietica. Inoltre l'acqua del Pacifico, essendo meno salata dell'acqua atlantica, gela più facilmente per effetto del freddo vento subartico, mentre in prossimità della superficie l'acqua che non gela non è sufficientemente ricca di sali e quindi non abbastanza densa per scendere fino al fondo.

La situazione dell'Atlantico è completamente diversa. L'Atlantico ha due fonti sostanziali di acqua salata: l'acqua del Golfo del Messico, che viene trasportata verso nord dalla Corrente del Golfo, e la corrente intermedia proveniente dal Mediterraneo. Questi flussi d'acqua rendono l'acqua superficiale dell'Atlantico settentrionale più salata di quella del Pacifico settentrionale. La parte più prossima alla superficie dell'acqua dell'Atlantico setten-



Le correnti abissali (in giallo) trasportano acqua fredda e densa dalle regioni polari nei tre grandi bacini oceanici della Terra. Sebbene entrambi i bacini dell'Atlantico e del Pacifico siano collegati con il Mare Artico non è molta l'acqua profonda di quest'ultimo che riesce a fluire nel Pacifico attraverso lo Stretto di Bering poco profondo. L'Atlantico, invece, riceve forti correnti di acqua fredda abissale da entrambe le regioni polari. Le

correnti vengono convogliate dalla topografia del fondo oceanico e vengono deviate verso ovest dalla forza di Coriolis, una conseguenza della rotazione terrestre. Le aree con varie intensità di tratteggio blu rappresentano variazioni di energia cinetica associate a vortici vicini alla superficie. Le aree con il tratteggio incrociato più fitto corrispondono a un'energia cinetica (espressa in unità indicanti la fluttuazione di velocità attorno alla

velocità media) di oltre 20 centimetri quadrati per secondo quadrato. Il tratteggio orizzontale rappresenta valori compresi tra quattro e 10 nella medesima unità di misura; il tratteggio incrociato meno fitto rappresenta valori compresi tra 10 e 20. I valori di energia cinetica dei vortici sono basati in parte su osservazioni fatte da imbarcazioni, in parte su misurazioni compiute da satelliti delle variazioni del livello oceanico. L'energia dei vortici viene trasferita al fondo oceanico, dove interagisce con le correnti abissali per erodere e ridistribuire i sedimenti, creando grandi aree (in rosso) di sedimenti di deriva.



È possibile individuare i siti calmi e i siti attivi delle profondità marine con le fotografie del fondo. In un'area calma (a sinistra) la superficie appare modificata dall'azione degli organismi che vi vivono. Dal fondo sporgono anche di 10 centimetri i sottili tubi costruiti dai foraminiferi, animali con un minuscolo guscio. In una zona caratterizzata da forti



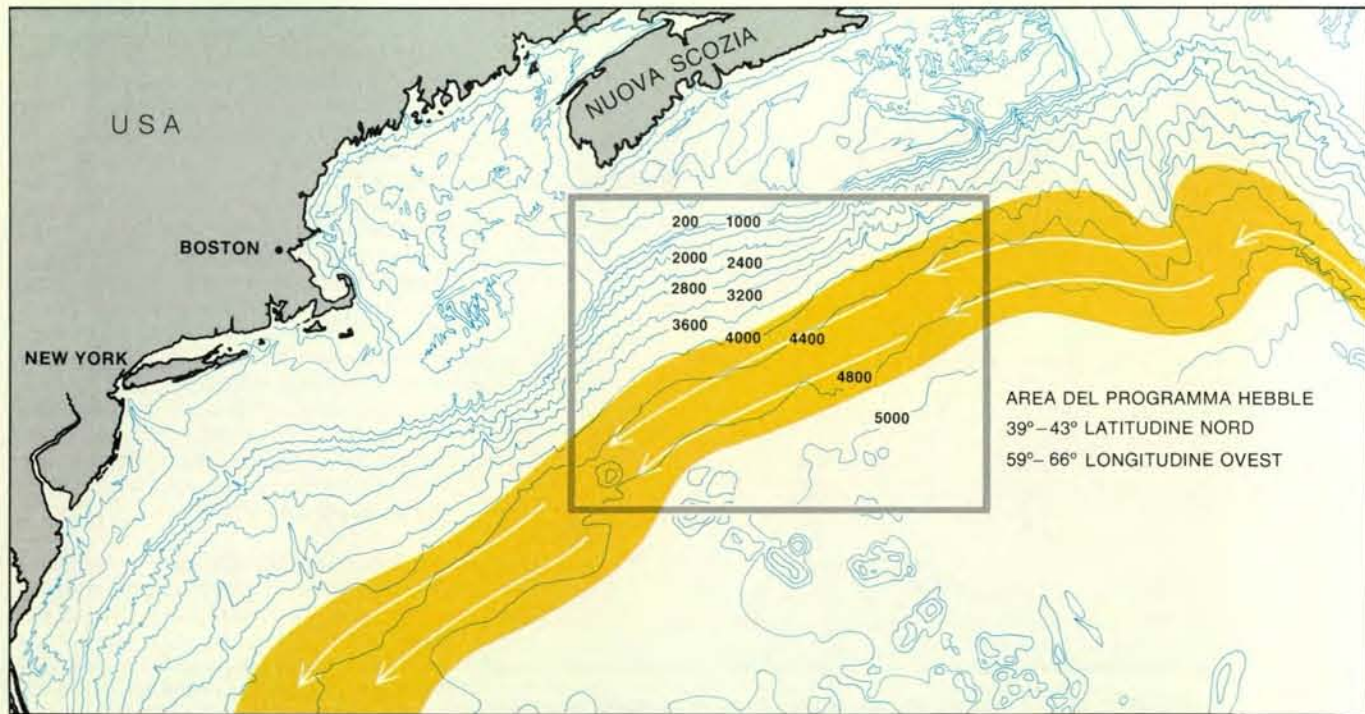
correnti abissali (a destra) il fondo è eroso e la direzione del flusso della corrente è rivelata da «code» di sedimenti sul lato riparato di piccoli monticelli. La fotografia a sinistra è stata ripresa sulla scarpata continentale a una profondità di 1000 metri, quella a destra alla base della risalita continentale del New England a una profondità di 5000 metri.

trionale si muove verso nord ed entra nel Mar di Norvegia, dove si raffredda a circa zero gradi centigradi senza congelare, diviene più densa e affonda. Raggiungendo il fondo, l'acqua fredda e densa inverte la direzione e scorre di nuovo verso l'Atlantico attraverso una serie di soglie strette e profonde nella dorsale che congiunge Groenlandia, Islanda e Scozia.

Questa grande massa d'acqua profonda, incanalata dalla topografia del fondo, si sposta verso sud ed è sospinta verso destra, nel senso della corrente, contro il margine continentale del Canada orientale e degli Stati Uniti orientali, formando la Corrente intermedia occidentale (la Western Boundary Undercurrent). L'esistenza di questa corrente fu scoperta

circa vent'anni fa. Attualmente si valuta che questa corrente trasporti circa 10 milioni di metri cubi d'acqua al secondo lungo la costa orientale degli Stati Uniti. Durante le tempeste abissali la corrente è in grado di trasportare circa una tonnellata di sedimenti al minuto.

Con questa premessa, si possono studiare gli oceani e tentare di scoprire dove



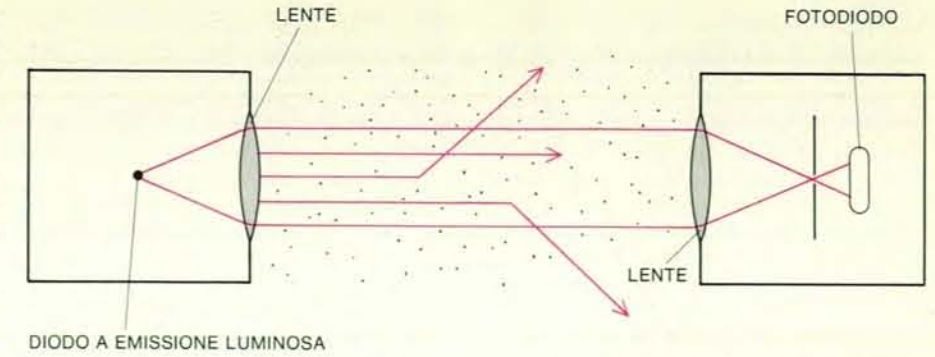
L'area di fondo oceanico studiata dal programma HEBBLE si trova a una profondità di 4800 metri, 450 miglia a est di Boston, alla base della risalita continentale. La banda in giallo mostra come le correnti abissali fredde che hanno origine vicino alla Groenlandia e nel Mar di Norvegia

vengano deviate verso ovest contro la scarpata continentale dell'America Settentrionale e si sollevino scorrendo verso l'equatore. Le tempeste periodiche elevano la velocità della corrente nel sito HEBBLE fino ad un nodo e oltre, sollevando «tormente» di silt e di fango.

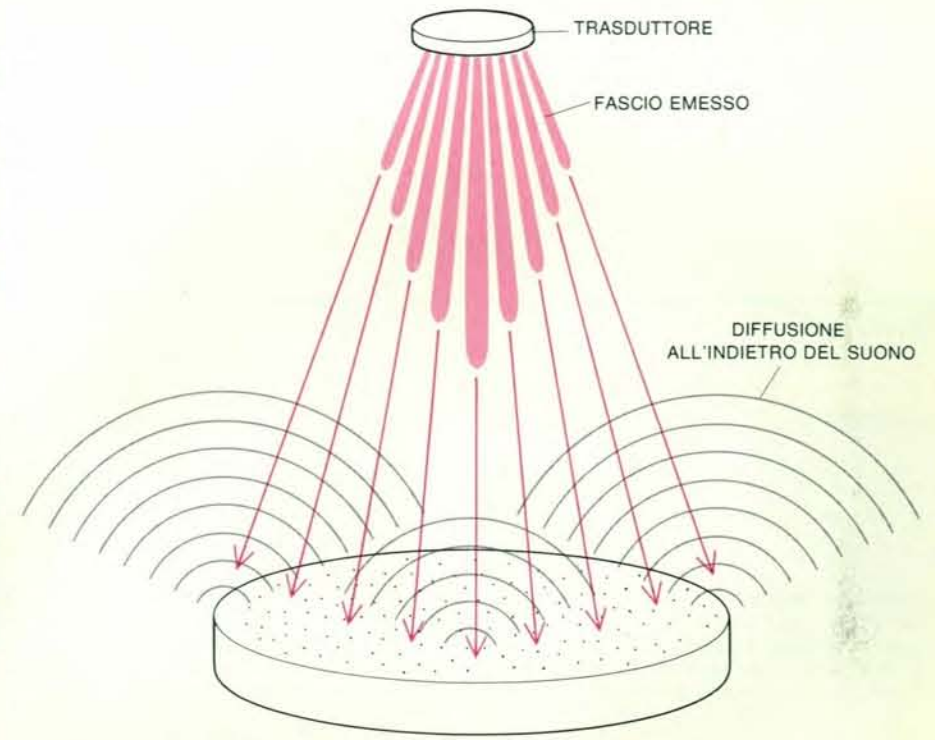
possano esistere depositi di fango spostati dalle correnti ed eventualmente modificati da tempeste abissali. Chiaramente le maggiori masse di fango inorganico e le più intense correnti abissali debbono trovarsi nell'Atlantico occidentale. Le latitudini elevate del lato occidentale dell'Atlantico, sia settentrionale sia meridionale, sono potenzialmente la sede ideale per tempeste abissali: in queste zone infatti le fotografie rivelano un fondo plasmato e modificato da forti correnti. Questi sedimenti costituiscono i più estesi depositi di tutta la Terra: fino a 200 chilometri di larghezza, 1000 chilometri di lunghezza e con spessori di 2000 metri. Esempi notevoli di questo fenomeno sono le dorsali esterne Blake-Bahama, al largo della Florida. Sembra plausibile che l'intera risalita continentale atlantica a profondità comprese fra 3000 e 5000 metri al largo dell'America Settentrionale e dell'America Meridionale sia stata modellata in larga misura dalle correnti abissali che ne seguono il contorno. Grazie alle perforazioni in mare profondo, si è scoperto che le correnti abissali hanno eroso estese superfici dell'Atlantico occidentale fino a circa 20-15 milioni di anni or sono e che da allora la deposizione controllata dalle correnti abbia prevalso, con la formazione di grandi depositi sedimentari di deriva.

Altrove le correnti abissali hanno modellato la distribuzione di materiale a granulometria fine lungo i bordi dell'Africa, dell'Antartide, dell'Australia, dell'India e della Nuova Zelanda. In tutta questa regione la formazione di grandi depositi di materiale prevalentemente biogenico è favorita dalla concomitanza dell'elevata produttività biologica lungo la divergenza circumpolare e lungo la corrente circumpolare da vento. Nonostante che molti di questi depositi di deriva non siano sempre attribuibili direttamente alle tempeste abissali, non vi è dubbio che aree veramente estese di fondo oceanico siano modificate dal materiale trasportato dalle correnti profonde. Può darsi che il ruolo principale delle tempeste periodiche sia quello di rimescolare i sedimenti, che successivamente vengono trasportati dalle correnti abissali meno intense, ma più persistenti.

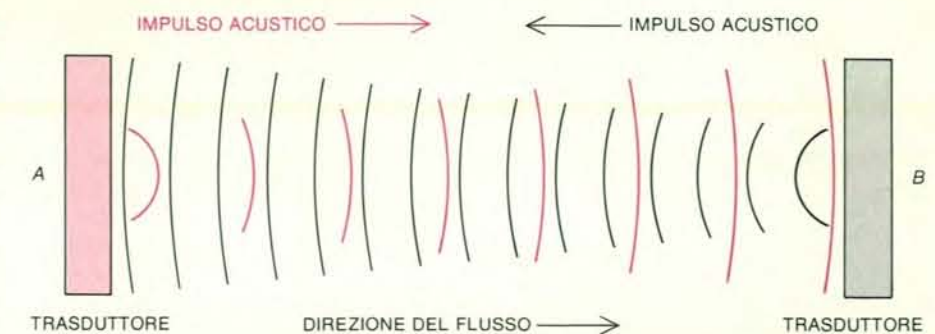
Grazie agli strumenti ormeggiati sul fondo oceanico, altri importanti particolari si sono potuti aggiungere al quadro intuitivo dei fenomeni di erosione, di trasporto e di rideposizione dei sedimenti. Il processo non può essere facilmente riprodotto in laboratorio. L'acqua in movimento esercita sostanzialmente una sollecitazione sul fondo, la quale è in definitiva responsabile dello spostamento del materiale a granulometria fine e del modellamento dei depositi. Un flusso costante sopra un letto orizzontale assume una struttura caratteristica. All'interfaccia tra questa acqua e il fondo esiste uno strato di molecole d'acqua aderente a quest'ultimo e che è essenzialmente immobile. Direttamente al di sopra di questa interfaccia, se il flusso è lento e il fondo è liscio, un sottile strato di acqua si



Questo turbidimetro è impiegato nel programma HEBBLE. La quantità di luce che raggiunge il fotodiode a destra è strettamente correlata con la concentrazione di particelle presenti nel percorso luminoso. La sorgente luminosa (a sinistra) emette a una lunghezza d'onda di 660 nanometri.



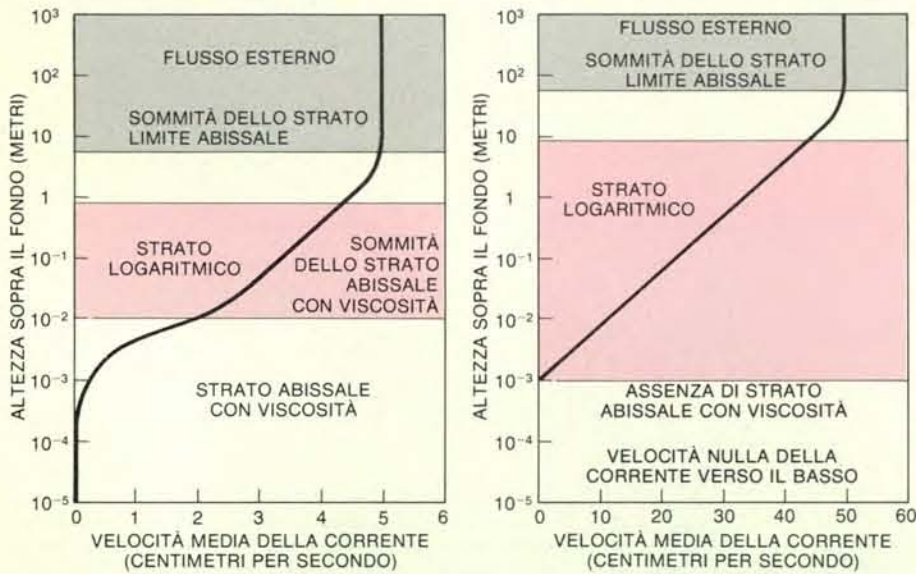
Un altro strumento che fornisce dati sulla concentrazione di particelle nell'acqua presso il fondo oceanico è un sistema che misura l'energia acustica diffusa all'indietro: emette impulsi sonori e misura la quantità di energia riflessa all'indietro dalle particelle presenti in un dato volume di acqua.



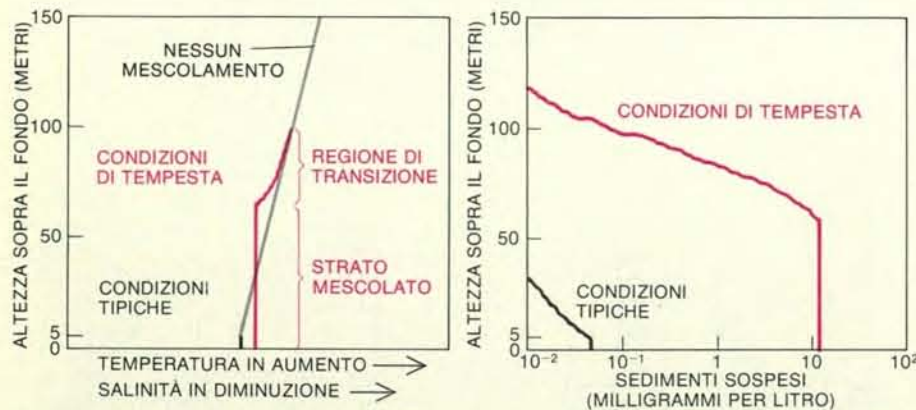
La velocità e gli sforzi del flusso possono essere determinati misurando il tempo necessario a un impulso sonoro per percorrere una distanza nota. I ricercatori del programma HEBBLE chiamano l'apparecchiatura qui raffigurata schematicamente «sensore di sollecitazione acustica». Se la corrente fluisce verso destra, il tempo che un impulso sonoro impiega per viaggiare da A a B sarà minore del tempo che impiega per percorrere la distanza fra i due punti in direzione opposta. La differenza di tempo fornisce la velocità del fluido. La grandezza nota come «sforzo turbolento» è calcolata nella massa d'acqua in esame mediante due misurazioni ad angolo retto l'una rispetto all'altra.

sposterà sopra il letto in un flusso laminare, in cui la forza dominante è la viscosità dell'acqua. All'altezza d'un centimetro o meno, le forze viscosive vengono superate dall'inerzia del fluido, e perciò predomina il rimescolamento turbolento.

All'interno dello strato di flusso laminare, quando e dove questo strato esiste, la velocità di flusso aumenta linearmente con l'altezza. Nei pochi metri di flusso turbolento vicino al fondo il rimescolamento è intenso e la velocità aumenta esponenzialmente con l'altezza. Se si rappresenta la velocità media di flusso in funzione dell'altezza al di sopra del fondo,



Le proprietà di una massa d'acqua (come salinità, temperatura e contenuto di solidi sospesi) sono rimescolate attivamente nello strato «logaritmico», in cui la velocità media aumenta linearmente con l'altezza, quando l'altezza al di sopra del fondo è riportata su una scala logaritmica. Quando la corrente è lenta e il fondo è liscio (a sinistra), un sottile strato d'acqua si sposta sopra il fondo in un flusso viscoso, ossia laminare. Se la corrente è rapida oppure il fondo è accidentato (a destra), lo strato logaritmico si estende per tutto il tragitto da un millimetro a circa 10 metri sopra il fondo.



La presenza di tempeste influenza il modo in cui salinità, temperatura e quantità di materiale sospeso sono distribuite a seconda dell'altezza sopra il fondo marino. Poiché la salinità e la temperatura seguono andamenti analoghi, sono rappresentate da un'unica curva (a sinistra). In condizioni di flusso tipico abissale (in nero) le due grandezze aumentano linearmente con l'altezza al di sopra del fondo, se si eccettuano i cinque metri circa inferiori, dove si mantengono su valori sostanzialmente costanti. In condizioni di tempesta (in colore) nel flusso abissale viene introdotta una quantità d'energia sufficiente a estendere lo strato rimescolato fino a un'altezza di 60-70 metri, con, al di sopra, una zona di transizione che si estende grosso modo per 30 metri. L'effetto delle tempeste abissali sulla quantità di sedimento sospeso nelle acque profonde è visualizzato a destra, dove le condizioni tipiche sono in nero mentre quelle di tempesta sono in colore.

usando una scala logaritmica per l'asse delle ordinate si ottiene una retta; per questo motivo si parla di strato logaritmico. In questa regione le proprietà globali dell'acqua, come la salinità, la temperatura e il contenuto di solidi in sospensione, sono attivamente rimescolate.

A grandi scale verticali i moti dell'acqua vengono influenzati dalla forza di Coriolis: nell'emisfero boreale i fluidi la cui accelerazione aumenta vengono deviati verso ovest, quelli la cui accelerazione diminuisce sono deviati verso est. Man mano che le correnti vengono frenate per l'attrito esercitato dai fluidi sottostanti

che si muovono più lentamente e, infine, anche per l'attrito esercitato dal fondo la loro direzione vira verso est, con un andamento caratteristico, noto come spirale di Ekman (dal nome dell'oceanografo svedese V. Walfrid Ekman). La forza di taglio tra gli strati di fluido che si muovono a velocità diverse continua ad alimentare un rimescolamento turbolento.

Gli effetti dell'attrito esercitato dal fondo non si estendono oltre i 100 metri verso l'alto della colonna d'acqua. Questa regione, in cui le proprietà sono ben rimescolate dalla turbolenza, viene di solito chiamata «strato mescolato abissale». Il suo spessore è limitato dall'attenuazione del rimescolamento turbolento, un processo che estrae energia dal flusso. A una certa altezza, proporzionale sia alla velocità della corrente sovrastante sia all'entità della forza di Coriolis, la forza di taglio non è più sufficiente a provocare un rimescolamento turbolento sopra i livelli oceanici di base.

La sommità dello strato mescolato abissale è di solito considerata il limite dello spessore di questo strato. A differenza dell'atmosfera, le acque abissali o non sono stratificate in base alla densità oppure, al massimo, sono stratificate in minima misura (in altre parole, le acque più fredde e più salate stanno sotto quelle più calde e più dolci). Di conseguenza, le perturbazioni non si sviluppano allo stesso modo di quelle atmosferiche. Ogni stratificazione stabile avrebbe l'effetto di diminuire lo spessore dello strato mescolato, poiché sarebbe necessaria più energia per il rimescolamento.

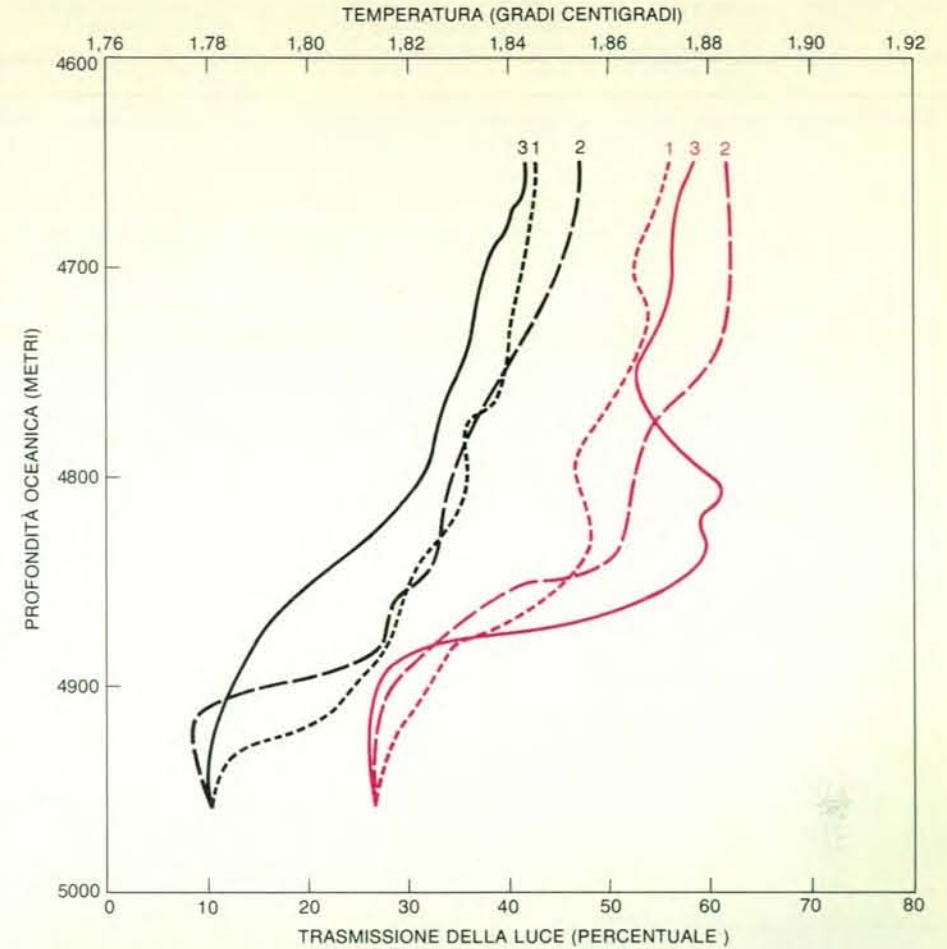
L'erosione può essere causata solo dall'energia che raggiunge il fondo. Tanto più è accidentata la topografia del fondo oceanico e più rapido è il flusso sovrastante, tanto maggiore è la sollecitazione di taglio che agisce sui sedimenti del fondo. Al di sopra del livello critico della sollecitazione di taglio in grado di iniziare il processo di erosione, il tasso dell'erosione cresce esponenzialmente al crescere della sollecitazione di taglio. Una volta che i sedimenti sono stati sollevati dal fondo, cadono sotto l'azione di due forze opposte; una è la turbolenza, che tende a diffondere le particelle verso l'alto, l'altra è la gravità, che invece le fa sedimentare. Quanto maggiore è la velocità di sedimentazione, tanto maggiore è il gradiente di concentrazione nei pressi del fondo. Quanto più è veloce la corrente e quanto più è accidentato il fondo, tanto più uniformemente varia la concentrazione con l'altezza al di sopra del fondo. I sedimenti di mare profondo sono costituiti tipicamente da argille e da silt, le cui particelle hanno dimensioni inferiori a 30 micrometri. Tali particelle tendono a depositarsi con tale lentezza che le concentrazioni di sedimento sono quasi uniformi in tutto lo strato mescolato abissale, in presenza di correnti moderate. Di conseguenza, non appena il materiale finissimo viene sollevato da una tempesta abissale intensa, è possibile che esso venga trasportato anche a grandi distanze da correnti relativamente deboli.

Le particelle sedimentarie fini hanno un rapporto tra superficie e volume alto, perciò hanno anche una elevata capacità di adsorbimento delle sostanze disciolte presenti nell'acqua in cui le particelle vengono rimescolate. Per esempio i granuli delle argille naturali hanno un'area superficiale compresa tra 10 e 100 metri quadrati per grammo di sedimento. Tra le sostanze raccolte dalle correnti abissali vi sono isotopi radioattivi sia naturali sia artificiali. Se ammettiamo che il tasso di immissione di isotopi radioattivi di origine naturale e il tasso di sedimentazione siano grosso modo costanti, come sembra avvenga per molti isotopi in molti ambienti marini abissali, il tasso di diminuzione della radioattività nei sedimenti con la profondità può servire come orologio naturale per valutare i tassi di sedimentazione.

Ammettendo che i sedimenti non vengano rimescolati dopo la deposizione, la concentrazione d'un isotopo radioattivo adsorbito diminuirà esponenzialmente a un tasso inversamente proporzionale alla sua vita media. Gli isotopi naturali studiati comunemente in queste cronologie sedimentarie hanno una vita media che va da 24,1 giorni (per il torio 234) a 5730 anni (per il carbonio 14). Parimenti utili per la datazione assoluta degli strati sedimentari sono parecchi isotopi radioattivi introdotti dalle esplosioni delle bombe atomiche nell'atmosfera; tra questi possiamo citare il cesio 137 (vita media di 30,1 anni), il plutonio 239 (24 000 anni) e il plutonio 240 (6580 anni).

Benché il rimescolamento dei sedimenti complichino la stima del tasso netto di sedimentazione, esso permette di trarre altre conclusioni. I primi 10 centimetri circa della maggior parte dei sedimenti marini a tutte le profondità oceaniche sono rimescolati in modo più o meno continuo dall'attività di alimentazione e di scavo degli organismi che vi abitano. Gli organismi che vivono sul fondo sono anche responsabili della conformazione topografica a piccolissima scala del fondo e di quelle «colle» biologiche (per esempio muco) che influenzano sia l'entità della sollecitazione di taglio del fluido sia la sua efficacia nell'erosione i sedimenti. La maggior parte degli animali che agitano il fondo si nutrono digerendo batteri e altri componenti organici dei sedimenti proprio alla stregua dei comuni lombrichi.

L'abbondanza di questi organismi generalmente diminuisce con la profondità dell'acqua e la distanza dalla costa, seguendo così da vicino la stessa velocità di sedimentazione. L'accumulo netto di sedimenti diminuisce in modo tipico da alcuni centimetri ogni 1000 anni alla profondità di 1000 metri, fino ad alcuni millimetri ogni 1000 anni nell'oceano aperto. Scegliendo opportunamente isotopi con particolari vite medie, si possono ricavare stime sia della profondità dello strato rimescolato da attività biologiche sia dell'intensità del rimescolamento di origine biologica. Di regola in condizioni di rimescolamento biologico costante, due isotopi con vita media diversa non forniscono profili di forma identica.



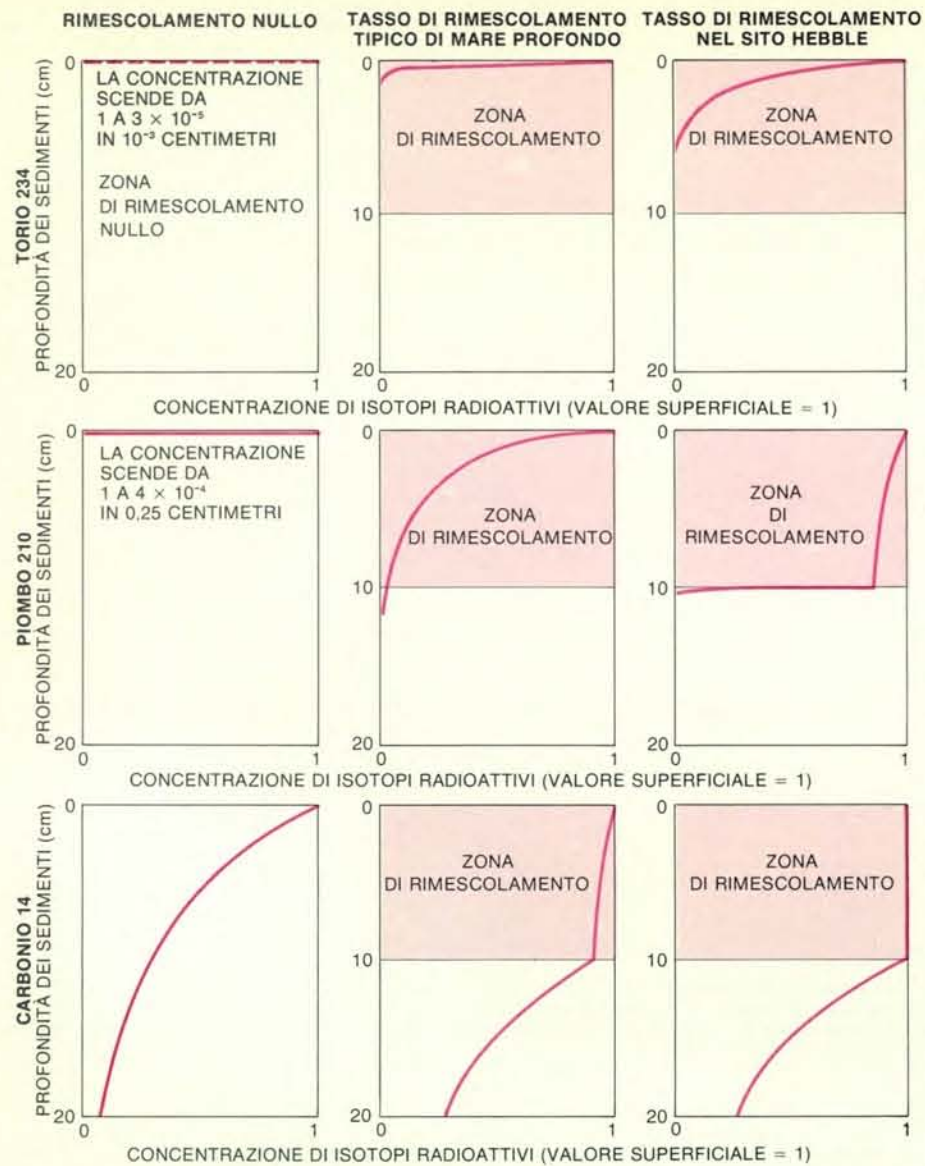
In condizioni di tempesta, nel sito HEBBLE, sono state registrate in un'ora rapide variazioni di temperatura (in nero) e di trasmissione di luce (in colore). La trasparenza dell'acqua varia con la concentrazione di particelle solide sospese indicando, quindi, cambiamenti nelle quantità di sedimento sollevato dalla corrente. I profili sono stati registrati a intervalli di 20 minuti.

Nella scala temporale degli isotopi radioattivi impiegati in stratigrafia, l'erosione fisica e la rideposizione sono quasi istantanee. Uno strato ridepositato dovrebbe quindi mostrare concentrazioni uniformi di tutti gli isotopi lungo tutto il suo spessore. Perciò in linea di principio è possibile valutare la profondità di un deposito influenzato da tempeste abissali e, confrontando i profili di due o più isotopi, distinguere gli impulsi di rimescolamento fisico da quelli di rimescolamento biologico, più costante.

Tuttavia queste interpretazioni vengono complicate dal fatto che la rideposizione fisica spesso lascia strati graduati di sedimenti, in cui le particelle più grossolane, avendo una velocità di deposizione più elevata e un'area superficiale minore, hanno una minore concentrazione di isotopi adsorbiti. Inoltre, poiché è difficile raccogliere un campione immediatamente dopo una tempesta, il rimescolamento biologico del materiale depositato di solito finisce per modificare notevolmente le tracce della tempesta avvenuta. Si può dedurre il rimescolamento biologico del sedimento dalla presenza di piccoli cumuli di escrementi e di tracce del comportamento alimentare e del movimento

strisciante degli organismi. Poiché queste attività distruggono rapidamente i depositi prodotti dal flusso della corrente, si può concludere che un aspetto morfologico del fondo (per esempio un'increspatura) registrato in una fotografia presa in profondità non è una conformazione antica, ma si è formato di recente.

A grandi profondità marine le osservazioni di ogni genere sono difficili. Gli strumenti usati per osservare le intense tempeste abissali devono essere non solo abbastanza robusti per resistere alla pressione di circa 500 atmosfere (circa mezza tonnellata per centimetro quadrato!) che esiste alla profondità di 5000 metri, ma anche abbastanza resistenti per far fronte alle notevoli forze di trazione esercitate dalle correnti abissali. Allo scopo di definire le caratteristiche delle tempeste abissali e delle loro conseguenze, i ricercatori del programma HEBBLE hanno progettato solidi strumenti in grado di misurare le seguenti variabili: la velocità del flusso e le sollecitazioni prodotte dalla corrente e che sollevano il sedimento del fondo, la concentrazione di sedimento nell'acqua, la temperatura e la salinità dell'acqua. Gli ultimi due valori ci consentono di calcolare la densità di una particolare massa d'acqua e ricavarne l'origine; mediante



La distribuzione degli isotopi radioattivi nei sedimenti del fondo indica l'entità del rimescolamento prodotto dalle correnti e dagli organismi. Le nove curve mostrano come tre isotopi caratterizzati da vita media estremamente diversa siano distribuiti secondo la profondità quando non vi è rimescolamento (a sinistra), quando gli isotopi sono soggetti a un tasso di rimescolamento tipico del mare profondo (al centro) e quando sono soggetti a un tasso di rimescolamento cento volte più elevato, come si è osservato nel sito HEBBLE (a destra). Per il torio 234 la vita media è 24,1 giorni; per il piombo 210 è 21 anni e per il carbonio 14 è 5730 anni. Le curve riportate sono basate sull'ipotesi che il sedimento si vada accumulando a una velocità di un centimetro per 1000 anni.

fotografie riprese a intervalli possiamo visualizzare il modo in cui il passaggio d'una tempesta ha alterato le caratteristiche microtopografiche del fondo. L'effetto delle tempeste sulle caratteristiche biologiche e chimiche dei sedimenti viene determinato mediante analisi di carote prelevate dal fondo.

Per misurare le velocità della corrente ci basiamo su rotori a palette simili, ma molto più robusti, agli anemometri degli aeroporti. Questi strumenti registrano internamente sia la velocità sia la direzione della corrente, mediate su intervalli di tempo scelti. Su un cavo verticale, ancorato sul fondo e mantenuto teso da galleggianti, vengono distanziati di solito sei o sette correntometri. La strumentazione si può recuperare mettendo in moto un

meccanismo di sganciamento a comando acustico. Le sollecitazioni esistenti nel fluido vengono ricavate misurando le variazioni del tempo necessario a un impulso sonoro a percorrere una distanza nota. Nel sistema progettato da Albert J. Williams III della Woods Hole Oceanographic Institution, la scansione degli impulsi sonori fornisce tutte e tre le componenti della velocità cinque volte al secondo, rendendo così possibile risolvere le fluttuazioni di velocità che rappresentano la turbolenza del flusso.

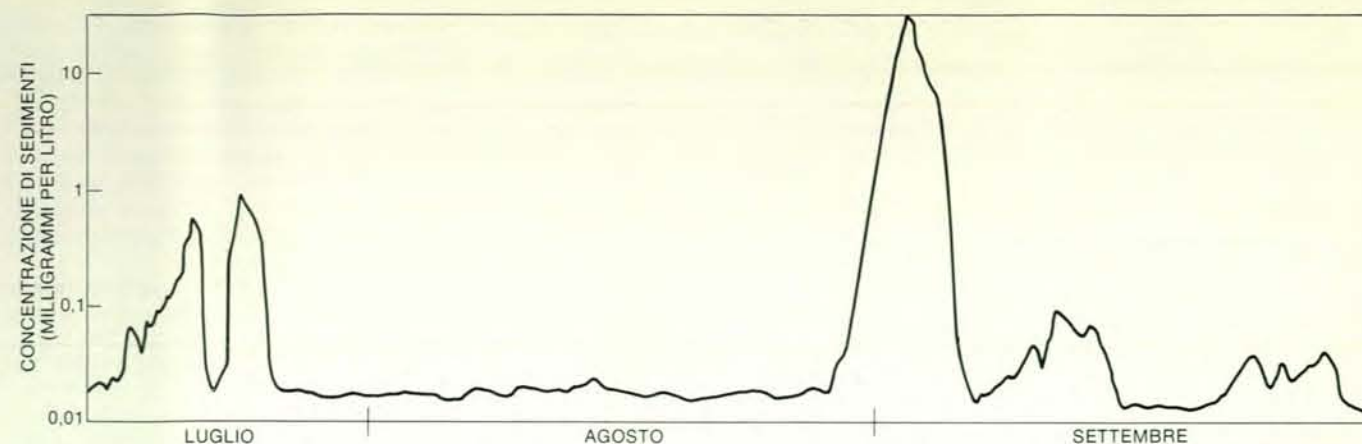
Le misurazioni di temperatura e di salinità nei pressi del fondo sono difficili perché le variazioni di queste proprietà sono minime, tipicamente inferiori a 50 millesimi di gradi centigradi per quanto riguarda la temperatura e a un millesimo di

salinità. La temperatura viene misurata con termistori attentamente calibrati e la salinità con sonde sensibili alla conducibilità elettrica dell'acqua. La concentrazione del sedimento sospeso può essere misurata direttamente prelevando campioni d'acqua in contenitori. Per misurazioni di routine della durata di giorni o settimane, ci basiamo, però, su metodi acustici e ottici. La diffusione all'indietro (*back-scattering*) di onde acustiche di alta frequenza può risolvere la struttura di nubi di sedimento, mentre la concentrazione del sedimento può essere misurata accuratamente sulla base dell'attenuazione d'un raggio luminoso.

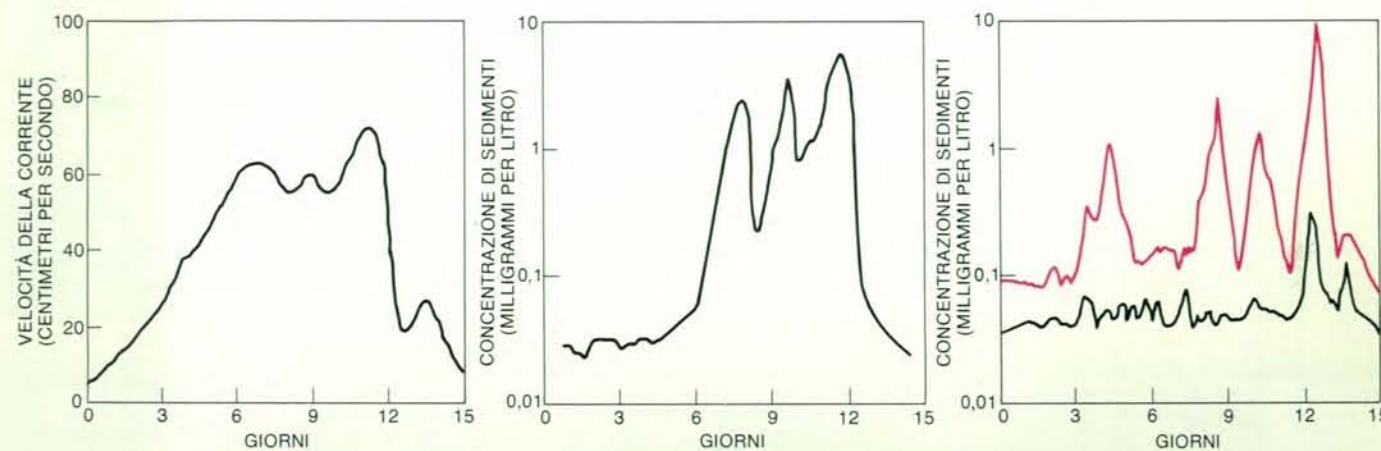
Un sistema stereoscopico di macchine fotografiche fornisce immagini particolareggiate del fondo oceanico. Aree più estese possono essere cartografate trainando un sistema sonar a scansione laterale a 50 metri dal fondo. Tra gli strumenti usati per esaminare il campo della corrente e il fondo vi è poi un velocimetro Doppler a diffrazione progettato e costruito a Woods Hole da Yogesh C. Agrawal. Questo strumento misura la velocità della corrente a 50 centimetri dal fondo in un profilo continuo, analizzando otticamente l'ambiente con un raggio laser e misurando lo spostamento di frequenza della luce riflessa all'indietro dalle particelle che si spostano passando attraverso il raggio.

Con l'aiuto di questi strumenti abbiamo cominciato dal 1979 a studiare le correnti abissali nel sito HEBBLE, situato nell'Atlantico settentrionale. Durante le tempeste abissali la velocità delle correnti può aumentare da un decimo di nodo fino a un nodo e più (ossia da circa 0,2 chilometri a 2 chilometri all'ora). Contemporaneamente la concentrazione di sedimento sospeso aumenta da 10 a 100 volte. L'elevato quantitativo di materiale sospeso assume durante una tempesta configurazioni diverse che ricordano le sfumature di grigio più o meno intenso che si osservano in un cumulonembo. La nube di sedimento sospeso in moto mostra chiazze evidentemente collegate con una durata di 20 minuti e corrispondenti a una dimensione lineare di 500 metri. Questo valore è circa 10 volte lo spessore dello strato rimescolato abissale durante una tempesta. Sulla base di esperimenti di laboratorio una scala di lunghezza di 500 metri è all'incirca la misura prevedibile per una turbolenza di grande scala nello strato limite. Durante una tempesta il livello di rimescolamento locale vicino al fondo cresce rapidamente, accompagnato da un aumento di energia turbolenta attraverso tutto lo strato abissale, mantenendo in sospensione il sedimento.

Le tempeste di cui abbiamo seguito lo svolgimento durano di solito da alcuni giorni a due settimane; alla fine di questo periodo la velocità della corrente cade rapidamente a circa cinque centimetri al secondo (circa 0,1 nodi), con una caduta simultanea del livello di sedimento sospeso. Il tasso di diminuzione della concentrazione di sedimento sospeso sta a indi-

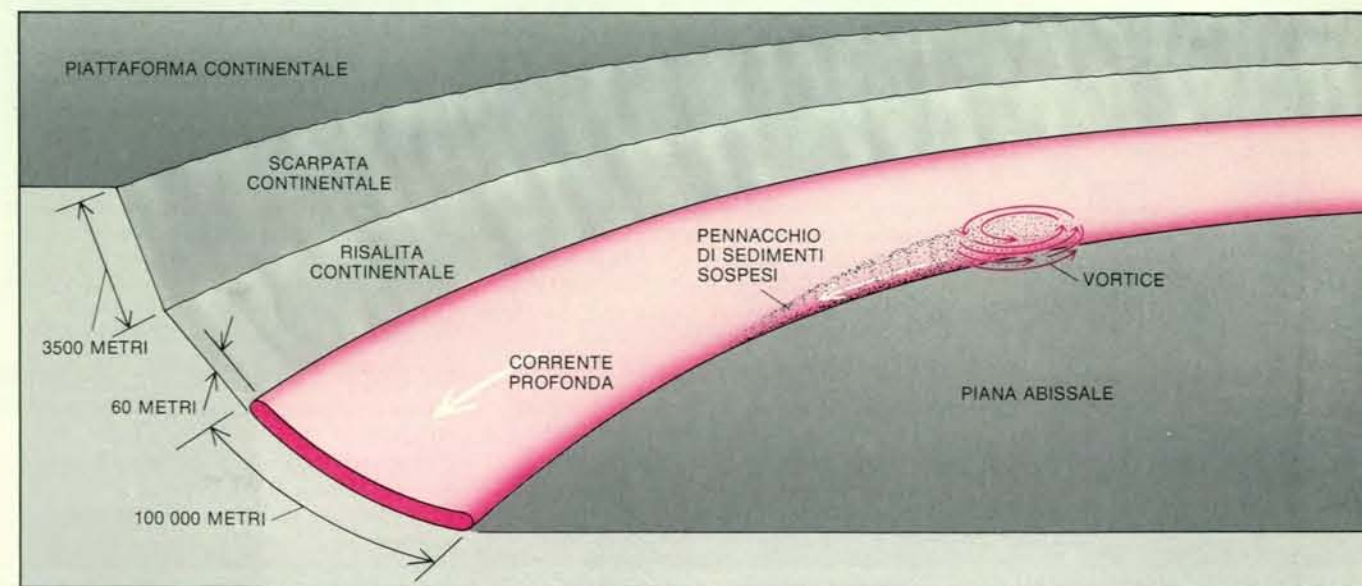


Nel sito HEBBLE la concentrazione di sedimenti è variata di un fattore 1000 circa, un metro sopra il fondo oceanico, in 10 settimane. Le massime concentrazioni hanno coinciso con violente tempeste alla base della risalita continentale a fine luglio e a inizio settembre.



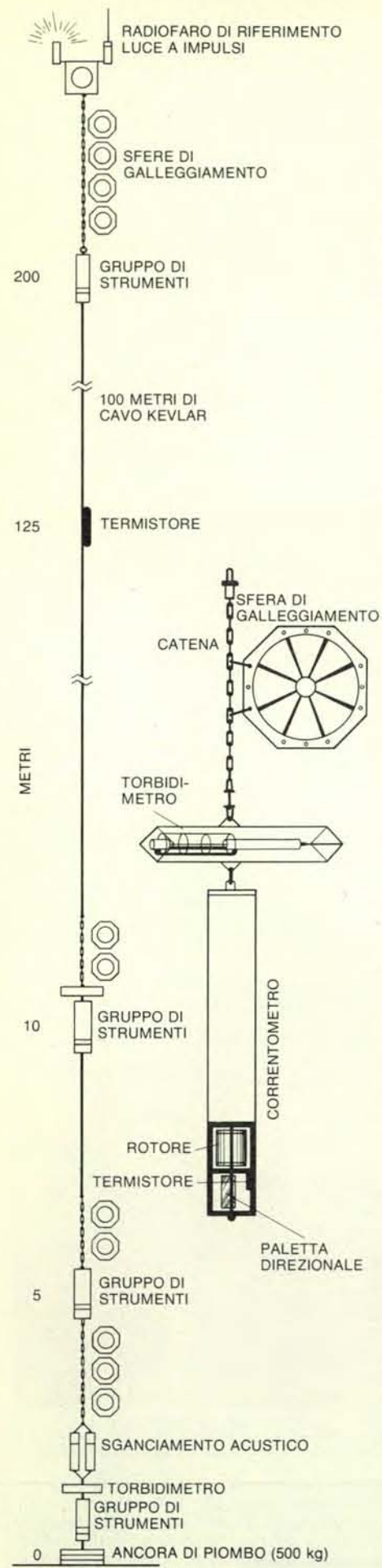
Durante una tempesta della durata di circa una settimana la velocità della corrente abissale (a sinistra) è aumentata di circa 10 volte, superando per breve tempo i 60 centimetri per secondo (1,17 nodi). Contemporaneamente la concentrazione di sedimenti nella corrente raggiunge un massimo di 100 volte il livello normale (al centro). Le

misurazioni compiute in una scala temporale breve (a destra) mostrano come la concentrazione di sedimenti fluttui rapidamente a due metri al di sopra del fondo (in nero) e più ampiamente mezzo metro al di sopra (in colore) non appena il sedimento viene sollevato dalla corrente ed è trasportato all'interno del campo d'azione dei torbidimetri.



Qui è rappresentato schematicamente un vortice di mesoscala mentre interagisce con la corrente abissale che fluisce verso sud alla base della risalita continentale, 450 miglia al largo della costa del New England. La corrente viene spinta contro la risalita dalla forza di Coriolis. Sembra che questi vortici, o tempeste abissali, siano collegati con la presen-

za in superficie di vortici di mesoscala che si dipartono dalla Corrente del Golfo. Il vortice abissale assume la forma di un'ellisse lunga circa 30 chilometri e larga forse cinque; la sua altezza non è stata determinata. L'azione combinata del vortice e della corrente abissale media erode il fondo, trascinando fango, poi ridepositato in letti sedimentari di deriva.



care che tutte le particelle, salvo le più fini, si depositano più rapidamente di quanto si potrebbe prevedere in base alla velocità individuale di sedimentazione; questo fenomeno è evidentemente provocato dalla aggregazione delle particelle presenti nella nube. L'inizio delle tempeste abissali è imprevedibile pressappoco come lo è l'arrivo di tempeste nell'atmosfera. Un dato che invece potrebbe sorprendere è che la circolazione dell'oceano profondo, a differenza della circolazione atmosferica, non mostra affatto un netto andamento stagionale. Attualmente possiamo solo dire che le tempeste abissali hanno una probabilità di avvenire ogni due o tre mesi.

La turbolenza che le tempeste generano nello strato limite dissipa energia a un tasso elevato. Il tasso a cui l'energia meccanica viene trasformata in calore raggiunge valori di un centesimo di watt per metro quadrato di fondo oceanico, valori che vengono superati solo negli estuari di scarsa profondità e sulla piattaforma continentale quando presso il fondo interagiscono ondate e correnti forti. Da dove proviene tutta questa energia?

Due sono le fonti: una è la forza pressoché costante associata alla circolazione termoalina, mentre l'altra appare collegata con i vortici che si dipartono in superficie dalla Corrente del Golfo. George L. Weatherly della Florida State University ha recentemente raccolto dati che dimostrano una forte correlazione fra la presenza di vortici a nucleo caldo che hanno origine dalla Corrente del Golfo e la presenza di tempeste a circa 5000 metri di profondità, nel sito HEBBLE. Nell'ipotesi che questi risultati possano essere estesi ad altre aree, sarebbe possibile usare carte geografiche marine che indicano la distribuzione dei vortici superficiali (deducibili mediante misurazioni fatte da satellite dell'altezza della superficie oceanica) per identificare altre regioni del fondo oceanico in cui è probabile un fenomeno di erosione attiva prodotta da tempeste.

Nel prossimo futuro, in collaborazione con Nelson G. Hogg della Woods Hole, speriamo di poter installare una schiera di correntometri sopra un'area sufficientemente estesa per valutare direttamente la frequenza e le dimensioni spaziali dei vortici che influenzano effettivamente il fondo. Attualmente alcuni studi sulla torbidità, eseguiti da J. Ronald V. Zaneveld e Hasong J. Pak dell'Oregon

Il sistema di ancoraggio delle apparecchiature impiegate per studiare le correnti abissali nell'ambito del programma HEBBLE regge una schiera di sette gruppi di strumenti che misurano velocità di flusso, temperatura e trasmissione della luce (un'indice della concentrazione di sedimenti) a diverse distanze dal fondo: meno d'un metro, cinque metri, 10, 30, 60, 125 e 200 metri. Gli ormeggi sono a distanza, l'uno dall'altro, variabile da cinque a 100 chilometri. Essi vengono recuperati facendo scattare un meccanismo acustico presso la base dell'ormeggio.

State University indicano che chiazze torbide di acqua rimescolata dalle tempeste abissali si estendono anche per 30 chilometri, una scala ragionevole per vortici oceanici. Non esistono tuttora stime relative alla loro dimensione trasversale rispetto alla corrente. È probabile tuttavia che sia inferiore ai 30 chilometri, poiché i vortici tendono a essere piuttosto allungati nel senso della corrente.

Benché il passaggio di acque torbide sia chiaramente evidente da misurazioni torbidimetriche, non è per ora possibile ricavare se per una determinata località si abbia nel breve termine un'erosione netta di materiale, una sedimentazione netta oppure una variazione trascurabile. Per esempio, un'erosione sostanziale nel breve termine potrebbe essere esattamente bilanciata dalla sedimentazione. Le sospensioni più dense di sedimento finora registrate nel programma HEBBLE hanno raggiunto la concentrazione di 12 milligrammi di solidi per litro, ossia circa 250 volte il valore tipico delle profondità oceaniche. Se si ammette che lo strato torbido sia spesso 60 metri, le particelle in sospensione produrrebbero uno strato sedimentario spesso mezzo centimetro se venissero improvvisamente e uniformemente depositate. Considerando il fenomeno sotto un'angolazione diversa, se si valuta che la larghezza della corrente torbida sia di cinque chilometri, il quantitativo di sedimento trasportato per minuto attraverso una sezione trasversale del percorso della tempesta supererebbe i sei tonnellate.

Se la tempesta avesse realmente scaricato uno strato nuovo di sedimento spesso mezzo centimetro nel sito HEBBLE, avrebbe eguagliato il quantitativo netto depositato normalmente in 500 anni, secondo studi compiuti con isotopi radioattivi a lunga vita. Un tale tasso di sedimentazione è in accordo con altre stime fatte sulla sedimentazione lungo la risalita continentale. Tuttavia, quando i sedimenti del sito HEBBLE vengono esaminati relativamente a isotopi di vita breve, emerge un quadro piuttosto diverso. Il cesio 137, generato dalle esplosioni atomiche nell'atmosfera solo 30 anni fa, si è già mescolato ai sedimenti anche a profondità di 12 centimetri. Il torio 234, con una vita media di soli 24,1 giorni, si può evidenziare a sette centimetri sotto la superficie del fondo.

In questi studi, eseguiti da David J. DeMaster e da Charles A. Nittrouer della North Carolina State University, si è avanzata l'ipotesi che il tasso di rimescolamento sia circa 100 volte superiore, nel sito HEBBLE, al valore medio di aree tipiche abissali, non interessate da tempeste. Evidentemente le tempeste abissali e gli organismi che vivono sul fondo agiscono insieme per rimescolare intensamente i sedimenti che si accumulano gradualmente. Il contributo relativo dei due fattori non è ancora stato calcolato, ma secondo una stima preliminare le tempeste hanno una maggiore influenza nel rimescolamento rispetto agli organismi. Sembra che gli organismi siano abbastanza seden-

tari, ben adattati a raccogliere sostanze nutritive dal flusso che passa, piuttosto che a scavare in cerca di cibo.

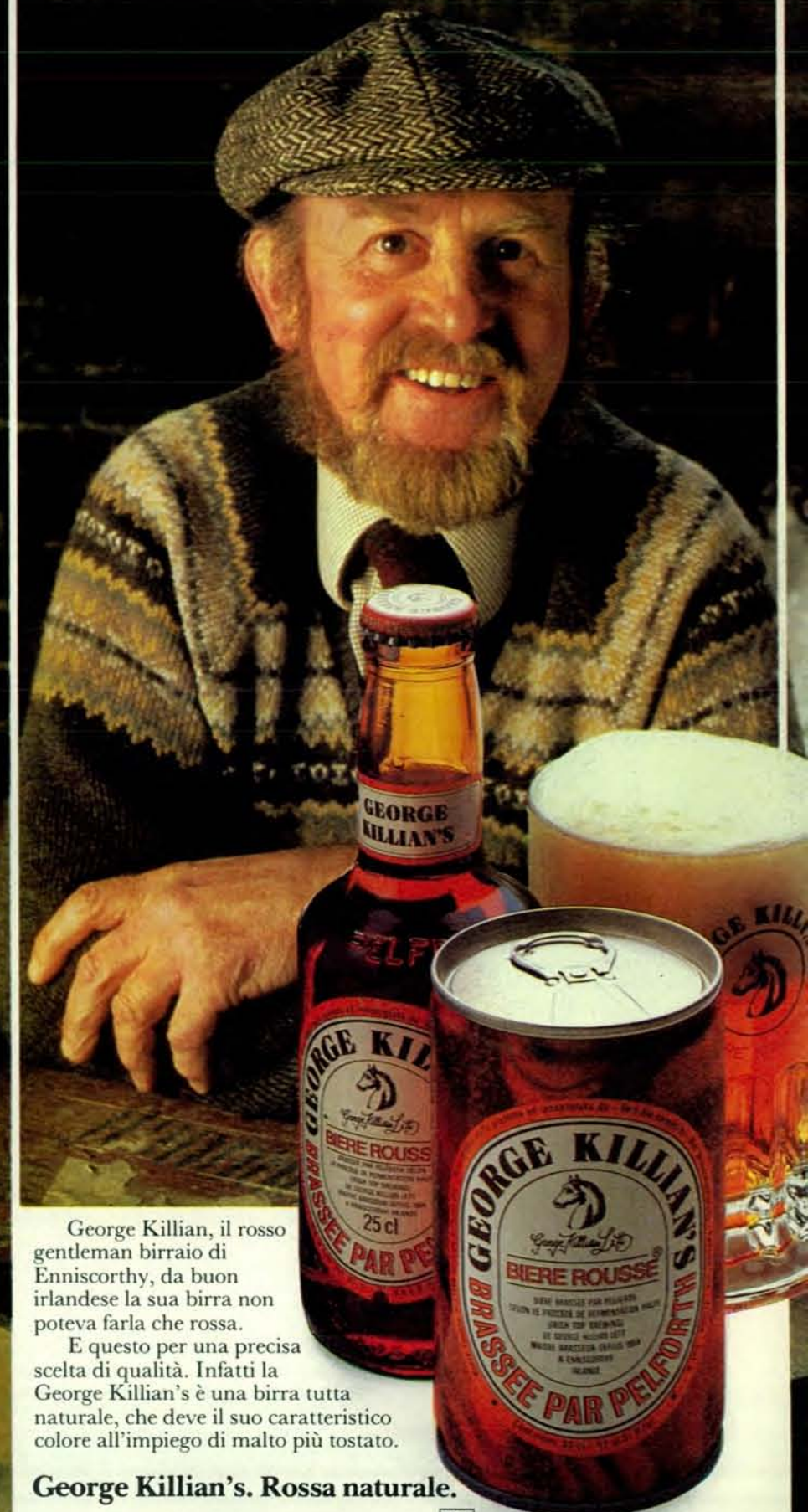
Finora il programma HEBBLE si è basato principalmente su osservazioni anziché su esperimenti. Il quadro provvisorio delle relazioni tra intensità del flusso, erosione e deposizione deve essere sottoposto a verifica sperimentale prima di poter essere sostenuto con una certa sicurezza o sostituito con un quadro più accurato. Attualmente le caratteristiche dell'erosione e della deposizione sono conosciute con un certo grado di sicurezza solo per le sabbie d'acqua poco profonda. Il comportamento dei letti fangosi abissali rispetto al flusso dell'acqua sovrastante risente sia dalle sollecitazioni imposte dalle correnti sia dalle modificazioni indotte dagli organismi viventi sulla loro microtopografia e sulla loro adesività. Poiché il fenomeno non è riconducibile a una simulazione di laboratorio, è necessario un esperimento in loco: solo così si può stabilire la complicata interconnessione tra intensità del flusso, tasso di erosione, profondità di erosione, tasso di deposizione e distribuzione degli isotopi radioattivi.

A questo scopo uno dei nostri colleghi della Woods Hole, Clifford L. Winget, ha progettato uno strumento speciale che, immerso sul fondo, può imporre sollecitazioni note e controllate sui sedimenti. Il dispositivo fa circolare un campione di acqua abissale a velocità varie, ma costanti, attraverso una sezione di fondo marino, esposta ma ben delimitata. Le sollecitazioni del fluido e il numero delle particelle raccolte dal fluido circolante sono misurate da uno speciale correntometro a laser progettato in modo da registrare lo spostamento Doppler nella luce diffusa da particelle sospese che si muovono assieme al flusso. L'entrata in attività di questo strumento, chiamato SeaDuct, è programmata per quest'anno.

La scoperta che le correnti abissali possono raccogliere e ridistribuire i sedimenti sta inducendo i geologi a riesaminare le idee a proposito delle località dove cercare nuove riserve di petrolio. Da tempo si sa che le grandi concentrazioni di petrolio dipendono dalla presenza di certe condizioni geologiche: una fonte sedimentaria, una roccia porosa che possa funzionare da serbatoio e alcuni tipi di barriera, o di struttura confinante. La fonte è fornita da carbonio organico in sedimenti a granulometria fine, ricchi di carbonio. Il serbatoio è spesso fornito da rocce sedimentarie porose e permeabili, come l'arenaria. Svariate strutture geologiche funzionano infine da barriere impermeabili per racchiudere il petrolio nel serbatoio sotterraneo.

Secondo le teorie convenzionali, la forza principale che agisce sui sedimenti è la gravità. Da tempo si suppone che tutti i sedimenti si muovano inesorabilmente verso il basso. Benché questa convinzione rimanga probabilmente vera per la sabbia grossolana e per la ghiaia, ora pare evidente che il materiale a granulometria più

George Killian l'irlandese vi invita alla sua birra rossa.

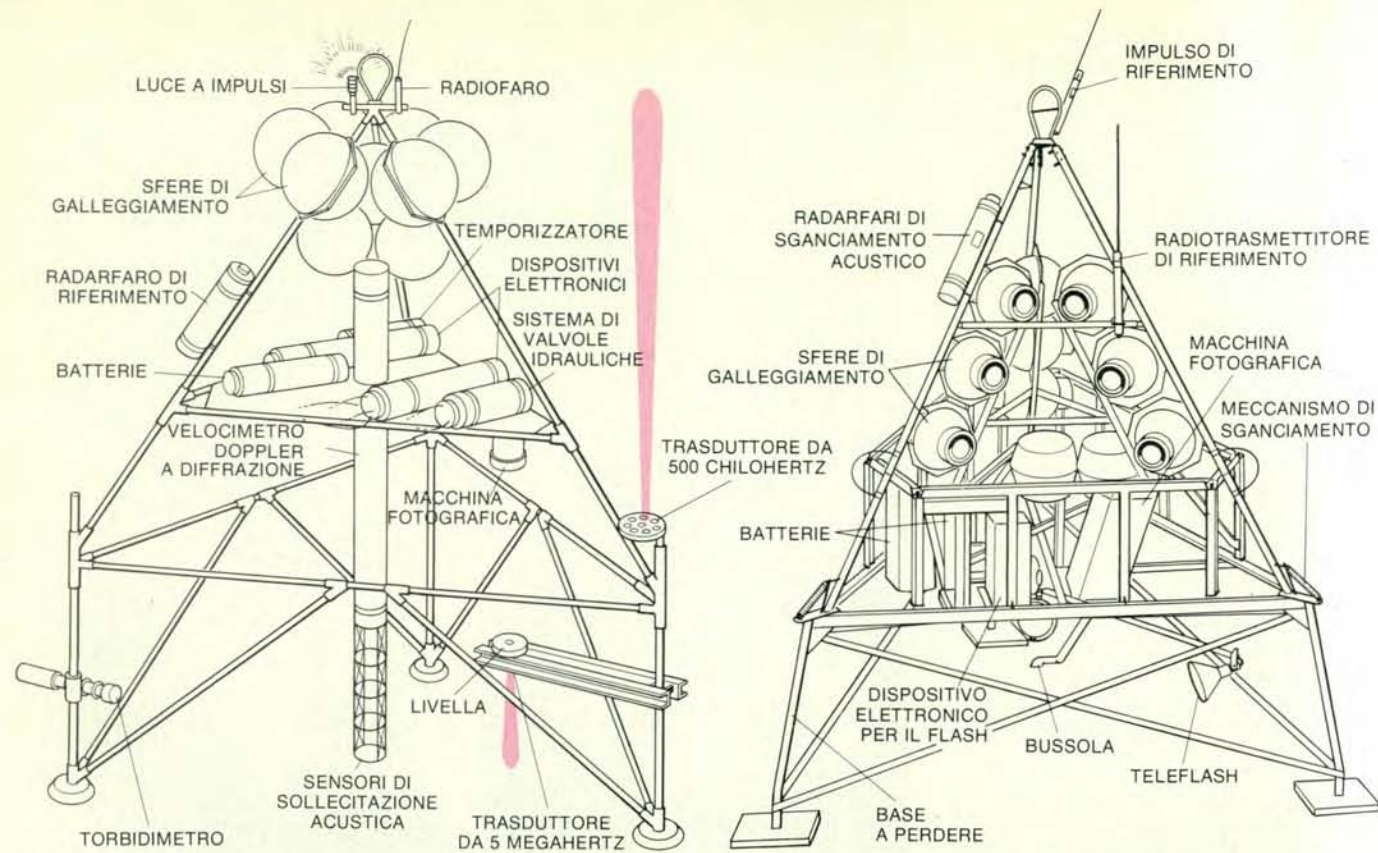


George Killian, il rosso gentleman birraio di Enniscorthy, da buon irlandese la sua birra non poteva farla che rossa.

È questo per una precisa scelta di qualità. Infatti la George Killian's è una birra tutta naturale, che deve il suo caratteristico colore all'impiego di malto più tostato.

George Killian's. Rossa naturale.

Birra prodotta dalla Pelforth su ricetta George Killian. Distribuita da Wax & Vitale. Linea Wax.



I ricercatori del programma HEBBLE hanno messo a punto due tipi di treppiede per strumenti destinati allo studio particolareggiato delle correnti abissali e della topografia del fondo oceanico. Il treppiede raffigurato a sinistra porta tre sistemi di misurazione. I sensori di sollecitazione acustica determinano la variazione temporale e spaziale di velocità e la sollecitazione turbolenta nella corrente abissale, misurando le variazioni nel tempo necessario a un impulso sonoro per percorrere una distanza nota. Un velocimetro Doppler a diffrazione

determina anche la velocità del flusso misurando lo spostamento Doppler nella luce riflessa dalle particelle in movimento. Un terzo sistema misura la concentrazione di sedimenti sospesi registrando le onde sonore diffuse all'indietro da onde sonore emesse a due frequenze: cinque megahertz e 500 chilohertz. Il secondo tipo di treppiede (a destra) porta una macchina fotografica stereoscopica ad alta risoluzione programmata per riprendere immagini ogni 30 minuti. Entrambi i treppiedi sono riportati in superficie per mezzo d'un comando acustico.

fine può essere trasportato a distanze significative dalle correnti abissali. Presumibilmente la maggior parte delle sostanze organiche in grado di formare petrolio è concentrata nel materiale particellato fine che ha origine principalmente nelle acque superficiali. Da quando questo materiale si è depositato sul fondo lungo il lato occidentale degli antichi bacini oceanici, potrebbe essere stato trasportato parallelamente ai contorni del bacino, ossia ad angolo retto rispetto alla direzione comunemente presunta.

Perché il materiale organico si trasformi in petrolio, si pensa sia necessaria almeno una delle due condizioni seguenti: il tasso di accumulo deve essere talmente elevato oppure la concentrazione di ossigeno nelle acque abissali deve essere talmente bassa che il materiale non possa essere ossidato prima di venir sepolto. Se i geologi che si occupano dei giacimenti petroliferi potessero sapere quali sono le località degli antichi mari in cui prevalse una, o meglio prevalsero entrambe queste condizioni, potrebbero anche scoprire più facilmente nuovi depositi di petrolio. Perciò le aree in cui le tempeste abissali hanno prodotto tassi elevati di accumulo di sedimenti ricchi di materiale organico dovrebbero essere siti favorevoli alla

formazione di rocce petrolifere. Di conseguenza i geologi potrebbero cercare a valle delle antiche scarpate le rocce-serbatoio, a grana grossa, e a valle delle paleocorrenti i depositi di sedimenti ricchi di carbonio.

Le nuove conoscenze acquisite sulle tempeste abissali hanno immediate implicazioni pratiche per i tecnici responsabili della stabilità e della sicurezza dei cavi e delle strutture situate sul fondo oceanico. Nelle fotografie sottomarine si è potuto osservare come le correnti siano in grado di asportare i sedimenti dalla zona sottostante ai cavi sottomarini e di lasciare i cavi stessi sospesi tra aree non erose del fondo. In alcuni casi le conseguenti vibrazioni del cavo ne hanno provocato la rottura.

L'esistenza di tempeste abissali presenta anche nuovi problemi per coloro che sono interessati alle strategie antisommergibile. Il rumore associato a grandi tempeste turbolente può confondere gli strumenti di ascolto passivo sistemati sul fondo per rilevare il passaggio di sommergibili. Sarebbe addirittura possibile far sì che il rumore dei sommergibili fosse analogo a quello delle tempeste, rendendo estremamente difficile rilevare la pre-

senza di queste unità in regioni di elevata turbolenza come l'Atlantico settentrionale occidentale.

Forse però la questione più importante sollevata dall'esistenza delle tempeste abissali è il loro effetto potenziale sui rifiuti tossici che possono essere depositati sul fondo marino. Prima di scegliere il fondo oceanico come il luogo definitivo per liberarsi di rifiuti pericolosi, è indispensabile disporre di previsioni attendibili sulle condizioni a lungo termine della località sottomarina prescelta. Se il metodo di eliminazione dei rifiuti si basa sul confinamento, il luogo di scarico dovrebbe essere una zona dove avviene un trascurabile trasporto di sedimenti. D'altra parte, se il metodo prescelto è quello di diluire e disperdere il materiale di rifiuto, sarebbe anzi meglio scegliere aree soggette a tempeste. In ogni caso però si dovrebbe essere consapevoli che alcuni rifiuti tossici vengono facilmente adsorbiti sulla superficie delle minute particelle sospese e potrebbero perciò trovare facile entrata nella rete alimentare degli organismi che si cibano sul fondo. Queste sono solo alcune delle questioni sollevate dal nostro studio sulle correnti e sulle tempeste abissali, compiuto in un remoto angolo di un oceano.

I «dischi da getto», armi femminili dell'antico Lazio

Secondo recenti studi, nell'Età del ferro, tra il IX e il VII secolo a.C., le donne di comunità agricolo-pastorali usavano in combattimento dischi di bronzo affilati e taglienti, ritenuti in precedenza oggetti di ornamento

di Rita Cosentino e Guido Devoto

Si dalla più remota antichità l'uomo ha indirizzato spesso la propria genialità creativa nel concepire e realizzare armi sempre più micidiali, perfezionate e insidiose per sopraffare e uccidere i propri simili.

Le armi tradizionali in asta, da botta e da getto, quali la lancia e l'ascia, l'alabarda e il martello d'arme, il giavelotto e la mazza, per quanto di origine contadina, sono armi «nobili», permeate di spirito cavalleresco; così come tra le armi da punta e da taglio la spada fu sempre avvolta di sacralità. Al contrario il pugnale e la daga, solo per il fatto di essere armi corte, che si potevano facilmente nascondere indosso, sotto le vesti, negli stivali, persino in seno, ebbero il marchio di armi «insidiose». Non per caso quel particolare tipo di daga che nei duelli del XV e XVI secolo si usava in appoggio alla spada o striscia fu denominato «mano sinistra», non soltanto perché la daga veniva impugnata proprio con quella mano, mancini esclusi, ma per sottolinearne con l'appellativo «sinistro» il carattere di pericolosa insidiosità e di funesta, bieca, minacciosa, traditrice potenzialità offensiva.

Tra le armi cosiddette manesche da lancio, l'arco e la balestra ebbero fama di strumenti scellerati e insidiosi; la spada era arma onorevole, con la quale un duello o «singolar tenzone» si risolveva in favore del più forte o del più abile. La freccia o il bolzo di balestra volavano alla cieca e colpivano a distanza di sorpresa il cavaliere corazzato.

Tuttavia la storia delle armi è ricca di particolari strumenti d'offesa «insidiosi», di forma, struttura e foggia inusuali o addirittura strani: dal brandistocco o buttafuori del XV secolo a lama rientrante nel manico, che trasformava una scure in un acuminato spiedo, ai coltelli genovesi scanalati a doccia e forati in punta per simulare una lesina per cuoio, agli spiedi con lame ripiegabili, ai martelli d'arme che nascondevano nel fusto una pistola a pietra focaia. Ma forse la fantasia creativa

delle popolazioni orientali nel progettare e realizzare i più singolari strumenti di morte non fu mai superata.

Nacque così in Indonesia quello straordinario oggetto d'arte che è il *kriss*, arma micidiale ma anche, e soprattutto, espressione culturale e amuleto, scultura e simbolo religioso, esibizione tecnologica e strumento di magia. In India, il famoso pugnale di corte Moghul con impugnatura a «staffa», il *katar*, era un simbolo gerarchico di particolare ceto e levatura sociale, ma anche una micidiale daga da usarsi esclusivamente di punta come «sfondagiaco» o come un pugno «diretto» verso la gola scoperta dell'avversario. Tra le armi indiane più fantasiose bisogna ricordare il *bagh nakh* (letteralmente, «artiglio di tigre»), una specie di tirapugni munito di numerose punte a unghia ricurve e taglienti; arma preferita da assassini professionisti, produceva ferite del tutto simili a quelle di una zampata di tigre.

L'ingegnosità creativa degli armaioli indiani raggiunge forse il suo vertice in un'arma da getto assolutamente fuori del comune, il *chacarani* o *chakram*, disco anulare appiattito in acciaio forgiato, con il bordo esterno affilato come un rasoio. Lo studio di quest'arma insidiosa straordinaria ci ha stimolato a prendere in esame per un possibile confronto comparativo i cosiddetti «anelli da sospensione», diffusi nell'Italia peninsulare durante l'Età del ferro.

Gli anelli bronzei definiti «da sospensione», diffusi nelle necropoli dell'area laziale, falisco-capenate, in Etruria (Cerveteri, località Sorbo) e nell'Italia centromeridionale (Terni, Cuma, Pontecagnano ecc.) sono dischi anulari rinvenuti sempre in corredi riferibili a sepolture femminili e ritenuti elementi di abbigliamento con scopi semplicemente decorativo-funerari. Tali manufatti, compresi in un arco cronologico dell'Età del ferro che va dal II periodo laziale (IX secolo a.C.) sino agli inizi del IV periodo (fine VIII-

-inizi VII secolo a.C.), venivano inseriti sull'ardiglione di fibule bronzee o di altro materiale, agganciate a loro volta, come spille di sicurezza, sull'abito della defunta, all'altezza delle spalle, del petto o del ventre. Ogni fibula poteva inoltre sostenere, come hanno dimostrato i rinvenimenti, più anelli disposti in serie dimensionale decrescente.

Prima di procedere all'analisi delle caratteristiche tecnologico-costruttive di questa classe di manufatti, si vuole sottolineare quale dato significativo, con un rapido *excursus* dei corredi delle necropoli ove tali oggetti compaiono, la loro presenza in sepolture esclusivamente femminili. Tali sepolture sono caratterizzate e distinte da quelle maschili per peculiarità patenti e nettissime. Nella necropoli di Castel di Decima (Roma), costituita da sepolture a fossa di dimensioni e profondità variabili, riempite da grandi scaglie di tufo dopo la collocazione del defunto e con orientamento di massima est-ovest, i corredi maschili sono contraddistinti dalla presenza di fibule «a drago» e da armi tipiche (lance singole o raggruppate con altre, talora soltanto una spada). I corredi femminili evidenziano invece fibule «a navicella», fuseruole in gran quantità e spirali per capelli; in alcuni corredi di questa necropoli compaiono «anelli da sospensione», per esempio nelle tombe CXXXIII e III. Tra i corredi è particolarmente degno di nota quello della tomba CLIII sia per la quantità sia per la qualità degli oggetti (circa 90 pezzi tra materiali fittili di grandi e piccole dimensioni, bronzi, collane di vaghi d'ambra, paste vitree, pendenti configurati a scarabei egittizzanti). In questo corredo compare un «anello da sospensione» di eccezionale diametro (circa 350 millimetri), agganciato a una fibula bronzea situata sul ventre della defunta.

Nella necropoli dell'Osteria dell'Osa, dove compaiono tombe sia a inumazione a fossa che a cremazione (costituite da un pozzetto piuttosto profondo contenente

un dolio, entro cui erano deposte l'urna cineraria con un corredo ceramico e bronzo miniaturizzato), le sepolture femminili sono caratterizzate dalla presenza di una o più fibule bronzee, ad arco ingrossato e staffa simmetrica, vaghi d'ambra, paste vitree, fuseruole, rocchetti (oggetti legati probabilmente ad attività di filatura e tessitura) e talora da spirali bronzee, interpretate come fermatrecce, in quanto rinvenute invariabilmente ai due lati del cranio della defunta.

Come per la necropoli di Castel di Decima, non tutti i corredi riferibili a sepolture femminili hanno restituito «anelli da sospensione», tuttavia in quelli dove tali oggetti sono stati rinvenuti, si è notata una maggiore ricchezza rispetto ai corredi coevi che ne sono privi. A tale proposito, basterebbe ricordare il corredo della tomba a fossa n. 117, datato alla seconda fase laziale A (prima metà del IX secolo a.C.), costituito da circa 18 oggetti tra cui ceramica, più abbondante e accurata del solito, fibule, collanine di dischetti di faïence, anelli fermatrecce; si tratta di un corredo decisamente ricchissimo in riferimento sia al momento cronologico che a deposizioni coeve.

Sarebbe lecito ipotizzare, nell'attuale stato di conoscenze e di studio, che gli «anelli da sospensione» fossero appannaggio solo di quelle donne che potessero rivestire, per status o condizione gerarchica, determinati ruoli all'interno della comunità di appartenenza.

Nella necropoli di Tivoli, costituita da tombe «a circolo», cioè delimitate da lastre di travertino o da una o più file di pietre disposte in cerchio e contenenti all'interno sepolture a profondità variabile (talora segnalate da un cippo-segnacolo posto in un'area del circolo risparmiata dalla copertura di strati di materiale litico), i corredi femminili sono caratterizzati dalla presenza di fuseruole, fibule e oggetti d'ornamento personali quali collane di pasta vitrea e ambra, orecchini, anelli e talora spilloni. I corredi maschili sono contraddistinti da pugnali, lance, talora da coltelli in ferro. Anche in questa necropoli non tutte le deposizioni femminili hanno restituito «anelli da sospensione»; tra quelle dove tali oggetti compaiono, va evidenziata per la ricchezza del corredo la tomba XLIII, datata tra la fine del II e l'inizio del III periodo laziale.

Discorso analogo vale per la necropoli della Riserva del Truglio, per la distinzione dei corredi maschili da quelli femminili, possibile per la presenza di elementi guida: armi per le deposizioni maschili e strumenti legati ad attività muliebri per quelle femminili. Degna di nota la ricchezza del corredo della tomba XXX della necropoli, comprensiva di «anelli da sospensione» di diametro cospicuo. Questa sepoltura contiene infatti, il corredo più ricco della necropoli per la presenza di grandi quantità di bronzi, ambre e paste vitree; nell'ambito di una comunità piuttosto povera, questo elemento potrebbe evidenziare, come già è stato sostenuto, sia un processo di differenziazione sociale che il ruolo particolare della defunta.

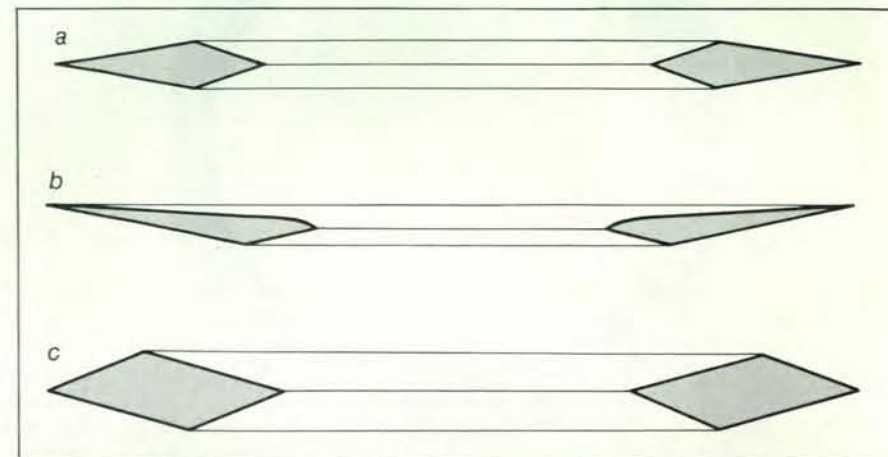


Questo «disco da getto» in bronzo, il cui diametro reale è di circa 83 millimetri, è caratterizzato da un notevole spessore e da una sezione a «parallelogramma» appiattito (si veda nell'illustrazione in fondo alla pagina). Tra gli esemplari studiati è il più massiccio e di esecuzione meno accurata. Lo stesso «disco da getto» è visto, in basso, di profilo. Il leggero ingrandimento (1,5 volte) mette in evidenza la sezione a «parallelogramma» e la fusione irregolare poco curata e mal rifinita.

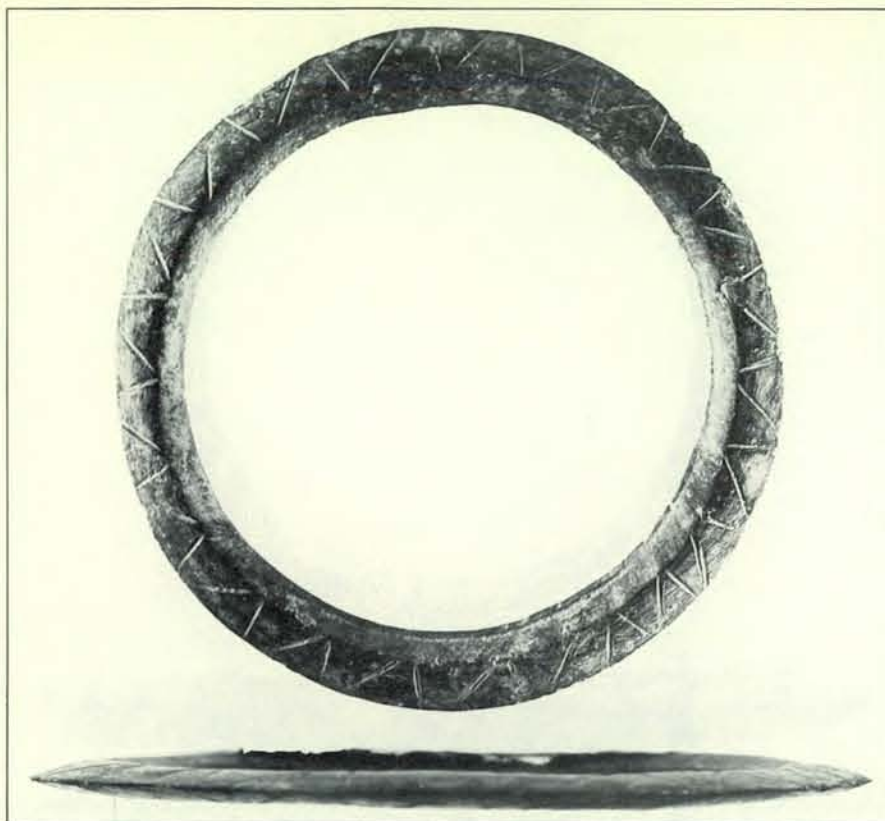
Elementi distintivi caratterizzano i corredi femminili e maschili delle sepolture delle necropoli di Caracupa, Anzio, Satrium, La Rustica e dell'Esquilino, come del resto di altre che non interessano ai fini di questo studio; in queste necropoli sono stati rinvenuti «anelli da sospensione» che potrebbero essere riferiti non a

re femminili, ipotesi che deve essere confermata da indagini più approfondite.

L'analisi tecnologico-costruttiva dei manufatti ha messo in luce che si tratta, più che di anelli, di «dischi anulari» in bronzo appiattiti, con sezione definita di solito «romboideale» o «quasi piatta» spesso arricchiti da una decorazione inci-



Il tipo di sezione che si riscontra con maggior frequenza nei «dischi da getto» è quello a due triangoli isosceli accoppiati per la base (a). Un altro tipo, caratteristico dei dischi anulari più appiattiti e, in realtà, più raffinati, è un vero e proprio profilo «alare» aerodinamico (b). Assai diffuso è anche un tipo di sezione a «parallelogramma» appiattito (c). L'analisi microchimica compiuta su tre dischi aventi le tre sezioni riportate qui ha dato una composizione analoga con piccole variazioni nel contenuto di rame (rispettivamente, dall'alto in basso, 70, 75, 80 per cento), di piombo (25, 20, 15 per cento) e lo stesso contenuto di stagno (circa 5 per cento).



Questo «disco da getto» in bronzo molto appiattito ha una sezione a triangoli isosceli contrapposti (si veda a nell'illustrazione in basso della pagina precedente). Si nota in questo esemplare una notevole accuratezza di esecuzione e di finitura e, a tratti, sul margine esterno sono visibili le striature da «affilatura», mentre il motivo ornamentale è rappresentato da brevi solchi incisi a zig-zag. Il medesimo «disco da getto», il cui diametro reale è di 90 millimetri circa, è visto in basso di profilo e leggermente ingrandito. Il margine tagliente acuto è nettamente profilato e si nota la regolarità e l'accuratezza della sezione esterna a triangolo isoscele.



Questo «disco da getto» di piccole dimensioni (circa 80 millimetri di diametro) con sezione a triangoli isosceli contrapposti e rotto in tre frammenti presenta una marcata «usura» differenziata asimmetrica. In alto a destra, la larghezza della fascia anulare è ridotta a soli sette millimetri, rispetto agli 11-12 millimetri della restante circonferenza. Il fatto, inoltre, che il margine abbia assunto in un breve tratto un andamento quasi rettilineo può far supporre, in questo caso, un impiego del piccolo disco più come utensile di uso domestico (raschiatoio per pelli?) che come arma.

sa a zig-zag sulle superfici esterne. Il diametro esterno dei dischi varia notevolmente ed è compreso, nell'ambito degli esemplari analizzati e misurati, tra 235 e 30 millimetri; eccezionalmente, come abbiamo già fatto osservare in precedenza, il diametro esterno può raggiungere i 350 millimetri, ad esempio negli esemplari di La Rustica, Satricum e Castel di Decima. Meno variabile è invece il diametro interno in rapporto alla dimensione totale del disco; ciò comporta per la larghezza della fascia discoidale anulare valori molto più costanti, compresi tra 17 e 19 millimetri. Solo in rari casi (si veda l'illustrazione nella pagina a fronte in alto) la larghezza della fascia anulare raggiunge i 23-25 millimetri. Il peso varia notevolmente in funzione del diametro e dello spessore; in media, per esemplari di diametro esterno compreso tra 200 e 90 millimetri, il peso oscilla tra 210 e 45 grammi.

L'analisi geometrica delle sezioni trasversali tipiche dei dischi rivela alcune caratteristiche peculiari. In primo luogo si nota che solo raramente la sezione trasversale può essere assimilata a quella figura geometrica definita rombo (il rombo è una figura piana equilatera con angoli adiacenti supplementari). La sezione che si riscontra più frequentemente è rappresentata da due triangoli isosceli accoppiati per la base; il triangolo che forma il margine esterno del disco ha una altezza spesso quasi doppia di quello interno. Un secondo tipo di sezione assai diffuso è un parallelogramma appiattito (parallelogramma definito come una figura geometrica piana con i lati a due a due uguali e angoli adiacenti supplementari). Un terzo tipo di sezione è caratteristico dei dischi anulari più appiattiti (in realtà più perfezionati) ed è definibile come un vero e proprio profilo alare portante aerodinamico (profilo che si ottiene sezionando ortogonalmente un'ala di aeroplano o di aliante); la sezione è praticamente riconducibile a un triangolo scaleno estremamente appiattito, con lati appena leggermente curvi.

È evidente che questi diversi tipi di sezioni e di profili intenzionalmente concepiti e realizzati conferiscono all'oggetto una precisa funzionalità; l'analisi metallografica microstrutturale e macrografica (eseguita su quattro esemplari completi disponibili e su alcuni frammenti di altri) rivela inoltre che la lega bronzea adoperata ha subito un processo di incrudimento meccanico per martellatura, lungo i margini esterni del disco anulare. È noto che i metalli duttili non ferrosi, ai quali è impossibile conferire una tempra per raffreddamento brusco o per normalizzazione, aumentano la loro resistenza meccanica e la loro durezza per incrudimento a freddo. In particolare alcuni bronzi possono raggiungere con questo trattamento a martellatura la durezza dell'acciaio dolce. Nel caso in esame una misura di durezza «attuale» non avrebbe alcun senso, in quanto la mineralizzazione spinta della lega e soprattutto la sua ricristallizzazione interna per «invecchiamento» naturale

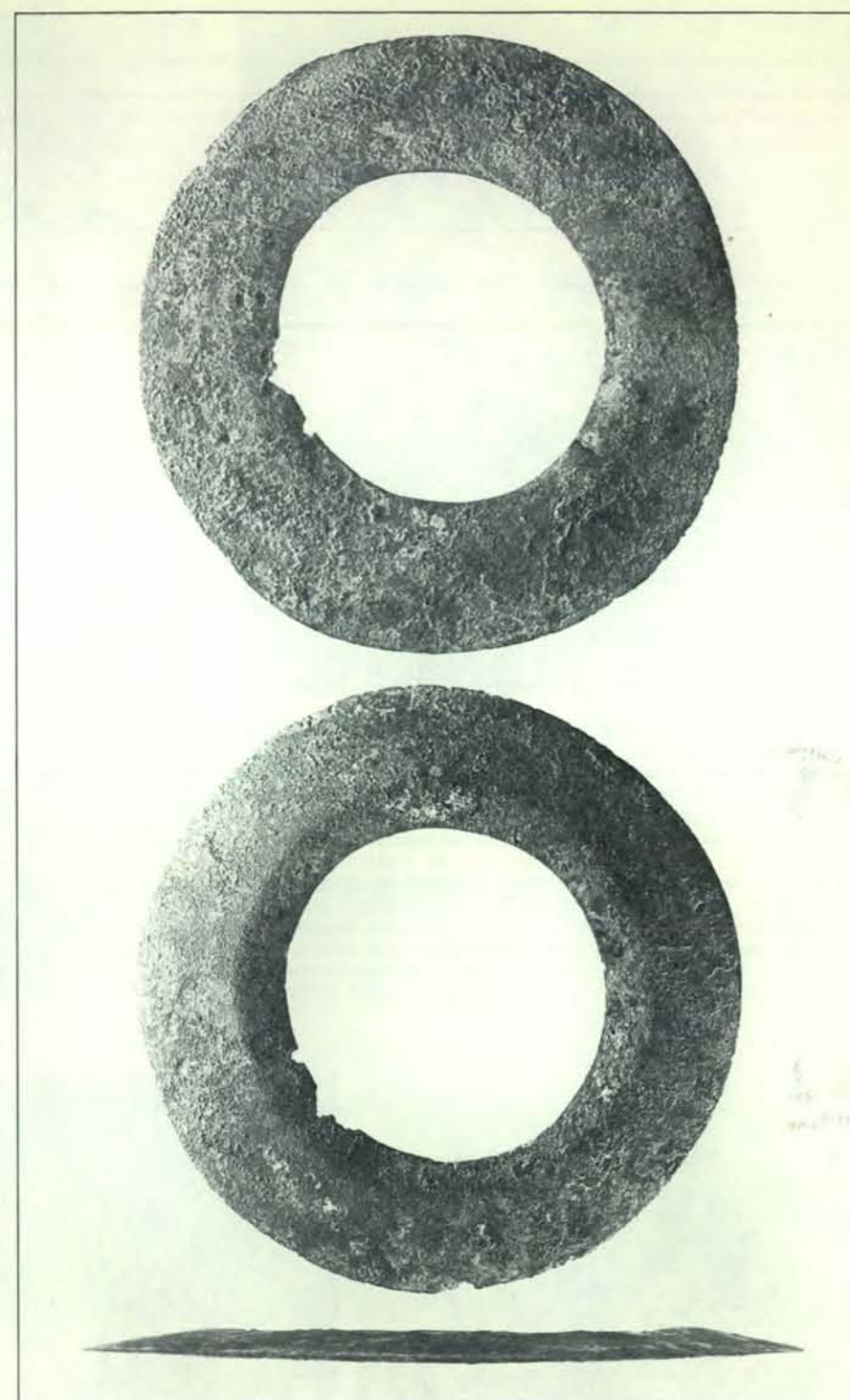
ha completamente alterato questo parametro fisico. All'indurimento meccanico è seguita una levigatura per abrasione e una affilatura a tagliente che ha prodotto un filo di taglio esterno identico a quello di una lama di coltello o di spada.

Tale affilatura, eseguita certamente a mano con una pietra da coti, è documentata microscopicamente dalla presenza di sottili solchi subparalleli al margine esterno del disco o inclinati, in modo caratteristico, di circa 10-20 gradi rispetto al bordo tagliente (si veda l'illustrazione a pagina 39). Questa inclinazione dei solchi è pressoché costante sulle lame di falci, roncole o coltelli affilati a mano con un caratteristico movimento di «va e vieni» longitudinale della coti sui due lati del tagliente. Il margine interno dei dischi mette invece costantemente in luce solchi più profondi, grossolani e spaziosi, quasi normali alla circonferenza, ottenuti molto probabilmente con l'uso della lima nel corso della rifinitura del disco dopo la sua fusione.

Una serie di analisi microchimiche, eseguite presso il Centro di studio per la geologia dell'Italia centrale (Istituto di geologia e paleontologia dell'Università di Roma), ha rivelato inoltre alcune caratteristiche composizionali della lega metallica estremamente interessanti. I campioni, ripuliti dalle patine di ossidazione, sono stati osservati al microscopio a scansione munito di rivelatore di raggi X a stato solido (SEM-EDS). La microanalisi, non quantitativa, ma comunque idonea a fornire una stima visiva percentuale, ha permesso l'identificazione degli elementi costituenti le leghe adoperate e la loro caratterizzazione. I risultati, relativi a tre dischi anulari di sezione differente - a triangoli isosceli contrapposti, appiattiti «alare», a parallelogramma appiattito - hanno mostrato come i tre dischi abbiano composizione analoga con piccole variazioni nel contenuto in rame (dal 70 all'80 per cento) e in piombo (dal 15 al 25 per cento) e lo stesso contenuto in stagno (5 per cento). Si tratta quindi di una lega bronzea molto particolare, caratterizzata da una elevatissima percentuale di piombo, intenzionalmente aggiunto. È logico supporre che tale caratterizzazione dovesse rispondere, oltre alla migliore fusibilità della lega, alle particolari esigenze di impiego di un'arma da getto di ridotte dimensioni e peso, che doveva associare al modesto volume il massimo peso specifico, per conservare a distanze utili una sufficiente energia balistica offensiva. Ciò vale ai giorni nostri per i proiettili ad alta velocità con nucleo di piombo fuso o pressofuso, incamicciato di rame o di acciaio.

Si può dunque ragionevolmente credere che i cosiddetti «anelli da sospensione» rappresentino una speciale categoria di armi bianche da getto (almeno gli esemplari di maggior peso e dimensioni), di specifico impiego femminile. Per essi proponiamo di adottare il nome di «dischi da getto».

Il disco anulare tagliente, opportunamente profilato, se maneggiato con de-

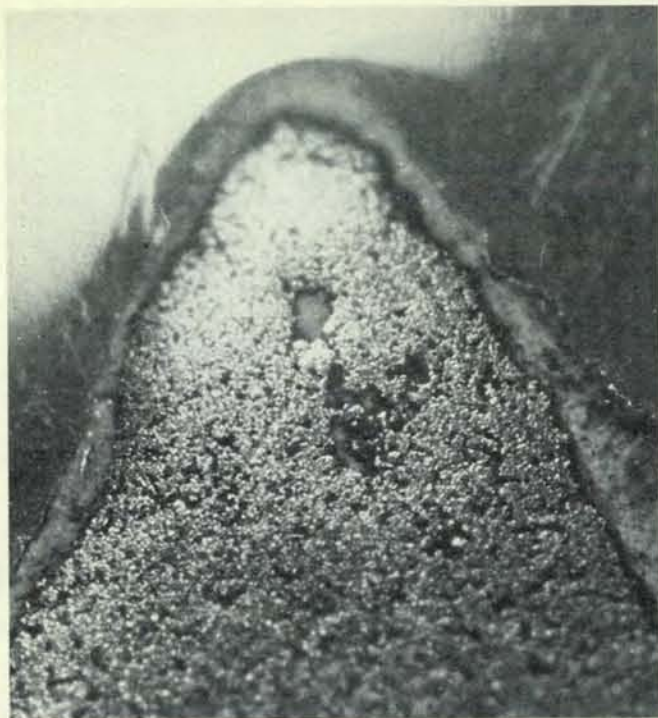


Questo «disco da getto» in bronzo, il cui diametro reale è di circa 105 millimetri, ha un profilo «alare» appiattito a lama (si veda nell'illustrazione in basso di pagina 35). Le fotografie mettono in evidenza la diversa forma delle due facce del disco, con costolatura a spigolo smussato (al centro) e superficie quasi piatta (in alto). Si tratta del tipo di disco da getto più perfezionato e certamente più efficace dal punto di vista offensivo. In basso l'esemplare è visto di profilo. Il margine esterno affilatissimo a lama mette in risalto la tremenda potenzialità offensiva dell'arma. Il profilo aerodinamico asimmetrico è la caratteristica più saliente di questi dischi altamente perfezionati.

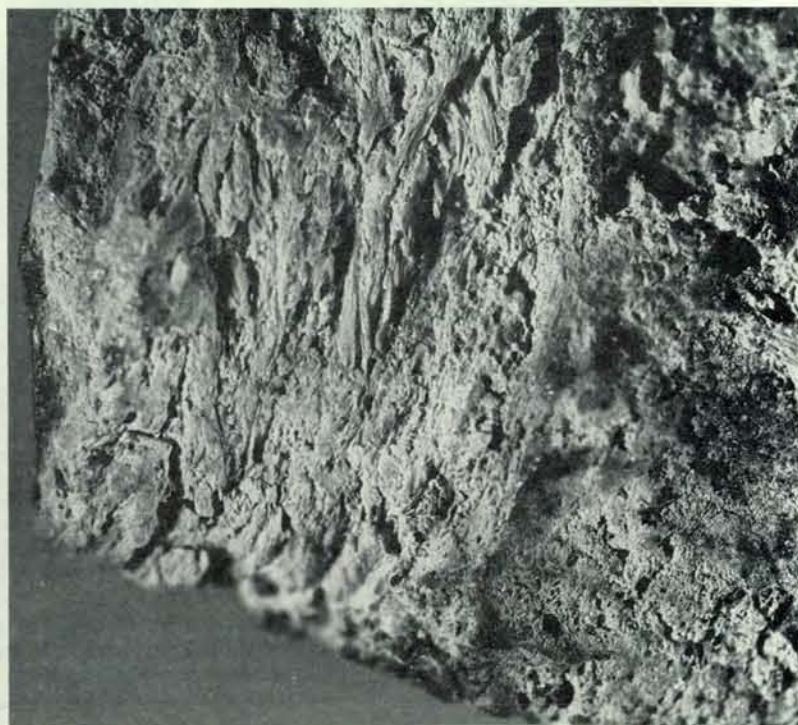
strezza, era uno strumento d'offesa particolarmente efficace alle medie distanze, capace di provocare ferite mortali da percussione e taglio, penetrando in profondità come una lama di spada con energia concentrata in un solo punto, con effetti devastanti sui tessuti e sulle strutture os-

see. Bisogna infatti tenere presente che alla forza di percussione normale alla superficie di impatto, il disco rotante affilato sommava la componente di taglio tangente alla superficie-bersaglio.

L'arma, che veniva portata in gruppi di diametro differente graduato (forse pro-



La fotografia a destra è una «sezione lucida» trasversale metallografica di un «disco da getto» con sezione a parallelogramma appiattito, ingrandita circa 15 volte. L'attacco microchimico con soluzione cloridrica al 30 per cento evidenzia la differente tessitura cristallina del margine tagliente e dei bordi induriti a martellatura (zone più chiare) rispetto a una tessitura microgranulare indisturbata e omogenea del corpo centrale (zona più scura). Qui sopra, nel dettaglio del tagliente a ingrandimento maggiore (circa 45 volte) si nota la tessitura cristallina del bronzo, deformata irregolarmente dagli effetti della martellatura. Dalle ricerche compiute su quattro esemplari risulta così che la lega bronzea era sottoposta a un processo di incrudimento meccanico lungo i margini esterni del disco allo scopo di indurirne il tagliente e aumentarne l'efficacia.



Una conferma al fatto che i dischi da getto fossero agganciati alle vesti, probabilmente in gruppi e mediante una fibula, è venuta dal ritrovamento di frammenti di tessuto di lino che si sono conservati per minera-

lizzazione e incrostazione sulla patina malachitica di un disco. Nella fotografia a sinistra (ingrandita circa 40 volte) è visibile l'intreccio ortogonale dei fili. La fotografia a destra è ingrandita circa 20 volte.



Nella fotografia a sinistra (ingrandita circa 11 volte) la superficie di un «disco da getto», opportunamente «decorticata» dello strato di mineralizzazione malachitica, mette in risalto sul lato del tagliente le sottili striature da affilatura «a pietra» subparallele o inclinate di 10-20 gradi rispetto al bordo esterno. Il margine interno, visibile in basso a sinistra, presenta invece solcature e rigature molto più profonde, spaziate e grossolane, orientate quasi ortogonalmente alla circonferenza del disco, causate certamente dall'uso della lima in fase di rifinitura. Nella fotografia qui sopra (ingrandita circa 15 volte) sono visibili le striature da affilatura «a pietra» sul margine tagliente, inclinate di circa 10 gradi, direttamente «replicate» dalla mineralizzazione in cuprite (Cu_2O). Si tratta in questo caso di un disco da getto quasi completamente mineralizzato, il cui colore naturale è rosso mattone.

prio in funzione di un impiego a distanze diverse), agganciati alle vesti con una fibula all'altezza del seno o del ventre, doveva essere scagliata con una particolare flessione elastica a scatto del polso, dal basso verso l'alto o viceversa, distendendo contemporaneamente in avanti il braccio con un'energica spinta di spalla. Per rendere meglio l'idea, più o meno come si può lanciare un sasso appiattito per farlo rimbalzare sull'acqua. Una conferma al fatto che questi dischi da getto fossero agganciati alle vesti è venuta dal ritrovamento di frammenti di tessuto di lino mineralizzati, incrostati nella patina malachitica di un disco da getto.

Certamente importante era la posizione e l'orientamento tra le dita della lancia-trice, dato il profilo talora asimmetrico del disco stesso. La sezione aerodinamicamente realizzata a triangoli contrapposti, a parallelogramma piatto o a profilo alare, assicurava al disco rotante la portanza necessaria per una traiettoria il più possibile tesa ed efficace. La pratica del lancio doveva richiedere particolari doti psicofisiche, ottima vista, prontezza di riflessi e comportare un allenamento assiduo e metodico.

Questo particolare tipo di arma da getto della civiltà del ferro dell'Italia peninsulare sembra trovare un puntuale e inte-

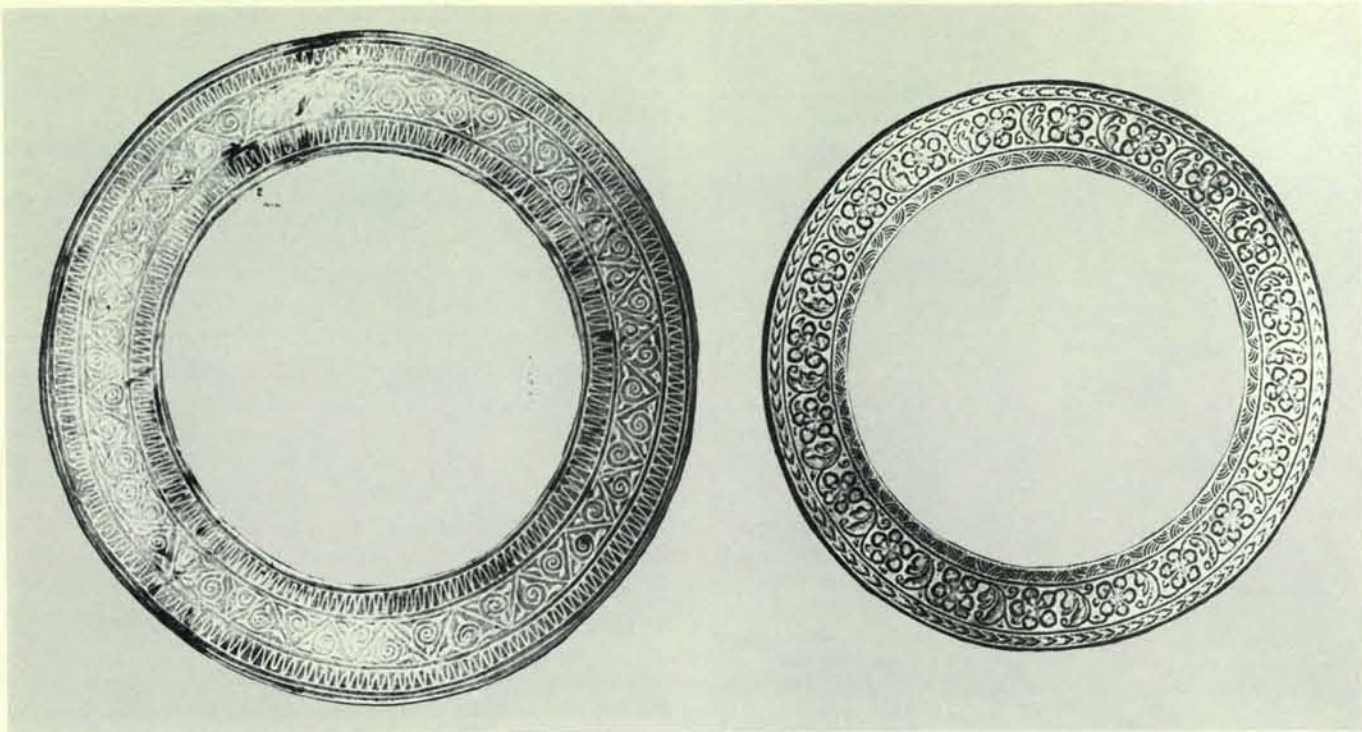
ressantissimo confronto nel già ricordato *chakram* indiano del Punjab.

Il *chakram* (si vedano le illustrazioni alla pagina successiva), che spesso era riccamente inciso a freddo con motivi decorativi geometrici e floreali, fu arma micidiale e insidiosissima, tipica dei guerrieri indiani Sikh, potente comunità del Punjab durante le rivolte contro l'Islam tra il 1738 e il 1780 e, successivamente, nelle lotte contro gli inglesi sino al 1849. I *chakram* venivano portati a gruppi di sei-otto, con diametro variabile a scalare in media tra 230-220 e 120-100 millimetri, infilati su di una specie di lungo bastone di forma leggermente conica rivestito di stoffa. Talora si portavano anche infilati sul braccio o su di un particolare turbante conico a «pan di zucchero». La tecnica di lancio del *chakram* variava a seconda delle particolari attitudini di ogni lancia-tore; tuttavia sembra che due fossero i metodi comunemente adottati.

Il primo consisteva nel far ruotare velocemente il disco su un dito, incrementando la velocità di rotazione fino a quando l'arma, uscendo verso l'alto, poteva assumere una traiettoria rettilinea opportunamente indirizzata verso il bersaglio. Il secondo metodo, che forse era anche quello di più semplice esecuzione, consisteva nell'imprimere una rotazione al disco, tenuto

stretto tra due o tre dita, al momento del lancio in direzione del bersaglio. Da quanto ci è noto sembra che un abile lancia-tore Sikh potesse colpire un uomo in punti vitali a oltre 50-60 metri di distanza.

Per quanto eccezionale e strettamente «endemico» possa essere l'impiego dei «dischi da getto» come proiettili da usarsi in combattimento in campo aperto e ancor più in agguati e attacchi di sorpresa, il *chakram* non rappresenta l'unico esempio di armi siffatte. Nell'antico Giappone, dal 400 circa a.C. sino all'epoca Tokugawa (1600 circa), i famigerati *ninja*, esperti in azioni di guerriglia e di sabotaggio, nonché in assassinii su commissione e azioni di «commando», reclutati spesso tra i delinquenti comuni o tra i samurai rinnegati in qualità di spie, terroristi e sicari, usarono, nella gamma vastissima di armi e strumenti di insidia particolarissimi loro propri, dischi stellati in acciaio muniti di numerose punte acuminate e affilate. Questi dischi, detti *shuriken* o *shaken* («stelle della morte») venivano lanciati con un movimento circolare del polso dall'alto verso il basso, che imprimeva loro un moto rotatorio e giungevano sul bersaglio con effetti letali. Se si tiene conto che spesso le punte venivano cosparse di veleno, è facile immaginare la micidialità di tali strumenti di morte.



Questi *chakram* del XVIII-XIX secolo, una micidiale arma da getto indiana, sono costruiti in ottimo acciaio *Wood* forgiato e temprato. Decorati con incisioni geometriche e floreali hanno un diametro di 195

millimetri (a sinistra) e di 165 millimetri (a destra). La larghezza della fascia anulare è rispettivamente di 34 e di 25 millimetri. Entrambi gli esemplari sono conservati al Rijksmuseum voor Volkenkunde di Leida.



Questo *chakram* indiano del Punjab in acciaio forgiato è del XVIII secolo. La proporzione tra il suo diametro di 237 millimetri e la larghezza della fascia anulare, 25 millimetri, trova un preciso riscontro nei «dischi da getto» in bronzo dell'Età del ferro dell'Italia centromeridionale.

L'uso dei «dischi da getto» come armi tipicamente femminili delle culture peninsulari dell'Età del ferro pone ancora una volta il problema dell'utilizzazione di particolari strumenti e armi non solo sul piano paleontologico e culturale, ma anche e soprattutto all'interno delle società umane, quali elementi gerarchicamente differenzianti le classi sociali e, nella fattispecie, i due sessi.

L'ipotesi che i dischi da getto fossero adottati come arma tipicamente femminile in quanto facilmente maneggiabili dalle donne, dotate in media di una minore potenza muscolare rispetto al sesso maschile, cade immediatamente non appena si consideri che i già ricordati Sikh del Punjab erano e sono una tra le popolazioni indiane fisicamente e atleticamente più dotate. È assai più probabile che i dischi stessi celino un profondo simbolismo per noi oggi difficilmente accessibile e interpretabile. Non si può escludere d'altra parte che, oltre alla loro funzione offensiva primaria di armi da getto, le lame circolari dei dischi potessero proficuamente essere impiegate per usi domestici, ad esempio come raschiatoi per pelli, certamente più razionali e pratici delle lame lunghe e acuminate di normali coltelli o daghe. In realtà, l'uso del raschiatoio circolare in pietra scheggiata risale almeno al Paleolitico superiore (Musteriano). Ciò potrebbe spiegare meglio la vasta gamma dimensionale dei dischi, alcuni dei quali di diametro troppo ridotto e di massa-peso troppo esigua per poter essere adoperati efficacemente come proiettili e soprattutto renderebbe ragione dell'usura asimmetrica che spesso si riscontra sui dischi di minore diametro.

Per concludere, l'esistenza in diversi orizzonti culturali e società umane di armi speciali di dotazione e impiego tipicamente femminili non rappresenta certamente un fatto eccezionale. Molte antiche culture europee e asiatiche hanno visto il ruolo predominante della donna non solo nella conduzione della vita familiare e degli affari della tribù o del clan, ma anche nelle vicende militari e belliche. Per fare un solo esempio, nel Giappone medioevale fino all'epoca Meiji (1868) la donna del *buke* (samurai) veniva addestrata all'uso di armi tradizionali (spada, lancia), ma maneggiava con somma destrezza e coraggio alcune armi «speciali» che in un certo senso divennero tipiche del costume femminile dell'epoca.

Tra esse il *naginata*, o lancia a lama curva falciante, arma formidabile che la giovane donna nipponica di alto rango doveva imparare a maneggiare, in una scherma complicata e difficile, prima dei diciotto anni. Altra arma portata esclusivamente dalla donna-samurai era il *kai-ken* (una variante del pugnale d'armatura detto *tantō*), daga corta a un solo taglio leggermente curvo, che non abbandonava mai la manica o la fuscaccia del suo kimono e che ella sapeva lanciare con micidiale destrezza e precisione. Tale arma assunse il ruolo di «guardiano dell'onore» della donna e fu adoperata sempre per compiere il suicidio rituale o *jigai* (taglio delle arterie carotidi e giugulari).

Può a questo punto sorgere l'interrogativo se possa esistere un qualsiasi rapporto o collegamento sul piano storico-culturale ed etnologico tra i «dischi da getto» dell'Età del ferro e i *chakram* indiani. Potrebbe sedurre l'idea che i dischi da getto risentano di lontane influenze «orientali» e possano quindi rappresentare una categoria di oggetti di imitazione da prototipi asiatici. Purtroppo non si ha notizia di ritrovamenti, in strati archeologici dell'India o dell'Asia anteriore, di strumenti simili. Anche l'ipotesi contraria, che i dischi da getto dell'Italia antica peninsulare rappresentino i modelli dei *chakram* attuali, non si può certo sostenere meglio.

La questione è probabilmente oziosa e sembra riaprire le polemiche sorte in passato a proposito di una corta spada greca a lama falcata, la *kōpis*, che trova un preciso confronto attuale nella spada-coltello nepalese dei guerrieri *Nepari* e *Ghurka*, il *kukri*.

Se il *kukri* rappresenti l'«antenato orientale» della *kōpis* oppure se sia giunto fino al Nepal sulle ali della conquista macedone di Alessandro il Grande, forse non lo sapremo mai. Probabilmente, come spesso accade, la verità sta nel mezzo e tanto le spade falciformi greche e asiatiche quanto i «dischi da getto» italici e asiatici possono essere stati concepiti e realizzati indipendentemente nei diversi ambienti e contesti, per quei fenomeni di «convergenza culturale» e di «invenzione autonoma» tanto frequenti nel corso della preistoria e della storia dell'umanità.

I raccoglitori per il 1984

Questi raccoglitori corrispondono ai volumi XXXII e XXXIII de *LE SCIENZE*, e rispettivamente ai fascicoli da gennaio (n. 185) a giugno (n. 190) e da luglio (n. 191) a dicembre (n. 196).

Sono ancora disponibili i raccoglitori dal Vol. XXIV al XXXI e dei raccoglitori non numerati appositamente approntati per sostituire i raccoglitori esauriti.

I raccoglitori si possono richiedere direttamente all'editore usando l'apposita cartolina allegata alla rivista e unendo il relativo importo; gli ordini infatti vengono evasi solo a pagamento avvenuto.



I raccoglitori si trovano anche presso i seguenti punti di vendita:

BOLOGNA

Libreria Parolini
Via U. Bassi 14

FIRENZE

Libreria Marzocco
Via de' Martelli 22/R

GENOVA

Libreria Int. Di Stefano
Via R. Ceccardi 40/R

MILANO

Le Scienze S.p.A.
Via del Lauro 14

TORINO

Libreria Zanaboni
C.so Vittorio Emanuele 41

NAPOLI

Libreria Guida A.
Via Port'Alba 20/21

PADOVA

Libreria Cortina
Via F. Marzolo 4

PALERMO

Libreria Dante
Quattro Canti di Città

ROMA

Claudio Aranci
Viale Europa 319 (EUR)

Ogni raccoglitore
L. 3.600

Come i geni controllano un comportamento innato

Le tecniche del DNA ricombinante vengono utilizzate per definire un gruppo di geni che codifica per una serie di neuropeptidi affini, la produzione coordinata dei quali regola uno schema d'azione fisso

di Richard H. Scheller e Richard Axel

Certi schemi stereotipi di comportamento animale sono innati. Sono modellati dall'evoluzione ed ereditati da generazioni successive; in gran parte immutati dall'esperienza o dall'apprendimento, vengono esibiti da tutti gli organismi di una determinata specie e non da altre specie. Gli etologi hanno descritto questi schemi comportamentali stereotipi innati come «schemi d'azione fissi», schemi comportamentali che consistono di parecchi elementi indipendenti, i quali o formano tutti insieme una sequenza coordinata o non si verificano affatto. Ogni animale eredita questi schemi d'azione fissi, che caratterizzano il comportamento della specie cui appartiene.

In che modo un animale eredita questo repertorio comportamentale? E che cosa eredita? Eredita il DNA e i geni del DNA possono specificare il comportamento stereotipo in due modi. Innanzitutto essi possono specificare una ben precisa rete di cellule nervose e muscolari interconnesse, che vengono collocate nella loro sede e attivate tutte assieme nel corso dello sviluppo di quell'animale. Un comportamento stereotipo, però, viene suscitato solo in particolari situazioni o in particolari stadi del ciclo vitale di un animale e solo dall'attività coordinata di particolari parti della rete di neuroni e di cellule muscolari. Oltre alla rete, dunque, i geni devono specificare gli elementi di controllo, sostanze che eccitano connessioni specifiche preesistenti in un modo rigidamente determinato per generare al momento giusto uno schema d'azione fisso.

In qualità di biologi molecolari che studiano il sistema nervoso, noi cerchiamo di identificare questi elementi di controllo e i geni che ne governano la sintesi. Sperimentiamo così di capire qualcosa sui fattori che svolgono un ruolo nel generare repertori comportamentali innati, su come si evolvono in differenti specie e su come si sviluppano in una determinata specie.

I comportamenti specifici, però, diver-

samente da caratteri come il colore degli occhi o le alterazioni ereditarie dell'emoglobina, non hanno alcuna probabilità di riflettere direttamente lo stato di geni specifici. Il sistema nervoso centrale integra e filtra i comandi che provengono dai geni in modi che per la maggior parte sono inaccessibili alla sperimentazione. Quanto più complesso è il sistema nervoso centrale, tanto più elusivo è il rapporto tra una serie di geni (un genotipo) e i caratteri che si possono osservare (un fenotipo). È più facile studiare i geni che specificano il comportamento in un organismo che sia sufficientemente sofisticato da esibire repertori comportamentali interessanti, ma abbastanza semplice da permettere di attribuire il suo comportamento a cellule ben identificabili. Noi stiamo appunto lavorando su un animale del genere, il mollusco *Aplysia*, un gasteropodo marino privo di conchiglia, che può pesare anche cinque chilogrammi.

Il sistema nervoso centrale di *Aplysia* è numericamente semplice; consiste di circa 20 000 cellule nervose riunite in quattro paia di gangli cefalici e in un ganglio addominale. Ciò è in netto contrasto con il cervello di un mammifero, che ha forse un milione di volte più neuroni. Inoltre, i neuroni di *Aplysia* possono anche avere un diametro di un millimetro, cioè più di 1000 volte la dimensione di una tipica cellula cerebrale umana. La maggior parte di queste enormi cellule contiene quantità altrettanto grandi di DNA: fino a un microgrammo, cioè parecchie centinaia di migliaia di volte il contenuto in DNA di un tipico neurone di mammifero. Le funzioni che, in un sistema nervoso più complesso, sarebbero realizzate da una grande quantità di neuroni correlati tra loro, in *Aplysia* possono venir realizzate da una sola cellula di notevoli dimensioni. Una serie di ricerche, in particolare l'elegante lavoro di Eric R. Kandel e dei suoi collaboratori del Colle-

ge of Physicians and Surgeons della Columbia University, ha messo in relazione specifici schemi comportamentali di quel gasteropodo con il funzionamento di particolari cellule. Noi stessi con i nostri collaboratori abbiamo continuato sulla stessa strada esaminando l'attività di specifici geni in una cellula singola e abbiamo potuto attribuire uno schema di comportamento all'attività di singoli geni.

In età adulta, *Aplysia* è in gran parte impegnata nelle attività di alimentazione e di riproduzione. Vari aspetti del comportamento riproduttivo del gasteropodo sono marcatamente ritualistici e comportano una serie coordinata di schemi stereotipi, che realizzano il corteggiamento, l'accoppiamento e la deposizione di uova fecondate. *Aplysia* è un vero animale ermafrodito, è un organismo che funziona da maschio e da femmina, nella maggior parte dei casi simultaneamente, con i singoli individui che si accoppiano in lunghe catene di mezza dozzina di animali o più; talvolta la catena si richiude addirittura a formare un cerchio. La fecondazione è interna e ha luogo nel gonodotto; in seguito, le uova fecondate vengono deposte in mare e quindi si sviluppano all'esterno. I geni che stiamo studiando sono quelli interessati nella serie elaborata, ma completamente stereotipa, di comportamenti coordinati che permettono di realizzare la deposizione delle uova.

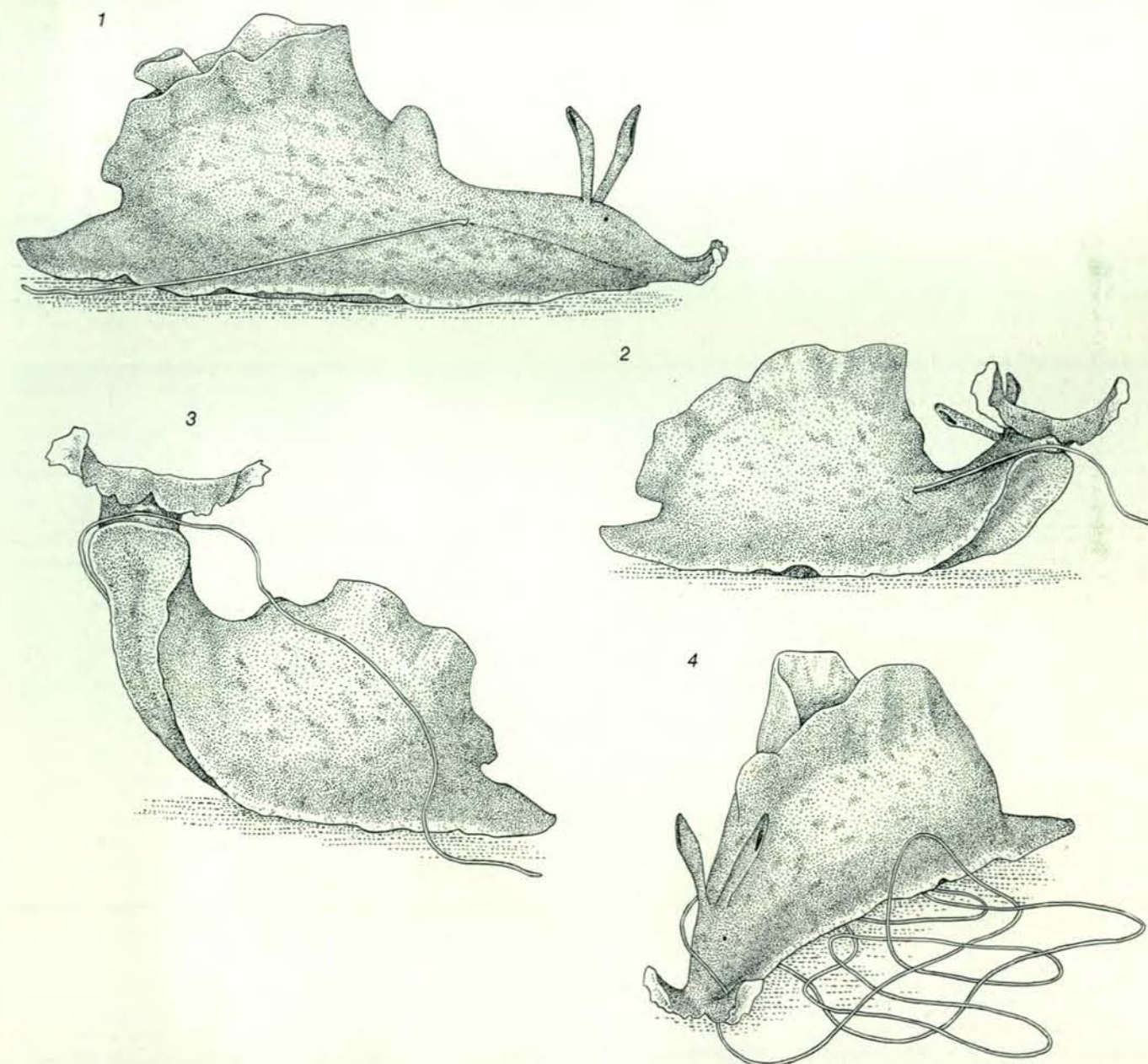
Le uova sono deposte in lunghi cordoni che ne comprendono più di un milione. Il gasteropodo, mentre uno di questi cordoni viene espulso per contrazione dei muscoli del gonodotto, si ferma e rimane immobile. Cessa anche di mangiare, mentre aumentano il suo battito cardiaco e il numero di atti respiratori. Afferra poi con la bocca il cordone di uova e, con una serie di caratteristici ondeggiamenti del capo, aiuta a estrarlo dal gonodotto e ad avvolgerlo in una massa irregolare. Una piccola ghiandola boccale secerne una sostanza mucoide appiccicosa, che si at-

tacca a tale massa. Quindi, con un forte ondeggiamento del capo, l'animale riesce ad attaccare quest'ultima, tutta intera, a un sostegno solido, per esempio uno scoglio. Numerose azioni disparate si sono riunite in una sequenza rigidamente coordinata per assolvere una funzione comune: la deposizione di uova fecondate in un modo che procurerà protezione durante lo sviluppo.

Quali sostanze controllano questo schema comportamentale coordinato e quali geni codificano per queste sostanze? Innanzitutto, dove queste sostanze vengono sintetizzate? Parecchi anni fa, Kandel e Irvin Kupfermann hanno identificato due ammassi di neuroni, le cellule a

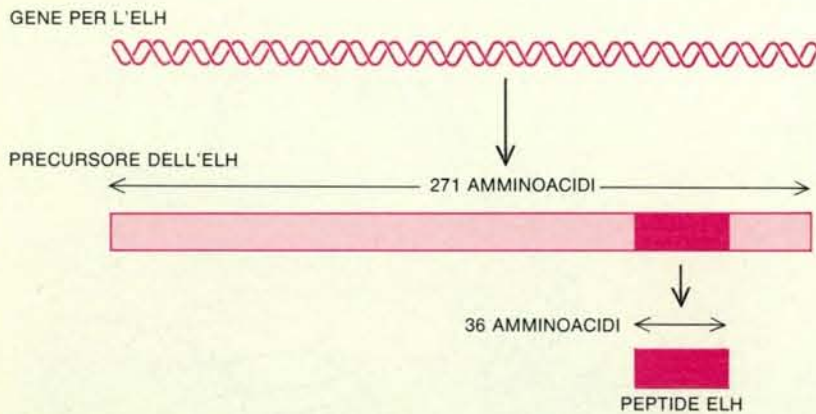
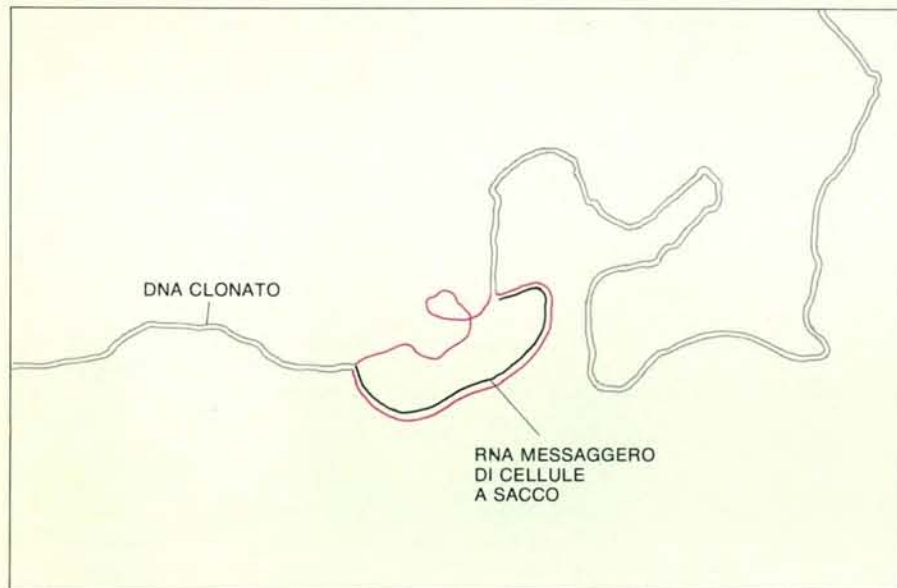
sacco, alla sommità del ganglio addominale. Quando un estratto di queste cellule è stato iniettato in esemplari vivi di *Aplysia*, ha scatenato l'intero repertorio di comportamenti associato con la deposizione delle uova anche se quegli animali, in effetti, non si erano accoppiati e le loro uova non erano di conseguenza fecondate. In studi successivi, Stephen W. Arch del Reed College e Felix Strumwasser del California Institute of Technology sono riusciti a isolare uno dei fattori attivi delle cellule a sacco e lo hanno identificato come peptide, o piccola proteina, costituito da soli 36 amminoacidi. Essi hanno anche trovato che questo peptide, quando veniva somministrato ai gasteropodi in

questione, induceva alcuni ma non tutti i comportamenti legati alla deposizione delle uova, suggerendo che esso doveva essere uno soltanto di diversi fattori che controllano l'intero repertorio comportamentale. Strumwasser ha determinato la sequenza lineare degli amminoacidi nel peptide, che è stato designato come ormone della deposizione delle uova (ELH, dall'inglese *egg-laying hormone*). Dato che esiste una relazione lineare tra la sequenza degli amminoacidi di una proteina e la sequenza nucleotidica del gene che codifica per essa, l'identificazione e la risoluzione della sequenza amminoacidica di un peptide che controlla il comportamento finiscono per introdurre un pro-



In un acquario di laboratorio, questo grosso mollusco gasteropodo marino (*Aplysia*) esibisce il comportamento che riguarda la deposizione delle uova. La contrazione dei muscoli del gonodotto fa espellere un cordone di uova, racchiuse ciascuna in un involucro (1). L'animale afferra il cordone con la bocca (2) e fa ondeggiare il capo (3), favorendo così la definitiva uscita delle uova dal gonodotto; alla fine attacca a un

substrato solido il cordone che si è nel frattempo tutto aggrovigliato (4). Nel caso rappresentato in queste immagini, il comportamento è stato indotto in un animale non fecondata, iniettandogli un estratto di cellule a sacco, neuroni in cui una «poliproteina» viene sintetizzata per poi dare origine, mediante scissione, a un ormone per la deposizione delle uova (ELH) e ad altri peptidi associati con questo comportamento.



Il gene per l'ELH è visibile nella microfotografia elettronica (*in alto*). Frammenti di DNA di *Aplysia* sono stati ricombinati con il DNA di un batteriofago e clonati. Un clone che contiene il gene per l'ELH è stato identificato, disteso e quindi posto in presenza dell'RNA messaggero per l'ELH, in condizioni in grado di promuovere la formazione di ibridi RNA-DNA. Come risulta nella mappa al centro, l'RNA (*in nero*) si è ibridato con il filamento complementare del gene per l'ELH (*in colore*), disgregando così la doppia elica del DNA clonato. Quando si è ricostruita la sequenza del gene per l'ELH, si è visto che essa è in grado di codificare per una proteina precursore, una catena di ben 271 amminoacidi, al cui interno è stato identificato il peptide ELH.

blema di biologia comportamentale nel campo della genetica molecolare.

In collaborazione con Linda B. McAllister, James F. Jackson, James H. Schwartz e Kandel ci siamo prefissi di isolare il gene che codifica per l'ELH dal genoma di *Aplysia*, cioè da tutto il complesso del DNA presente nei cromosomi di questo mollusco. Il procedimento usato per isolare geni specifici dipende dalle tecniche del DNA ricombinante. In breve, si costituisce un «archivio» di molecole di DNA ricombinante, ciascuna delle quali contiene un piccolo frammento del genoma di *Aplysia*, e quindi si analizza l'intera raccolta con una sonda che identificherà il gene per l'ELH. Abbiamo scisso il DNA del mollusco in molte migliaia di frammenti, calcolando che ogni frammento includesse un solo gene o forse alcuni geni. Abbiamo poi «ricombinato» i frammenti del DNA di *Aplysia* con il DNA di un batteriofago, il fago lambda, e abbiamo avvolto il DNA ricombinante nella guaina proteica del fago. I fagi ibridi così ottenuti sono serviti da vettori per introdurre il piccolo DNA in un batterio e clonarlo. Quando essi sono aggiunti a una coltura batterica, ciascuno infetta una singola cellula batterica e si moltiplica, uccidendo tale cellula; la progenie che ne deriva continua a riprodursi, uccidendo le cellule adiacenti e creando una «placca», cioè un foro, nella coltura. Ogni placca è occupata da milioni di fagi derivati da un unico fago e pertanto contiene un clone di milioni di copie di un singolo frammento di DNA di *Aplysia*.

Nello scegliere una sonda con cui scandagliare i cloni e trovare il gene per l'ELH, ci siamo basati sul fatto che l'ELH viene sintetizzato in abbondanza nelle cellule a sacco, che pertanto devono contenere una grande quantità di RNA messaggero per quell'ormone. A questo punto è necessaria una spiegazione. Il DNA è una molecola a duplice filamento e ciascuno dei due filamenti è una sequenza di quattro differenti subunità nucleotidiche. L'informazione genetica è contenuta in codice nella sequenza dei nucleotidi che definisce un gene. I gruppi di tre nucleotidi, o triplette, sono detti codoni, o parole del codice genetico: ciascun codone specifica un diverso amminoacido di una serie di 20 amminoacidi, che sono le subunità delle proteine. Il DNA non viene tradotto direttamente in proteina. Il suo filamento codificante, l'elica trascrivente, viene dapprima trascritto in un filamento complementare di un acido nucleico che gli somiglia e che è l'RNA messaggero, la cui sequenza riflette, infatti, quella del gene; l'RNA viene poi tradotto in proteina in base al codice genetico.

Dato che l'RNA messaggero per l'ELH è predominante nelle cellule a sacco, ma non nei tessuti che non siano tessuto nervoso, abbiamo isolato l'RNA messaggero sia da cellule a sacco sia da cellule che non appartenevano al sistema nervoso. Lo abbiamo poi sottoposto a una «trascrizione inversa» e abbiamo ottenuto DNA a esso complementare che abbiamo marca-

to con un isotopo radioattivo. Il DNA marcato è stato quindi esposto a tutto l'archivio di cloni ricombinanti in condizioni tali da permettere che il DNA complementare per l'ELH si «ibridasse», o si legasse, con segmenti di geni clonati, contenenti il codice per l'ELH; il DNA com-

plementare di cellule non appartenenti a tessuto nervoso non si sarebbe comportato così. L'autoradiografia ha permesso di rivelare i cloni con i quali la sonda costituita da DNA per l'ELH si è ibridata e con i quali, invece, non si è ibridata la sonda proveniente da cellule non nervose.

In questo modo siamo riusciti a identificare numerosi cloni contenenti geni per l'ELH. (Come spiegheremo, è risultato che esistono tre geni, in qualche misura diversi, ma essi sono così simili tra loro che la sonda proveniente dalle cellule a sacco si è ibridata con tutti e tre.) Dopo



In questa microfotografia si osserva un ganglio addominale di *Aplysia*, a un ingrandimento di circa 40 diametri. I due ammassi di cellule a sacco dove viene sintetizzato l'ELH sono ben visibili, disposti trasversalmente ai grossi fasci nervosi in alto a sinistra e in alto a destra.

Entro il corpo principale del ganglio sono distinguibili numerosi neuroni di notevoli dimensioni, molti dei quali, identificati individualmente, sono risultati invariati in tutti i membri della specie. L'ELH e i peptidi associati hanno effetti specifici sull'eccitazione di certi neuroni.

aver identificato i cloni contenenti il gene per l'ELH, li abbiamo moltiplicati in grandi colture, abbiamo isolato da essi il DNA ricombinante e lo abbiamo posto in presenza dell'RNA messaggero per l'ELH. Nelle microfotografie elettroniche che abbiamo ottenuto, abbiamo potuto scorgere, sulla lunga molecola di DNA ricombinante, i siti esatti in corrispondenza dei quali l'RNA si è ibridato con il DNA complementare e così abbiamo potuto concentrarci sull'esatto segmento di DNA che costituisce il gene per l'ELH.

Dopo aver isolato il gene per l'ELH, ne abbiamo determinato la sequenza nucleotidica, cioè abbiamo determinato l'ordine in cui i suoi nucleotidi sono disposti. Data tale sequenza, abbiamo potuto procedere alla sua «traduzione inversa», seguendo il codice genetico, e quindi abbiamo potuto dedurre la sequenza di amminoacidi della catena polipeptidica contenuta in codice in quel gene. Ne è risultato che la proteina prodotta dal gene deve avere 271 subunità amminoacidiche. Sapevamo, tuttavia, che l'ormone ELH ha solo 36 amminoacidi e, in effetti, potevamo riconoscere la sua corta sequenza all'interno della più lunga sequenza della proteina. Il peptide che costituisce l'ELH è fiancheggiato da ambedue le parti della stessa coppia di amminoacidi; lisina e arginina. È noto che questo binomio è un segnale di scissione. Serve, infatti, come sito in corrispondenza del quale enzimi specifici tagliano una grande molecola proteica che funge da precursore e ottengono una molecola proteica più piccola, o peptide, che è attiva. La spiegazione più ovvia è stata, quindi, che la catena costituita da 271 amminoacidi è un precursore da cui viene staccato l'ELH.

Era strano, tuttavia, che il peptide che rappresentava l'ormone ELH fosse solo una parte così piccola del prodotto proteico del gene che ne controllava la sintesi. La restante parte della catena precursore veniva semplicemente eliminata senza essere utilizzata, oppure poteva anch'essa contenere peptidi biologicamente attivi? Abbiamo esaminato tutta la sua sequenza alla caccia di altri segnali di scissione e ne abbiamo trovati in tutto dieci; se ogni sito veniva riconosciuto e tagliato si sarebbero prodotti dalla singola proteina precursore dell'ELH 11 peptidi ben distinti, che sarebbero anche potuti essere altri elementi in grado di controllare il comportamento relativo alla deposizione delle uova.

Earl M. Mayeri e collaboratori della San Francisco School of Medicine dell'Università della California avevano indagato per vario tempo sulle proprietà fisiologiche dei neuropeptidi secreti dalle cellule a sacco; noi ci siamo uniti al loro gruppo nel tentativo di vedere se qualcuno dei peptidi previsti dalla sequenza nucleotidica del gene per l'ELH fosse presente negli estratti di tali cellule e, in caso positivo, se risultasse attivo nel sistema nervoso. Mayeri e Barry S. Rothman hanno identificato altri tre piccoli peptidi nell'ammasso delle cellule a sacco e ne hanno ricostruito la sequenza: si tratta del

fattore alfa e del fattore beta delle cellule a sacco e del peptide acido. Abbiamo trovato che ciascuno di essi, delimitato da siti di scissione, è codificato assieme all'ELH nel gene che avevamo isolato. Esperimenti di elettrofisiologia hanno poi dimostrato che tre dei quattro peptidi (precisamente l'ELH e i fattori alfa e beta delle cellule a sacco) interagiscono con neuroni specifici del ganglio addominale, ben identificabili: nel ganglio ciascuno agisce da neurotrasmettitore, molecola che media il trasferimento dell'attività elettrica da un neurone all'altro.

L'ELH agisce localmente come trasmettitore dell'eccitazione, aumentando il grado di eccitazione del neurone del ganglio addominale designato con R15. Oltre ad agire come neurotrasmettitore, il peptide ELH si propaga nel sistema circolatorio ed eccita le cellule della muscolatura liscia del gonodotto, facendole contrarre e provocando l'espulsione del cordone di uova. In altre parole, l'ELH agisce non solo da neurotrasmettitore, ma anche da ormone. (L'effetto del peptide sulla muscolatura del gonodotto venne notato prima ancora della sua attività sul sistema nervoso; da qui il nome di ormone per la deposizione delle uova.) Il fattore beta delle cellule a sacco è anch'esso un trasmettitore dell'eccitazione e provoca l'eccitazione di due neuroni simmetrici, designati con L1 e R1, le cui funzioni non sono note. Il terzo neuropeptide, il fattore alfa delle cellule a sacco, è invece un trasmettitore dell'inibizione e inibisce l'attività elettrica di un gruppo di quattro neuroni, designati con L2, L3, L4 e L6. Risulta, inoltre, dotato di un effetto di retroazione positiva, cioè è in grado di eccitare le cellule a sacco dalle quali è stato secreto.

L'associazione di questi tre singoli peptidi, la cui sintesi è controllata da un unico gene, con l'attività di differenti serie di neuroni (e cellule muscolari) suggerisce che un interessante meccanismo possa essere responsabile della complessa schiera di comportamenti legati con la deposizione delle uova. Un unico gene risulta in grado di specificare una «poliproteina», una proteina la cui catena viene tagliata in numerosi piccoli peptidi biologicamente attivi. Può darsi che tutte le componenti del comportamento che riguarda la deposizione delle uova siano mediate da peptidi controllati da quest'unico gene. In caso positivo, vi sarebbe un effetto «tutto o nulla»: nessuna componente del comportamento verrebbe esibita in assenza delle altre. Inoltre, la sintesi della singola poliproteina coordinerebbe in modo rigido la scansione temporale di tutta una serie di comportamenti affini. Un unico gene che codifica per molteplici peptidi neuroattivi potrebbe ordinare così una schiera complessa di comportamenti innati: si avrebbe uno schema d'azione fisso.

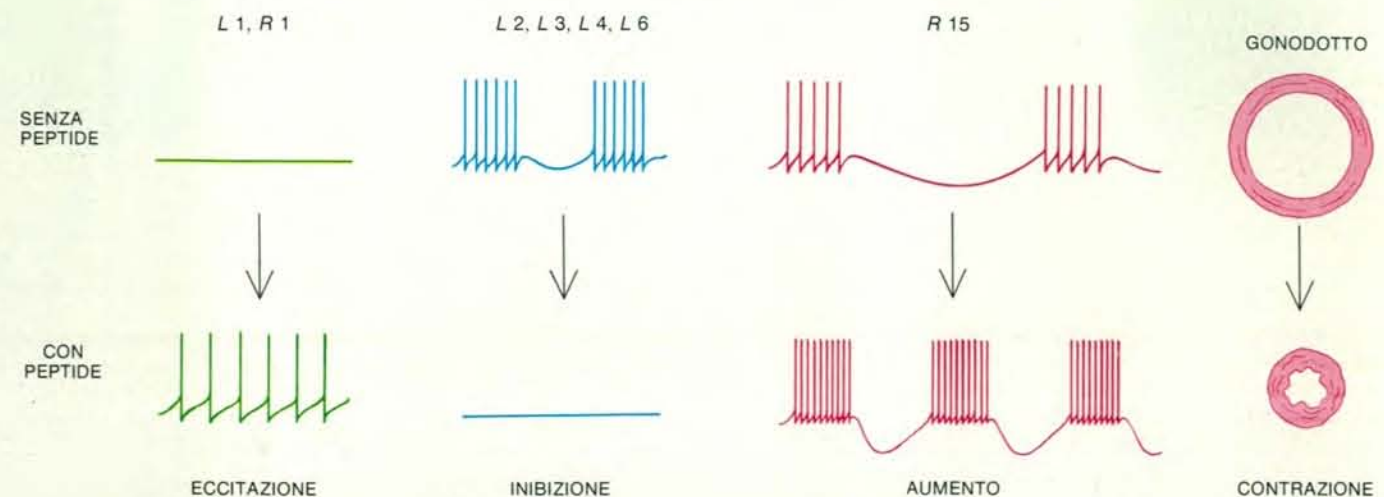
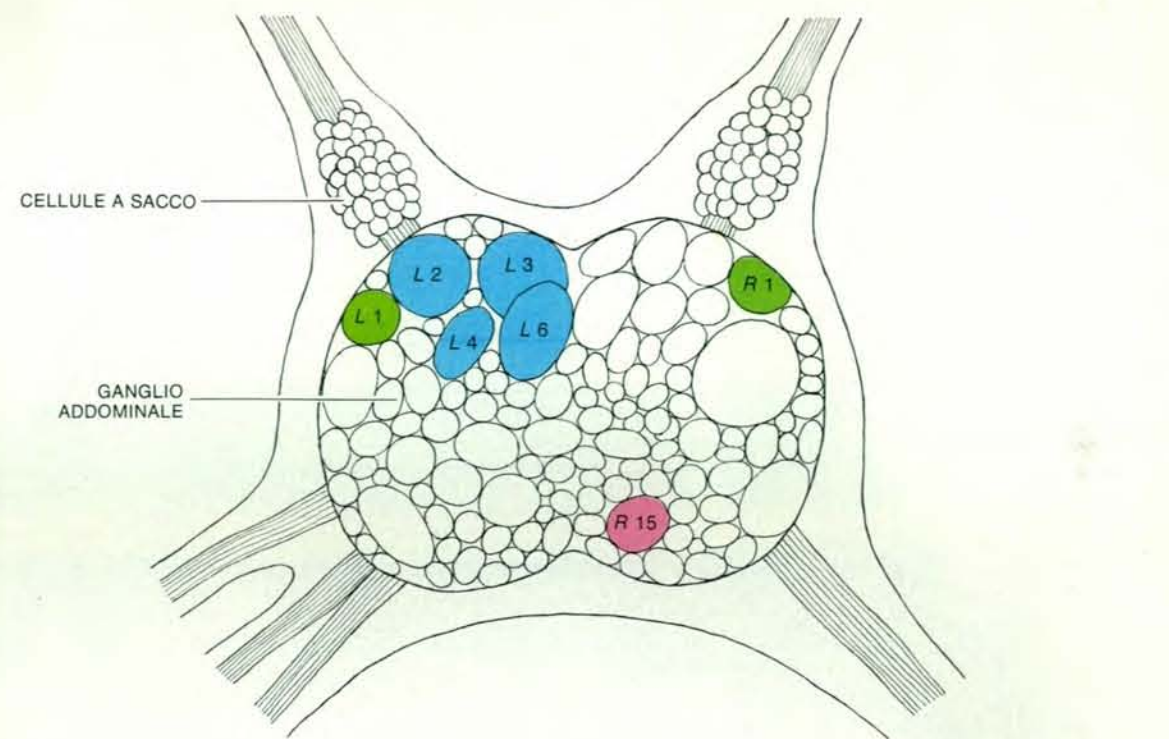
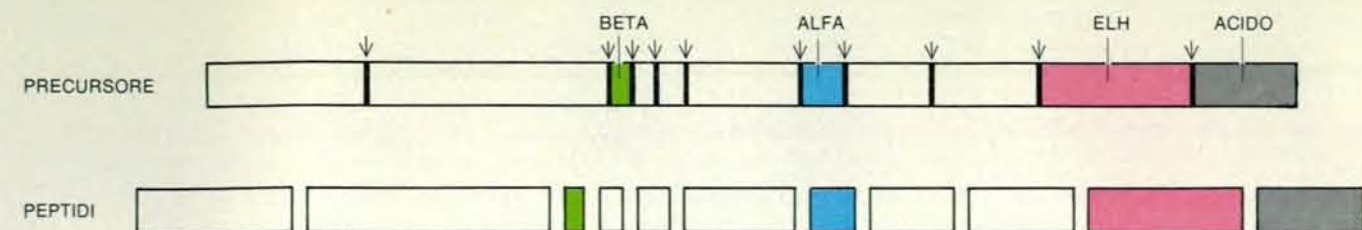
Se i peptidi derivati da una singola poliproteina mediano effettivamente i comportamenti relativi alla deposizione delle uova, che cosa controlla la secrezio-

ne di questi peptidi per dare inizio, in un momento appropriato della vita dell'animale, a tutta una schiera di comportamenti? La secrezione dell'ELH e degli altri ipotetici peptidi interessati nella deposizione delle uova ha luogo dopo una prolungata eccitazione elettrica delle cellule a sacco. Il gruppo di Strumwasser ha isolato due peptidi, designati con A e B e sintetizzati nella ghiandola atriale, un organo del sistema riproduttore di *Aplysia*; di essi hanno ricostruito le sequenze amminoacidiche. L'iniezione di questi peptidi in un animale dà luogo all'eccitazione dei gruppi di cellule a sacco, inducendo la secrezione dell'ELH e dei peptidi a esso associati. Non è stato ancora dimostrato se i peptidi A e B avviano davvero in natura il processo di deposizione delle uova, ma in laboratorio risulta che essi sono i fattori che controllano la secrezione dei peptidi delle cellule a sacco.

Quando abbiamo esaminato la sequenza amminoacidica dei peptidi A e B, abbiamo notato che nel nostro precursore dell'ELH brevi blocchi di amminoacidi erano omologhi di (cioè uguali o quasi uguali a) segmenti dei peptidi della ghiandola atriale. Il lettore ricorderà che, nei nostri esperimenti di clonazione, la sonda ottenuta dall'RNA messaggero per l'ELH si era ibridata con tre DNA in qualche misura diversi, dei quali solo uno codificava per l'ELH e per i peptidi a esso associati nelle cellule a sacco. Poteva darsi che i tre DNA fossero membri di una famiglia multigenica, costituita da un gene che codifica per l'ELH e da altri due geni che codificano per i peptidi A e B.

Con esperimenti di ibridazione, siamo riusciti a dimostrare che l'RNA messaggero derivato dai due geni associati al gene per l'ELH viene proprio sintetizzato nella ghiandola atriale, dove Strumwasser aveva trovato i due peptidi. Abbiamo determinato la sequenza nucleotidica di questi due geni e, mediante traduzione inversa, la sequenza amminoacidica per la quale ciascuno di essi codifica. Abbiamo così notato che, pur essendo i due geni strettamente correlati con il gene per l'ELH, nessuno dei due codifica per questo peptide che è attivo nelle cellule a sacco. Al contrario, essi codificano per i peptidi A e B della ghiandola atriale. I tre geni sono chiaramente membri di una piccola famiglia multigenica con una comune origine evolutiva, ma si sono differenziati per generare complessi di peptidi diversi, anche se funzionalmente affini.

Le somiglianze e le differenze risultano evidenti quando le tre sequenze nucleotidiche e amminoacidiche sono confrontate nei particolari (si vedano le illustrazioni a pagina 61). Ogni gene codifica per una proteina precursore, in cui i siti costituiti dalla coppia lisina-arginina (o talvolta da una singola arginina o da due arginine adiacenti) delimitano i blocchi di amminoacidi che vengono staccati per diventare peptidi attivi. Tutti e tre i precursori cominciano con una caratteristica sequenza segnale di circa 25 amminoacidi, che controlla la trasformazione della



Il precursore dell'ELH è una poliproteina che contiene numerosi peptidi attivi e che presenta (in alto) ben 10 siti (frecce), in corrispondenza dei quali la catena polipeptidica viene scissa da enzimi chiamati endopeptidasi. Se la scissione avvenisse in corrispondenza di tutti i siti, si avrebbero 11 peptidi (seconda fila dall'alto). Quattro di questi vengono secreti dalle cellule a sacco: si tratta dei fattori alfa e beta delle cellule

a sacco, del peptide ELH e del peptide acido. Tre (in colore) fungono da neurotrasmettitori, alterando l'attività di specifici neuroni del ganglio addominale (cellule in colore) in modi specifici (in basso). Il fattore beta eccita le cellule L1 e R1. Il fattore alfa inibisce le cellule L2, L3, L4 e L6. L'ELH aumenta l'eccitazione della cellula R15. Inoltre entra in circolo e funge da ormone, provocando la contrazione del gonodotto.

catena che costituisce la proteina. La catena appena tradotta penetra nel lume di un sistema di membrane, chiamata reticolo endoplasmatico ruvido, e qui comincia a essere modificata: la sequenza segnale viene tagliata e molecole glucidiche assieme a gruppi fosforati sono aggiunti alla proteina, che continua a spostarsi fino a raggiungere l'organello chiamato apparato di Golgi. Qui la proteina precursore viene scissa e i peptidi che la componevano vengono racchiusi in vescicole di piccole dimensioni o sacchi. In risposta a stimoli opportuni, le vescicole si fondono con la membrana esterna della cellula secernente e riversano all'esterno il loro contenuto per interagire con le cellule vicine, perché esso possa diffondersi attraverso il ganglio o possa entrare nella circolazione.

Le differenze tra i precursori che incorporano i peptidi A e B e il precursore che incorpora l'ELH cominciano dopo la sequenza segnale. Descriviamo dapprima i precursori di A e B: un sito costituito da un singolo residuo di arginina segnala l'inizio o del peptide A o del peptide B, costituiti ambedue da 34 amminoacidi. Alla fine di questi due peptidi vi è una sequenza glicina-lisina-arginina, che serve da segnale per la scissione, ma anche per la transamidazione, cioè l'aggiunta di un gruppo amminico (NH₂) all'estremità del peptide, in sostituzione del normale gruppo ossidrilico (OH). La transamidazione «blocca» l'estremità del peptide, forse rendendola più resistente alla degradazione. Segue, sia nel precursore di A sia in quello di B, un segmento di 47 amminoacidi che non ha alcuna attinenza con qualsiasi peptide noto. Viene quindi un altro sito di scissione lisina-arginina, seguito da quello che sembra l'inizio del peptide ELH.

L'ELH non viene sintetizzato, tuttavia, nella ghiandola atriale. L'esame della sequenza nucleotidica dei geni della ghiandola atriale mostra il perché. Nel caso del gene per il peptide A, esso codifica i primi 22 amminoacidi del peptide ELH in modo corretto. Quindi un'unica differenza che riguarda un singolo nucleotide nel codone per l'amminoacido in posizione 23 genera una sequenza arginina-arginina-arginina e stabilisce così un sito di scissione potenziale che potrebbe far scindere quello che altrimenti diventerebbe il peptide ELH. Un differente tipo di cambiamento a carico di un singolo nucleotide si nota nel gene per il peptide B. Si ha qui, infatti, una delezione di un nucleotide nel sesto codone. Pertanto il sistema di lettura delle triplette nucleotidiche si sposta e si genera così un codone di «fine»; la traduzione si arresta, così, dopo che è stato sintetizzato solo un troncione di sei amminoacidi del peptide ELH.

Si consideri ora il precursore dell'ELH che viene sintetizzato nelle cellule a sacco. La sequenza nucleotidica del gene che ne controlla la sintesi è molto simile a quella del gene che codifica per i peptidi A e B, eppure il gene non specifica quei peptidi; specifica l'ELH e numerosi altri peptidi interessati nel processo di deposi-

zione delle uova. Il precursore dell'ELH comincia con la stessa sequenza segnale notata nei precursori di A e di B; seguono poi i cinque primi amminoacidi del peptide B. A questo punto, però, il precursore dell'ELH differisce in modo marcato sia dal precursore di A sia da quello di B; il gene per l'ELH contiene, infatti, una sequenza di ben 240 nucleotidi che non è presente né nel gene di A né nel gene di B e che codifica per 80 amminoacidi, inserito che include quattro segnali di scissione che delimitano tre fattori delle cellule a sacco: beta, gamma e delta. Il fattore beta, come abbiamo già ricordato, ha un effetto specifico sui neuroni del ganglio addominale L1 e R1.

Dopo i 240 nucleotidi, la sequenza nel gene riprende normalmente, senza alcuna modificazione del sistema di lettura, e codifica per il sesto amminoacido del peptide B; prosegue quindi con una successione di nucleotidi che è molto simile alla sequenza dei due geni della ghiandola atriale. In questa regione vi è però una divergenza particolarmente significativa: viene introdotto un sito di scissione, che genera il fattore alfa delle cellule a sacco, costituito da nove amminoacidi, il quale, come visto, inibisce l'eccitazione di quattro neuroni del ganglio addominale.

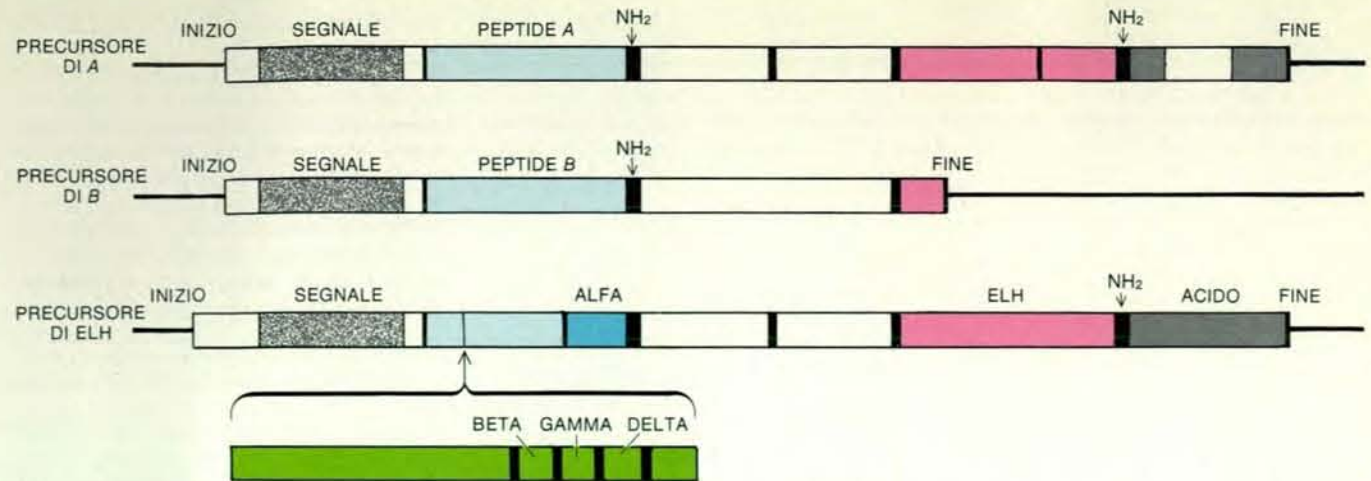
Segue un segmento che presenta un'omologia incompleta con i geni di A e B, dopodiché un sito di scissione segnala - nel precursore - l'inizio del peptide ELH, costituito da 36 amminoacidi. La fine di questo peptide, come nel caso di A e di B, è seguita da un segnale di scissione più transamidazione. Tra la fine della sequenza nucleotidica per il peptide ELH e il codone di fine che pone termine alla traduzione rimangono codoni per 27 amminoacidi, che sono quelli del peptide acido, il quale viene anch'esso secreto dalle cellule a sacco assieme all'ELH; di esso non si conosce ancora il bersaglio.

Vi sono quindi tre geni che hanno in comune tre regioni omologhe: la regione per i peptidi A e B, la regione per il peptide ELH e la regione per il peptide acido. Inoltre, all'interno di ciascuna di queste tre regioni, vi sono in posizioni fisse delle quasi-identità di sequenza. Ciò sottintende che tutti e tre i peptidi hanno avuto origine in un piccolo peptide ancestrale, il cui gene responsabile si è triplicato per generare una proteina più grande composta da almeno tre peptidi. Il gene che codifica per la proteina maggiore si è evidentemente triplicato a sua volta, dando origine a tre geni indipendenti che si sono differenziati allorché si sono specializzati per soddisfare differenti esigenze funzionali. A questo punto possono esservi state duplicazioni di minore entità a carico di alcune regioni dell'uno o dell'altro gene, come viene suggerito dal fatto che i peptidi beta e gamma, che sono adiacenti, sono praticamente identici. Tali eventi hanno presumibilmente reso possibile l'evoluzione di varianti senza una significativa alterazione del gene originale. Le varie versioni del gene possono essere state trasposte in siti diversi del genoma e forse su cromosomi diversi.

La disponibilità di geni clonati per l'ELH ci ha permesso di chiederci, nel corso del nostro lavoro con la McAllister e Kandel, se i geni si esprimono (cioè se sono trascritti in RNA messaggero che viene quindi tradotto in proteina) non solo nelle cellule a sacco e nella ghiandola atriale degli adulti di *Aplysia*, ma anche in altre parti del sistema nervoso di questo mollusco, e di seguire lo sviluppo dei neuroni nei quali i geni si esprimono. Allo scopo ci siamo serviti di due tecniche: l'ibridazione *in situ* e l'immunocitochimica. La prima si basa sul fatto che, dovunque un gene particolare si esprima, si trova l'RNA messaggero da esso trascritto. Si ricava una sottile sezione di tessuto, la si monta su un vetrino portaoggetti per osservazione al microscopio e la si tratta in modo che l'RNA presente nelle cellule sia accessibile all'ibridazione molecolare. Quando si aggiungono al vetrino i geni per l'ELH marcati con un isotopo radioattivo, essi si ibridano con l'RNA messaggero complementare, la cui localizzazione viene rivelata mediante autoradiografia. Dal canto suo, l'immunocitochimica rivela la presenza del peptide ELH. Un anticorpo contro di esso viene aggiunto a un vetrino e si lega al peptide se questo è presente. Un secondo anticorpo, scelto in modo che possa legarsi al primo e associato a un colorante fluorescente, viene a sua volta aggiunto e l'osservazione a luce ultravioletta rivela la localizzazione del peptide nella cellula.

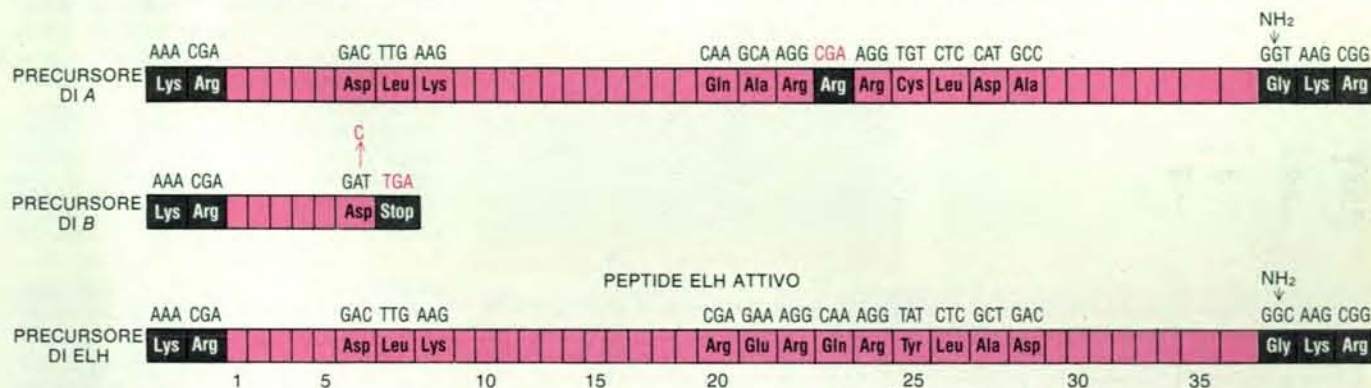
Abbiamo effettuato studi di ibridazione e di immunofluorescenza sull'intero sistema nervoso centrale di *Aplysia*. Le due tecniche scoprono singoli neuroni che esprimono un gene per l'ELH anche in gangli con diverse migliaia di cellule che non esprimono il gene. Le cellule a sacco e la ghiandola atriale sono chiaramente i siti principali di espressione dei geni per l'ELH, ma gli esperimenti mostrano che l'RNA messaggero per l'ELH e lo stesso peptide sono sintetizzati anche nell'ambito di un'ampia rete di cellule nervose, non solo nel ganglio addominale, ma anche in altri due gangli. In esperimenti analoghi, Strumwasser e Arlene Y. Chiu hanno identificato neuroni anche all'esterno degli ammassi di cellule a sacco e questi neuroni sintetizzano l'ELH. Pare che l'ELH possa avere un'estesa funzione di neurotrasmettitore in tutto il sistema nervoso di *Aplysia*.

La presenza di cellule diffuse che producono l'ELH in quattro gangli del sistema nervoso fa sorgere il problema di come le cellule compaiano nel corso dello sviluppo. Mediante ibridazione *in situ* in animali in via di sviluppo, abbiamo potuto conoscere quando i geni per l'ELH si esprimono la prima volta e dove hanno origine i neuroni in cui l'espressione avviene. In tutti gli animali il sistema nervoso si sviluppa come una specializzazione dell'ectoderma, cioè della superficie del corpo dell'embrione. I geni per l'ELH sono attivi molto precocemente negli stadi larvali dello sviluppo di *Aplysia* e in una zona di cellule primitive che tappezzano la parete corporea e sono destinate a diventare neuroni. In seguito, nel corso del-



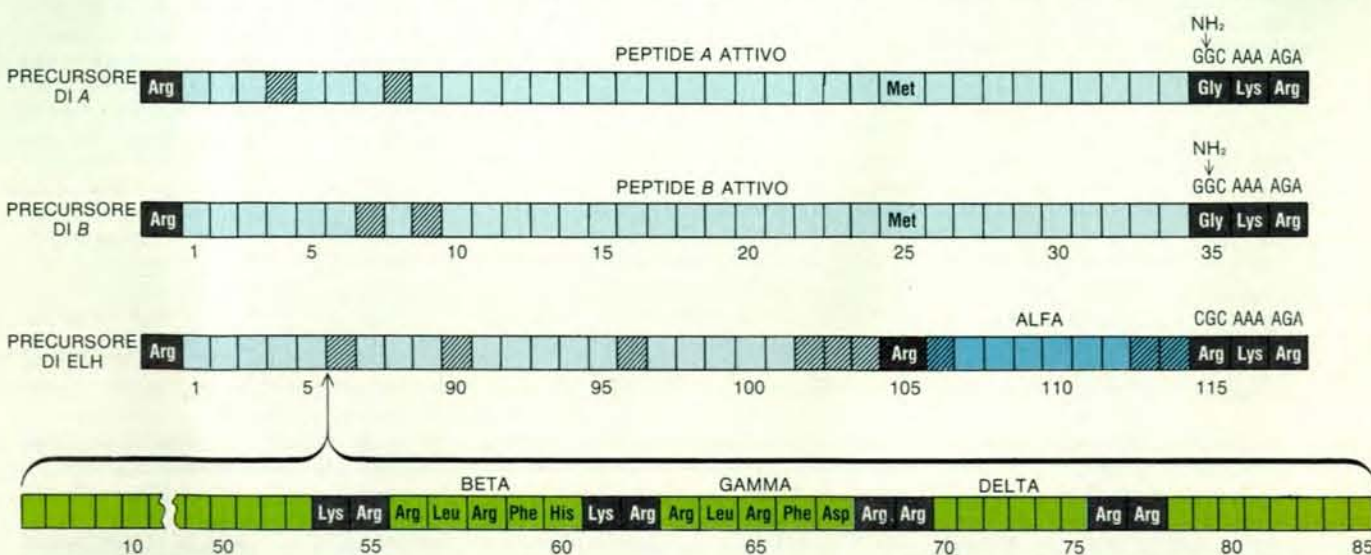
Sono confrontate tre proteine precursore, codificate dai tre membri della famiglia multigenica che comprende il gene per l'ELH. Esse sono scisse in siti specifici (sbarrette nere) e danno origine a peptidi attivi, alcuni dei quali subiscono un'ammidazione (NH₂). Hanno in comune una sequenza segnale (a sinistra) e diverse regioni omologhe che danno

origine, in un precursore o nell'altro, al peptide A o B o al peptide ELH e al peptide acido. Differenze di un unico nucleotide alterano i geni A e B cosicché l'ELH attivo non è sintetizzato. Un inserto di 240 nucleotidi, che codifica per 80 amminoacidi, nel gene per l'ELH interrompe le sequenze per i peptidi A e B e dà origine a molti altri peptidi.



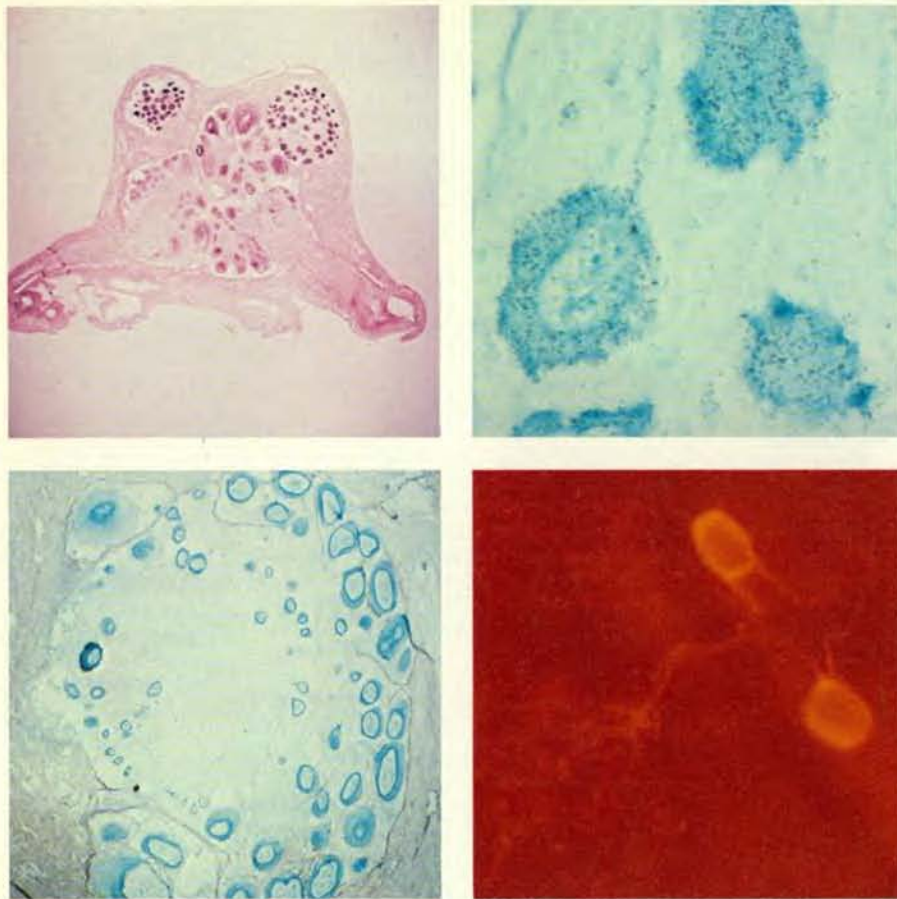
La regione omologa per l'ELH è molto simile in tutti e tre i geni, cioè la sequenza in cui sono disposti i quattro nucleotidi (A, G, C e T) è praticamente la stessa, cosicché i codoni costituiti da tre nucleotidi specificano gli stessi amminoacidi nella maggior parte delle posizioni. (Vengono illustrati qui solo alcuni codoni e amminoacidi tra i più significativi.) Mentre il gene per l'ELH dà origine a un peptide ELH attivo, gli altri due geni no. Nel gene per il precursore del peptide A, la sostituzione

di un nucleotide G con un nucleotide A nella posizione 23 dà origine a una arginina al posto di una glutammina: si genera in questo modo una sequenza di tre arginine e di conseguenza un potenziale sito di scissione. Nel gene per il precursore del peptide B, un nucleotide C scompare per delezione dal sesto codone, modificando il «sistema di lettura» in modo che un codone TGA venga introdotto in posizione 7. Il TGA funge da codone di «fine», che porta a termine la traduzione.



Nel gene per l'ELH un inserto di 240 nucleotidi, che codifica per 80 amminoacidi, altera la regione omologa di quelle presenti nei precursori per i peptidi A e B. Questa viene scissa in corrispondenza dei siti costituiti dalla coppia lisina-arginina e produce tre peptidi: i fattori beta, gamma e delta delle cellule a sacco. (Si noti la quasi identità di beta e gamma, forse dovuta a duplicazione di una corta sequenza di

DNA.) In corrispondenza della posizione 105 della regione omologa, la sostituzione di una arginina nel precursore dell'ELH al posto della metionina presente nei precursori di A e di B crea un sito di scissione, con formazione di un altro peptide attivo: il fattore alfa delle cellule a sacco. A parte l'inserto, i tre precursori sono molto simili. Le caselle tratteggiate indicano i siti dove uno di essi ha un amminoacido diverso.



I siti dove avviene la sintesi dell'ELH sono individuati mediante ibridazione *in situ* con l'RNA messaggero per quel peptide (*in alto a sinistra e a destra, in basso a sinistra*) e il legame degli anticorpi con il peptide stesso (*in basso a destra*). Una sezione del ganglio addominale è stata esposta a una sonda, costituita dal gene per l'ELH e marcata con un isotopo radioattivo (*in alto a sinistra*). La sonda si è ibridata con l'RNA messaggero per l'ELH, la cui localizzazione è stata identificata mediante autoradiografia. I punti scuri indicano l'avvenuta ibridazione nelle cellule a sacco (ammassi rotondeggianti nel margine superiore del ganglio) e inoltre in un singolo neurone verso il centro del ganglio. L'ingrandimento è di circa 17 diametri. Un ingrandimento di circa 400 diametri mostra l'RNA messaggero per l'ELH nel citoplasma di cellule a sacco e nei loro prolungamenti (*in alto a destra*). L'RNA messaggero per l'ELH è scoperto anche in una singola cellula di un differente gruppo di neuroni, il ganglio pleurale (*in basso a sinistra*). In uno studio di immunofluorescenza (*in basso a destra*) un anticorpo contro l'ELH è stato aggiunto a una sezione di ganglio addominale. Poi è stato aggiunto un secondo anticorpo, marcato con un colorante fluorescente, contro l'anticorpo precedente anti-ELH: l'intensa fluorescenza arancione rivela l'ELH nelle due cellule del ganglio addominale e nei prolungamenti.

lo sviluppo, queste cellule abbandonano la parete corporea e migrano, strisciando lungo briglie di tessuto connettivo, fino a raggiungere le loro localizzazioni definitive nell'insieme del sistema nervoso. Può darsi che un singolo neurone primitivo si divida e generi un gruppo di cellule che producono ELH, cellule che migrano non solo verso le cellule a sacco, ma anche verso siti sparsi in tutto il sistema nervoso.

In *Aplysia* perlomeno tre geni codificano per numerosi peptidi neuroattivi che hanno ruoli nei circuiti che controllano un repertorio comportamentale complesso, ma stereotipo. Nel cervello dei mammiferi sono stati anche identificati numerosi peptidi che, a quanto pare, mediano comportamenti specifici (si veda l'articolo *I neuropeptidi* di Floyd E. Bloom in «Le Scienze», n. 160, dicembre 1981). Quali proprietà dei neuropeptidi

e, in particolare, quali caratteristiche del loro modo di sintesi li rendono adatti come mediatori del comportamento?

Il potenziale comportamentale di un organismo è almeno in parte dettato da una rigida rete di cellule nervose di connessione. Buona parte della comunicazione tra neuroni si svolge localmente e viene mediata da un neurotrasmettitore, come l'acetilcolina o la noradrenalina, che viene prodotto da un neurone, attraverso lo stretto spazio sinaptico e produce un contatto da punto a punto con un altro neurone. Anche i peptidi neuroattivi possono funzionare localmente come neurotrasmettitori. Essi possono però essere secreti nel sistema circolatorio, fungendo da neuroormoni che agiscono su parecchi bersagli distanti e generano così varie attività ben distinte. Pertanto i peptidi rendono disponibile una ulteriore rete di comunicazione che integra la rete di co-

municazione, già costituita, delle cellule nervose interconnesse.

La diversità dei siti bersaglio di un neuropeptide può permettere al peptide di coordinare gli eventi fisiologici con particolari comportamenti. Per esempio, l'iniezione del peptide angiotensina II stimola nei vertebrati un bisogno spontaneo di bere agendo sui neuroni dell'ipotalamo. Il peptide agisce anche indirettamente sul rene per promuovere il riassorbimento del sodio e dell'acqua nel circolo sanguigno al posto della loro escrezione. Ambedue questi effetti ben diversi servono a reidratare l'animale. L'ELH costituisce un altro esempio. In *Aplysia* agisce localmente eccitando neuroni specifici nei gangli che possono provocare comportamenti come l'ondeggiamento del capo e cambiamenti fisiologici come un aumento del battito cardiaco. Nel contempo interviene a distanza, provocando una contrazione del gonodotto e l'espulsione delle uova.

Diversi peptidi di *Aplysia*, interessati nel processo di deposizione delle uova, vengono staccati da un unico precursore che è una poliproteina. È noto che esistono alcuni altri neuropeptidi che vengono sintetizzati nello stesso modo. La via che passa attraverso la poliproteina sembrerebbe avere molti vantaggi. Tanto per cominciare, essa fornisce un meccanismo semplice per il controllo sia della sintesi sia della secrezione dei peptidi. Differenti peptidi, codificati da un unico gene, possono venire sintetizzati simultaneamente sotto il controllo di un unico agente regolatore. Inoltre, i vari piccoli peptidi di una poliproteina possono, dopo il distacco dal precursore, venire ammassati nella stessa vescicola per essere poi secreti simultaneamente dalla cellula che li ha sintetizzati, mediante un singolo stimolo.

La produzione di molteplici peptidi da un unico precursore offre anche una soluzione a un problema di numero: il numero di geni in un genoma animale è semplicemente non abbastanza grande da specificare la varietà di comportamenti che una specie esibisce. Il potenziale d'informazione di un singolo gene può essere accresciuto se la proteina prodotta da quel gene viene scissa in maniera diversa in cellule diverse o in risposta a stimoli diversi. Un esempio di trasformazione alternativa di una proteina è stato notato da Edward Herbert dell'Università dell'Oregon e da James L. Roberts del Columbia College of Physicians and Surgeons, oltre che da Richard E. Mains e Betty A. Eipper della Johns Hopkins University School of Medicine. Essi hanno trovato che un unico precursore viene trasformato e dà l'ormone adrenocorticotropo (ACTH) nel lobo anteriore dell'ipofisi e un peptide simile all'endorfina nel lobo posteriore. Da parte nostra, è stato osservato in una singola cellula uno schema di scissione del precursore dell'ELH. In teoria, dati i siti di scissione che abbiamo notato, questa proteina potrebbe essere scissa in differenti modi e potrebbe dar luogo a più di 2000 differenti combinazioni di peptidi; ciascuna combinazione



Libri firmati JACKSON



Nicole Bréaud-Pouliquen
LA PRATICA DELL'APPLE

"Il Sistema APPLE II", il "BASIC Applesoft", il disegno e la grafica: arricchiti da esempi e esercizi. 130 pagine L. 10.000
Codice 341D

F. Franceschini - F. Paterlini
Voi e il vostro Commodore 64

Uno strumento fondamentale per la comprensione e programmazione del Commodore 64. Con consigli, programmi testati, glossario e utili accenni di BASIC. 256 pagine B L. 22.000 Codice 347

Alan Miller
PROGRAMMI SCIENTIFICI IN PASCAL

Un'opera base per chi desidera costruirsi una "libreria" di programmi in grado di risolvere i più frequenti problemi scientifici e ingegneristici. 372 pagine L. 25.000
Codice 554P

Carmine Elefante
L'home computer TI/99-4A

Il BASIC, il BASIC Esteso e il microprocessore dell'home computer della T.I. Con programmi di utilità e svago. 192 pagine L. 15.000
Codice 343B

Giacomino Baisini - Giò Federico Baglioni
IL FORTH PER VIC 20 E CBM 64

La programmazione in FORTH e la sua implementazione sul Commodore VIC 20 e CBM 64. 150 pagine L. 11.000
Codice 527B

Franco Filippazzi - Giulio Occhini
VOI E L'INFORMATICA

L'opera che il manager moderno non può ignorare. In 100 tavole: gli strumenti dell'informatica, l'informatica e l'Azienda, realtà e prospettive tecnologiche... 116 pagine L. 15.000
Codice 526A

Roland Dubois
CAPIRE I MICROPROCESSORI

Un fantastico viaggio alla scoperta del "cervello" elettronico: la funzione del microprocessore, delle memorie ROM e RAM, delle interfacce... 126 pagine L. 10.000
Codice 342A

Gaetano Marano
77 PROGRAMMI PER SPECTRUM

Dalla Grafica alla Business Grafica, dalla musica alle animazioni, dai giochi all'elettronica... tutte le possibilità offerte dal più piccolo dei computer. 150 pagine a colori L. 16.000
Codice 555A

Rita Bonelli-Daria Gianni
ALLA SCOPERTA DEL VIC 20

Un testo chiave per imparare a conoscere e usare uno dei Personal del momento. 308 pagine L. 22.000
Codice 338D
Cassetta Programmi L. 15.000
Floppy Programmi L. 25.000

La Biblioteca che fa testo

In busta chiusa, e senza impegno, inviate questo coupon a:
Gruppo Editoriale Jackson - Via Rosellini, 12 - 20124 Milano

Desidero ricevere gratuitamente il Catalogo Generale della Biblioteca Jackson e informazioni sulle 10 Riviste specialistiche da voi pubblicate. (allego L. 1.000 in francobolli per contributo spese di spedizione)

Desidero ricevere contrassegno il/i volume/i

(pagherò al ricevimento L.
più L. 2.000 per contributo spese di spedizione)

Nome _____ Cognome _____

Via _____

CAP _____ Città _____

potrebbe poi attivare un differente schema comportamentale. Il potenziale di questa diversità è ancora maggiore perché i geni per l'ELH costituiscono addirittura una famiglia di geni che si esprimono in tessuti differenti.

La via che passa attraverso la poliproteina offre una flessibilità temporale, permettendo ai vari peptidi di rimanere attivi per diversi periodi di tempo. La stabilità di un peptide può essere influenzata da certe modificazioni postsintetiche, come la transamidazione, e persino dalla dimensione (perché i piccoli peptidi non si possono ripiegare facilmente in strutture compatte e pertanto vengono probabilmente degradati più rapidamente di quelli di maggiori dimensioni). Sia il tipo di ammidazione sia la dimensione dei vari peptidi che intervengono nel processo di deposizione delle uova sono conformi con le loro esigenze funzionali. I peptidi (ELH e A e B), che possono fungere da ormoni, sono soggetti a una costante ammidazione e sono inoltre più lunghi; pertanto la loro vita attiva viene presumibilmente allungata. I peptidi (come i fattori alfa e beta delle cellule a sacco), che hanno apparentemente solo un'azione locale come neurotrasmettitori, sono più corti e non sono soggetti ad ammidazione. Vengono presumibilmente degradati più rapidamente, come la maggior parte degli altri neurotrasmettitori.

L'organizzazione dei geni che codificano per le poliproteine offre un sorprendente grado di flessibilità evolutiva. La disseminazione di sequenze che codificano per una serie di peptidi attivi all'interno di un gene, che codifica anche per una proteina non funzionale, lascia spazio per l'evoluzione di nuovi peptidi attivi senza alterare l'assetto originale. Per esempio, il gene per l'ELH ha un inserto di 240 nucleotidi che non è presente nei geni omologhi che si esprimono nella ghiandola atriale. Tale inserto codifica per tre peptidi (due dei quali riflettono apparentemente una piccola duplicazione interna), estendendo così la schiera dei peptidi coordinati della poliproteina, senza influire sulla sintesi dell'ELH attivo.

Infine, lo stesso peptide può essere incorporato in parecchi precursori di diversa natura, codificati da differenti geni. Si consideri l'ondeggiamento del capo in *Aplysia*: un caratteristico movimento di questo tipo ha luogo sia mentre l'animale si alimenta sia mentre depone le uova. Lo stesso peptide, o gli stessi peptidi, potrebbe indurre così la stessa componente comportamentale (cioè l'ondeggiamento del capo) in due contesti molto diversi. A questo fine il peptide (o i peptidi) che ne è responsabile può essere contenuto in codice in qualche altro gene, per esempio in quello interessato nel comportamento alimentare, oltre che nel gene per l'ELH. In questo modo, i comportamenti complessi potrebbero nascere dalla combinazione di semplici unità di comportamento, e ognuna di queste unità sarebbe mediata da un unico peptide o da un piccolo numero di peptidi.

nuovidea

Materia eccitonica

Gli eccitoni, prodotti nei semiconduttori dalla ricombinazione degli elettroni di conduzione con le buche cariche positivamente, possono formare molecole e liquidi, forse anche una nuova fase della materia

di André Mysyrowicz e James P. Wolfe

L'interazione della luce con la materia solida è un fenomeno di fondamentale importanza per esplorare le basi quantomeccaniche della natura. La scoperta di Albert Einstein che l'energia luminosa è trasportata da pacchetti di radiazione quantizzati, oggi chiamati fotoni, si basava sull'osservazione di una di tali interazioni: la fotoemissione di elettroni dalla materia. Più recentemente si è dimostrato che la distribuzione degli elettroni tra i loro stati energetici quantizzati in un solido può modificarsi in molti modi in risposta all'energia trasportata dai fotoni incidenti. Una dettagliata comprensione di questi effetti ha conseguenze significative sia per la tecnologia sia per la scienza. In particolare, le proprietà fotoelettriche di cristalli semiconduttori, che oggi vengono fatti crescere sinteticamente con purezze elevatissime nella manifattura di componenti elettronici, sono già state sfruttate per la produzione di dispositivi quali rivelatori di luce a stato solido, diodi emettitori di luce e celle solari. Con l'approfondimento delle conoscenze sulle interazioni tra elettroni e fotoni in un solido si potrebbero realizzare progetti nuovi e più efficienti per tali dispositivi.

Uno dei problemi più affascinanti nello studio dei solidi è offerto dai bellissimi effetti di assorbimento della luce in un cristallo semiconduttore. Per esempio, se si illumina con luce laser verde la superficie di un cristallo del semiconduttore ossido rameoso (Cu_2O), il cristallo emette un bagliore rosso, con una frequenza inferiore a quella della luce verde. L'energia di un fotone è proporzionale alla sua frequenza, perciò l'energia di ogni fotone verde incidente è maggiore dell'energia di ogni fotone rosso emesso. Si pone perciò la domanda: in che modo l'energia dei fotoni verdi viene convertita nell'energia dei fotoni rossi?

Quasi trent'anni fa si scoprì che l'energia dei fotoni incidenti si può convertire all'interno del cristallo in entità neutre di breve vita, chiamate eccitoni. In termini semplici, l'eccitone è un quanto di eccitazione elettronica del cristallo che viene prodotto dall'assorbimento di un fotone;

la luce rossa emessa dal cristallo è l'energia liberata quando cessa lo stato di eccitazione elettronica del cristallo. Ciò che rende interessante lo stato eccitato è che può essere descritto come una particella. In effetti, l'eccitone assomiglia all'atomo di idrogeno: è formato da due portatori con carica opposta legati uno all'altro dall'attrazione elettrostatica. Nell'atomo di idrogeno la carica positiva è un protone circondato dall'elettrone carico negativamente. Nell'eccitone la carica positiva ha una massa 1000 volte inferiore a quella del protone.

Per poter capire le conseguenze della scoperta dell'eccitone, immaginiamo il mondo all'interno di un cristallo semiconduttore, nel quale il leggerissimo eccitone ha assunto il ruolo dell'atomo di idrogeno. Dall'analogia nascono molti interrogativi. L'eccitone può propagarsi liberamente nel cristallo come fa un atomo di idrogeno libero in un gas? Due o più eccitoni possono combinarsi per formare una molecola? Gli atomi «eccitonici» o le molecole formate da essi possono dar luogo a una fase liquida o solida? Possono essere create nuove e più esotiche fasi di materia eccitonica condensata? Lo studio degli eccitoni all'interno dei cristalli fornisce risposte a tali domande.

In che modo gli eccitoni vengono creati dalla luce in un cristallo? O, ancora più importante, perché il cristallo assorbe la luce? Per rispondere a queste domande, esaminiamo brevemente gli elettroni nel cristallo, dal momento che sono tali particelle a rispondere alla luce incidente. Anziché presentare valori di energia ben distinti come gli elettroni in un atomo isolato, gli elettroni occupano in un cristallo larghe bande di energie permesse. Le bande sono separate una dall'altra da intervalli di energia proibiti a un elettrone. Ogni banda è formata da stati discreti i cui livelli energetici associati sono talmente fitti da non poter essere risolti. Il numero di stati in una banda è circa uguale al numero di atomi nel cristallo; per un cristallo di dimensioni macroscopiche tale numero è dell'ordine di 10^{23} . Quindi è

ragionevole pensare che, all'interno di una banda, le energie elettroniche siano distribuite in un continuo di stati.

Gli elettroni riempiono gli stati permessi di una banda di energia dal basso verso l'alto, allo stesso modo in cui l'acqua riempie una bottiglia. La distribuzione degli elettroni tra le bande di energia di un solido determina se questo solido è un metallo, un semiconduttore oppure un isolante. In un metallo, la banda più alta occupata è riempita solo parzialmente con elettroni. La corrente elettrica fluisce facilmente in un metallo perché il cambiamento di energia cinetica degli elettroni necessario per il passaggio di corrente si ottiene immediatamente con uno spostamento della distribuzione degli elettroni nella banda parzialmente vuota. Invece, in un semiconduttore o in un isolante a basse temperature, la banda più alta occupata, chiamata banda di valenza, è completamente riempita di elettroni come una bottiglia piena fino al tappo. La successiva banda di energia è chiamata banda di conduzione ed è vuota alle basse temperature; essa è separata dalla banda di valenza da un intervallo energetico «proibito». Non vi può essere alcuna corrente elettrica in una banda di stati energetici elettronici completamente piena. L'intervallo proibito in un semiconduttore è minore che in un isolante, pertanto in un semiconduttore si genera una certa conducibilità elettrica quando alcuni elettroni di valenza vengono eccitati termicamente ad attraversare l'intervallo proibito. A basse temperature, invece, un semiconduttore è altamente isolante.

L'assorbimento di un fotone da parte di un cristallo semiconduttore provoca il passaggio di un elettrone da una banda di valenza piena a una banda di conduzione vuota. In questo processo si deve conservare l'energia, e la conservazione dell'energia impone che, affinché il fotone possa essere assorbito, la sua energia debba essere maggiore o uguale a quella dell'intervallo proibito. Se l'energia del fotone è inferiore a questa soglia, nessun elettrone di valenza può assorbire la luce, perché altrimenti dovrebbe occupare l'intervallo

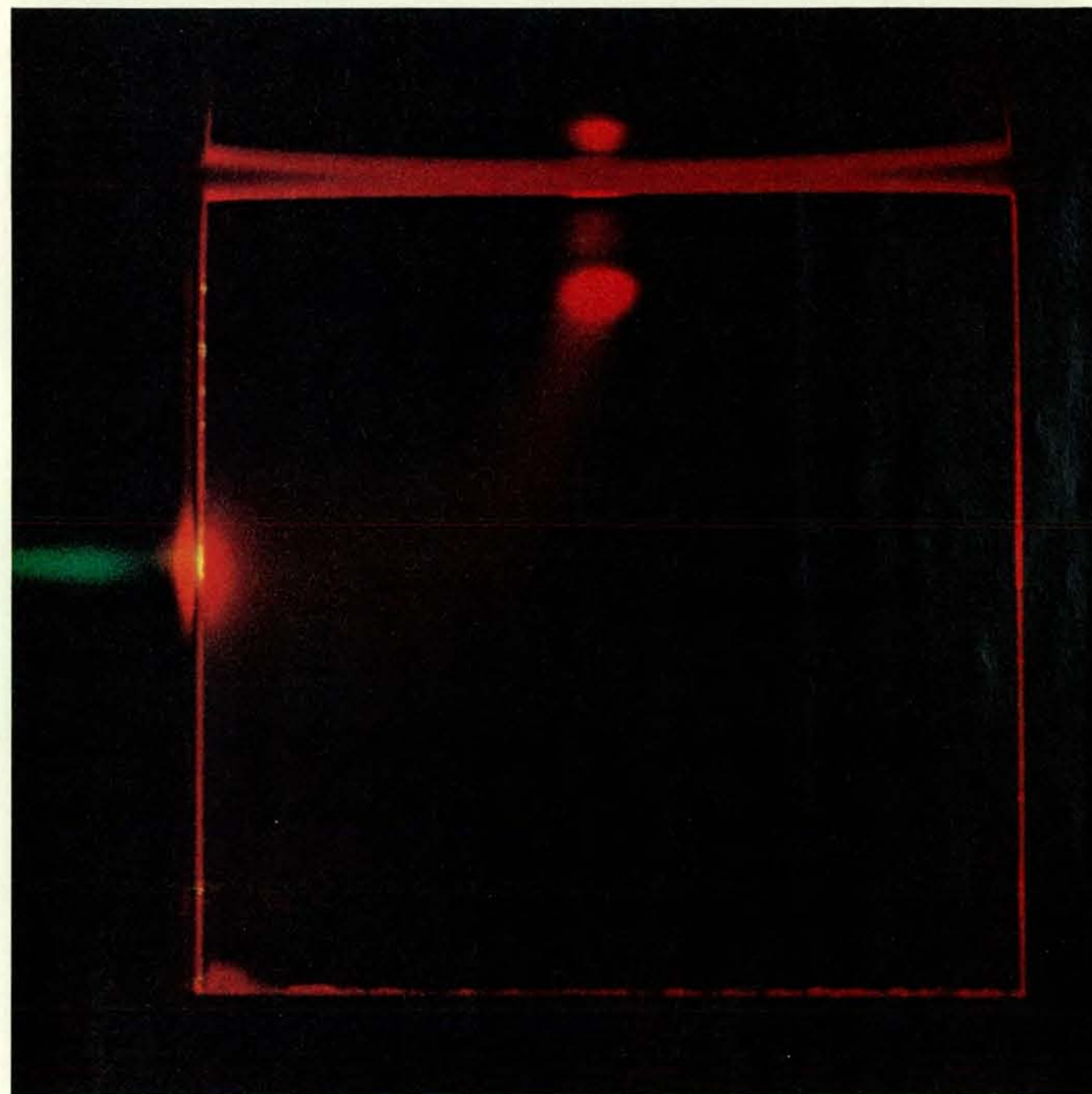
proibito. In assenza di altri effetti, un fotone con energia inferiore all'energia dell'intervallo proibito si propaga nel cristallo senza attenuazione.

Quando un elettrone viene espulso dalla banda di valenza da un fotone, lascia dietro di sé uno stato vuoto, allo stesso modo in cui si forma una bolla quando si

toglie una goccia da una bottiglia piena di liquido. Questo stato vuoto, o vacanza di un elettrone nella banda di valenza, si chiama buca. La buca può essere interpretata come una regione di carica totale positiva nel reticolo cristallino, e sotto molti aspetti le sue proprietà ricordano quelle di una particella ordinaria. Per

esempio, un elettrone da una regione vicina del reticolo si può spostare nella prima regione e la rende neutra, ma in tal modo la seconda regione acquista la carica positiva in eccesso. La regione di carica positiva si può propagare nel reticolo come se fosse una vera particella.

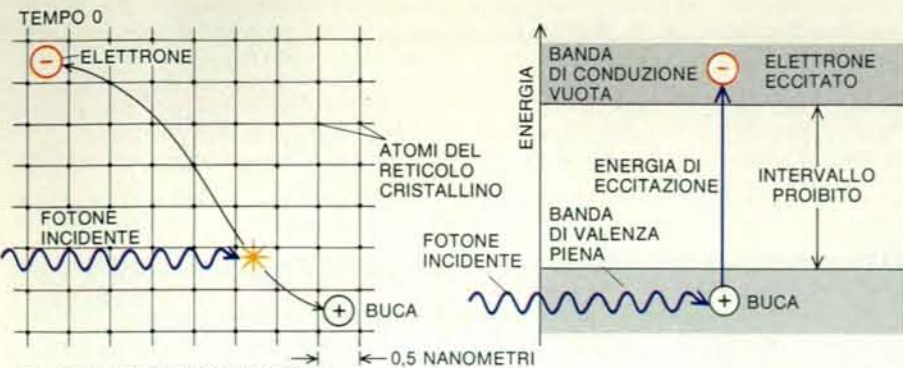
Una volta creati nelle corrispondenti



Un gas di eccitoni viene prodotto in un cristallo solido di ossido rameoso (Cu_2O) da un intenso fascio di luce verde emesso da un laser a ioni argo. Gli elettroni associati agli atomi del cristallo vengono eccitati dall'energia laser e vagano liberamente nel cristallo. Le buche, regioni di carica complessiva positiva, lasciate indietro dagli elettroni eccitati, si propagano a loro volta liberamente come elettroni carichi positivamente. Il cristallo, che misura due millimetri di lato e 1,5 millimetri di spessore, è immerso in un bagno di elio liquido a due gradi al di sopra dello zero assoluto. A questa temperatura tutti gli elettroni e le buche liberi possono formare uno stato legato a vita breve chiamato eccitone. La luce rossa emessa dal cristallo segnala la presenza di eccitoni; essa è

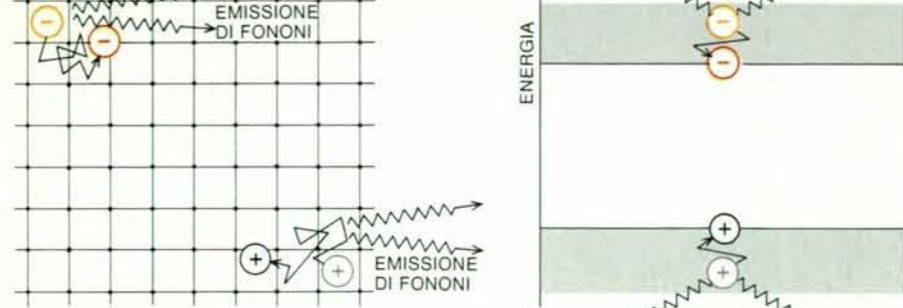
causata dalla ricombinazione degli elettroni e delle buche dopo pochi microsecondi dalla formazione dello stato legato. Nella parte superiore della fotografia un'asta arrotondata è premuta contro la superficie del cristallo, il che fa diminuire l'energia degli eccitoni che si trovano proprio sotto al punto di sollecitazione. Gli eccitoni prodotti nella parte sinistra del cristallo migrano rapidamente verso la regione a bassa energia, che appare come un punto rosso brillante perché gli eccitoni si concentrano appunto in quella sede. Il punto viene riflesso dall'asta di sollecitazione e la riflessione viene vista appena al di sopra del cristallo. La fotografia è stata eseguita da David P. Trauernicht e dagli autori di questo articolo all'Università dell'Illinois a Urbana-Champaign.

ECCITAZIONE NON RISONANTE



RILASAMENTO VIBRAZIONALE

TEMPO 1-100 PICOSECONDI



FORMAZIONE DI UN ECCITONE

TEMPO 100 PICOSECONDI-1 NANOSECONDO



RICOMBINAZIONE

TEMPO 1 NANOSECONDO-1 MILLISECONDO



L'eccitazione non risonante può dare origine a eccitoni in un semiconduttore quando l'energia di eccitazione è maggiore dell'energia dell'intervallo proibito, che è la differenza di energia tra gli elettroni più energetici nel cristallo semiconduttore non eccitato e lo stato quantomeccanico meno energetico al quale essi possono essere portati. Nel diagramma il moto delle particelle è illustrato a sinistra, mentre a destra sono mostrate le transizioni energetiche associate a ciascun moto. Lo stato iniziale di ogni fase della vita dell'eccitone è in colore arancione e grigio, mentre lo stato finale è in rosso e nero. Dopo l'eccitazione iniziale l'elettrone e la buca liberi si riportano rapidamente a un'energia uguale all'energia dell'intervallo proibito emettendo fononi, cioè quanti di energia vibrazionale, verso il cristallo circostante. Essi possono poi legarsi per formare un eccitone in uno dei livelli energetici discreti permessi dalla meccanica quantistica. Un eccitone formato in uno stato energetico eccitato può cadere in uno stato inferiore prima che l'elettrone e la buca si ricombinino. L'energia relativa di ogni transizione è indicata da un colore: le transizioni a bassa energia sono accompagnate da fotoni di bassa frequenza, rappresentati schematicamente in rosso, mentre le transizioni di energia più elevata lo sono da fotoni di frequenza maggiore, indicati in blu o in violetto. La distanza tra i livelli energetici eccitonici è stata esagerata per chiarezza.

bande, sia l'elettrone sia la buca si portano rapidamente verso il loro stato di minima energia. L'elettrone eccitato cade in fondo alla banda di conduzione, mentre la buca sale verso l'alto, come una bolla in una bottiglia d'acqua, verso la parte superiore della banda di valenza. La buca sale perché gli elettroni che rimangono nella banda di valenza la spostano mentre cadono nel più basso stato energetico possibile. Poi, entro un nanosecondo (10^{-9} secondi), l'elettrone e la buca possono diminuire ulteriormente la loro energia totale legandosi e formando così un eccitone.

È ben noto che l'elettrone e il protone nell'atomo di idrogeno possono assumere uno solo tra i valori di una serie di livelli energetici discreti. Allo stesso modo, esiste un insieme di valori dell'energia consentiti all'eccitone, corrispondenti ai diversi stati eccitati della coppia legata elettrone-buca. I livelli energetici eccitonici convergono verso un limite ad alta energia, coincidente con l'energia dell'intervallo proibito del semiconduttore. Tuttavia, l'energia di legame dell'eccitone è tipicamente da 100 a 1000 volte inferiore all'energia di legame dell'atomo di idrogeno.

Il legame debole implica che la distanza media tra elettrone e buca possa essere di uno o due ordini di grandezza maggiore del diametro dell'atomo di idrogeno. In effetti, l'orbita eccitonica può spaziare su molte distanze del reticolo cristallino. Per esempio, l'energia di legame dell'eccitone formato in un semiconduttore come il germanio è di soli 0,0041 elettronvolt e il suo raggio è di 13 nanometri (1 nanometro è pari a 10^{-9} metri). (In confronto, l'energia di legame dell'atomo di idrogeno è di 13,6 elettronvolt e il suo raggio è di 0,05 nanometri; la distanza tipica tra atomi in un cristallo varia da 0,3 a 0,5 nanometri.) Alle temperature ordinarie le vibrazioni termiche del reticolo sono sufficientemente violente da vincere il debole legame eccitonico. Per produrre eccitoni in quantità significativa è solitamente necessario raffreddare il semiconduttore fino a poche decine di gradi dallo zero assoluto.

Come si possono rivelare i livelli energetici eccitonici? Un metodo è quello di misurare la quantità di luce che viene trasmessa attraverso un cristallo semiconduttore per ogni colore della luce incidente. Nello spettro della luce trasmessa dal cristallo si osserva un certo numero di righe scure. Ognuna di queste righe è prodotta dalla conversione di luce in eccitoni in uno stato energetico distinto. I fotoni che provocano le transizioni sono assorbiti dal cristallo e quindi la posizione delle righe scure nello spettro della luce trasmessa dà il colore e quindi l'energia della luce assorbita.

Un classico esempio di uno spettro di assorbimento che fornisce le strutture energetiche degli eccitoni è lo spettro dell'ossido rameoso. Si osservano in esso ben otto diverse righe di assorbimento eccitonico, che formano una serie convergente simile alle righe dello spettro dell'idrogeno monoatomico. Lo spettro di assorbimento

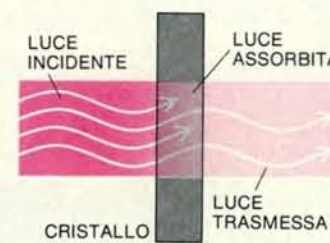
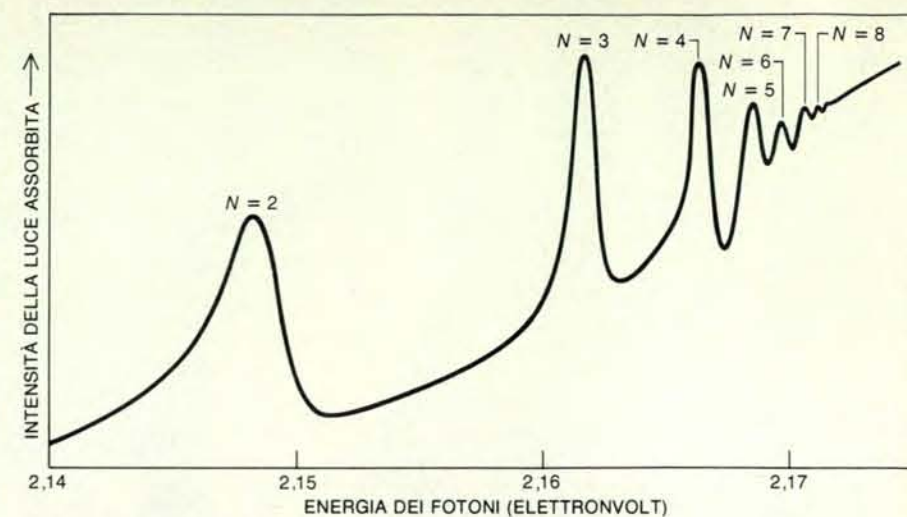
fornisce l'energia di legame dell'eccitone nell'ossido rameoso: è di 0,14 elettronvolt, corrispondente a un raggio dell'eccitone di 1 nanometro.

Un altro modo di ricavare informazioni sugli eccitoni sta nel rilevare la luce che essi emettono. Un eccitone, una volta formato, non vive in eterno. L'elettrone finisce col ricadere nella buca, viene emesso un fotone e l'eccitone scompare. Il processo si chiama ricombinazione. La probabilità di ricombinazione varia notevolmente da un cristallo all'altro e dipende dalle particolari caratteristiche della banda di valenza e della banda di conduzione. Nei cristalli puri la vita di un eccitone varia da qualche nanosecondo a qualche millisecondo.

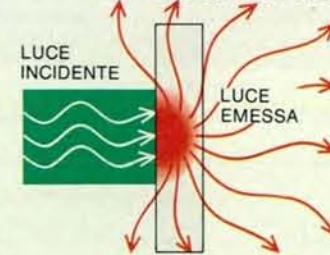
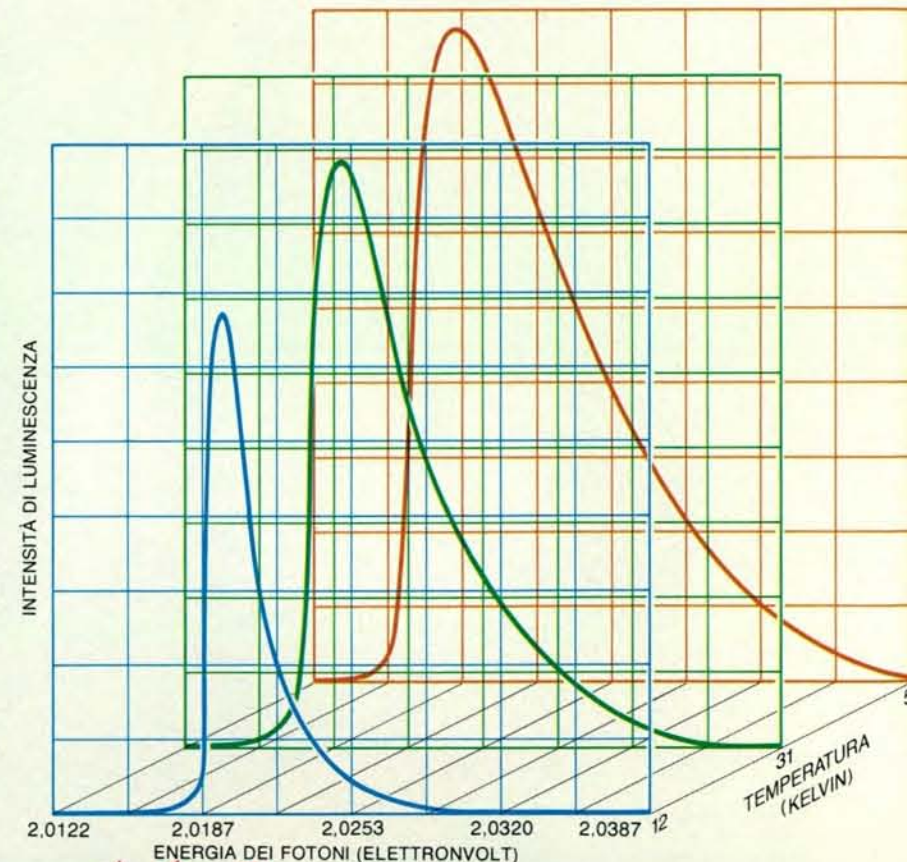
Analizzando lo spettro dei fotoni di luminescenza emessi nella ricombinazione si può ricostruire la distribuzione energetica di un insieme di eccitoni. Supponiamo, per esempio, che un cristallo di ossido rameoso sia illuminato con luce verde, la cui energia è maggiore dell'energia dell'intervallo di banda. Ricordiamo che in tali circostanze il cristallo emette una caratteristica luce rossa. L'emissione è causata dalla ricombinazione di eccitoni nello stato fondamentale, lo stato eccitonico di minima energia. La luce non viene però emessa a una sola frequenza, ma comprende un intervallo di frequenze la cui ampiezza aumenta con la temperatura del cristallo. L'intervallo delle frequenze emesse corrisponde all'intervallo delle energie eccitoniche presenti nel cristallo all'istante della ricombinazione. Dal momento che l'energia totale dell'eccitone è determinata dall'energia dell'intervallo proibito, dall'energia di legame e dall'energia cinetica associate al moto del centro della massa dell'eccitone, l'intervallo delle frequenze emesse dà la distribuzione dell'energia cinetica e quindi la distribuzione di velocità degli eccitoni.

Lo spettro di luminescenza dell'ossido rameoso mostra inequivocabilmente che gli eccitoni devono essere in moto all'interno del cristallo. La distribuzione delle velocità eccitoniche è in buon accordo con la distribuzione delle velocità degli atomi ordinari in un gas a bassa densità. In altri termini, alle basse densità un gruppo di eccitoni si può interpretare come un gas di particelle confinato all'interno di un mezzo solido!

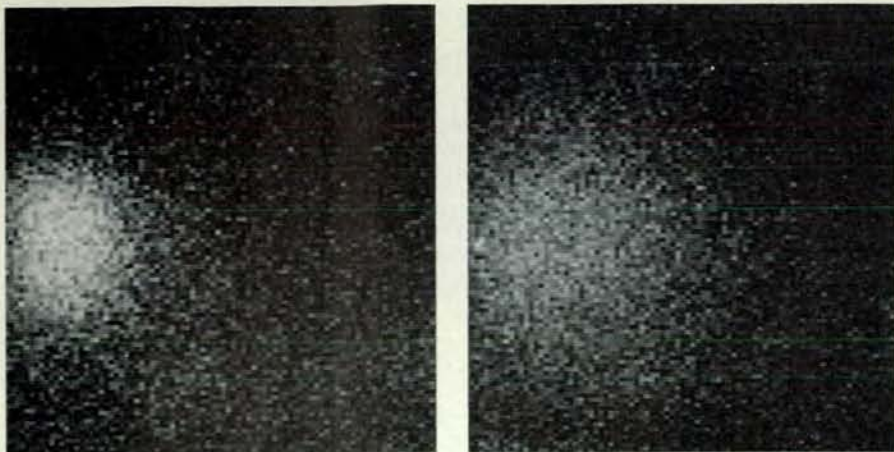
Per poter capire il moto di un gas di eccitoni, si pensi in primo luogo a un gas visibile, quale il vapore che fuoriesce dal becco di una teiera quasi al punto di ebollizione. Il vapore si diffonde dal becco nell'aria circostante e finisce col disperdersi. Allo stesso modo, gli eccitoni creati da fotoni sulla superficie di un cristallo migrano nel cristallo e finiscono col ricombinarsi. Gli eccitoni in migrazione non trasportano carica elettrica visto che sono neutri, ma trasportano energia, che si libera allorché l'elettrone e la buca si ricombinano. È il caso di notare che gli eccitoni si osservano anche nella materia vivente: la migrazione di eccitoni nella



L'eccitazione risonante degli eccitoni è causata dall'assorbimento di un fotone la cui energia corrisponde esattamente a uno dei livelli energetici eccitonici. Qui è rappresentato lo spettro di assorbimento dell'ossido rameoso cristallino. I picchi dello spettro, che vengono indicati da vari numeri N , corrispondono ai livelli energetici degli eccitoni, che sono prodotti nel cristallo dall'energia dei fotoni della luce incidente. Esperimenti di questo genere furono eseguiti per la prima volta da Evgenii F. Gross dell'Università di Leningrado e da Serge Nikitine dell'Università di Strasburgo.



Lo spettro di luminescenza emesso quando gli eccitoni si ricombinano dal loro stato fondamentale nell'ossido rameoso consente di ricavare la distribuzione delle energie cinetiche degli eccitoni subito prima della ricombinazione. L'energia di ricombinazione è la somma dell'energia dello stato fondamentale e dell'energia cinetica dell'eccitone. La larghezza dello spettro diminuisce con l'abbassarsi della temperatura del cristallo, il che vuol dire che si restringe anche la distribuzione delle energie cinetiche. Dati di Daniele Hulín, André Antonetti e A. Mysyrowicz dell'Università di Parigi.



È qui mostrata in due fasi la diffusione degli eccitoni originati nei pressi della superficie di un cristallo di ossido rameoso. Quando il gas di eccitoni si espande nel cristallo, alcuni eccitoni si ricombinano emettendo una luminescenza caratteristica. L'intensità di questa luminescenza è stata misurata da un fotomoltiplicatore per ogni piccolo elemento dell'immagine nelle due figure sopra riportate. Sulla sinistra è illustrata l'espansione 200 nanosecondi (miliardesimi di secondo) dopo l'eccitazione da parte di un fascio laser in corrispondenza dell'estremità sinistra del cristallo; sulla destra è mostrata l'espansione dopo 600 nanosecondi. La costante di diffusione, che misura la rapidità con la quale il gas di eccitoni si espande nel cristallo, si può ricavare dalle immagini; a causa della piccola massa dell'eccitone, la costante di diffusione di quest'ultimo è maggiore di quella di qualsiasi gas atomico o molecolare conosciuto. Le immagini sono state ottenute da Trauernicht.

materia vivente può essere un importante meccanismo di trasporto energetico per reazioni fotochimiche qual è, ad esempio, la fotosintesi.

Che aspetto ha, su scala microscopica, il moto di un gas di eccitoni? Lo spettro di luminescenza degli eccitoni indica che essi si trovano in equilibrio termico con gli atomi del cristallo. Per raggiungere l'equilibrio devono essere in grado di scambiare la loro energia cinetica con l'energia vibrazionale del reticolo. In altri termini, ogni eccitone deve venire diffuso dagli atomi in vibrazione termica e quindi è prevedibile che esso sia sottoposto a un moto browniano casuale. Si può dimostrare che la distanza radiale media, percorsa da una tale particella in un tempo t è proporzionale alla radice quadrata di t ; di conseguenza la sezione trasversale media di una nube di gas in espansione da un punto centrale è proporzionale al tempo di diffusione trascorso. La costante di proporzionalità è chiamata costante di diffusione.

In che modo si può misurare la costante di diffusione per un gas di eccitoni? L'idea sperimentale fondamentale è quella di creare gli eccitoni in un punto della superficie di un cristallo con un impulso di luce laser per esempio della durata di 100 nanosecondi. Si esegue poi una «istantanea» della luminescenza dell'eccitone in successivi istanti durante l'espansione del gas. Il «tempo di esposizione» necessario per una ripresa di questo tipo è circa 10 000 volte inferiore a quello più breve ottenibile con una comune macchina fotografica, e la quantità di luce emessa dagli eccitoni in un tale intervallo di tempo è modesta. Quindi, per poter effettuare l'istantanea, è necessario contare i singoli fotoni di luminescenza con un fotomoltiplicatore.

Gli impulsi dei fotoni vengono selezionati elettronicamente da intervalli di 100 nanosecondi centrati su vari tempi dopo ogni impulso laser. Il procedimento viene ripetuto circa un milione di volte; ogni impulso laser crea una nuova nube di eccitoni che si diffonde e decade prima che venga prodotto l'impulso successivo. Per ogni impulso il numero di fotoni emesso da una piccola regione del cristallo viene registrato mediante un calcolatore, e così, regione per regione, si costruisce un'immagine della luminescenza proveniente dalle diverse regioni del cristallo.

Si può visualizzare su un terminale video l'immagine risultante della luminescenza dell'eccitone per un dato intervallo di tempo successivo all'impulso laser. Quando si rappresenta graficamente l'area della sezione trasversale della regione brillante di luminescenza in funzione del tempo, si trova che essa aumenta in modo direttamente proporzionale al tempo trascorso dopo l'impulso. Tale relazione mostra che gli eccitoni si diffondono come un gas di atomi rarefatto. Però, la costante di diffusione per un gas eccitonico è molto maggiore che per un gas atomico ordinario. Per esempio, nell'ossido rameoso la costante di diffusione è di 1000 centimetri quadrati al secondo a una temperatura di 1,2 kelvin (gradi centigradi al di sopra dello zero assoluto), mentre per le molecole di azoto nell'aria a temperatura ambiente essa è di circa 0,2 centimetri quadrati al secondo. La diffusione estremamente rapida dell'eccitone è dovuta alla sua piccola massa e alla diffusione relativamente rara da parte di altre particelle a basse temperature.

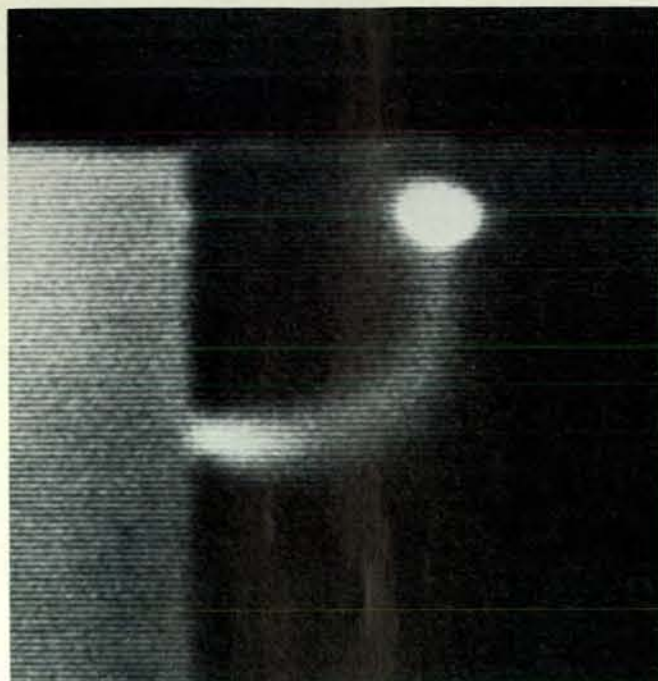
Un'altra fonte di informazioni sul moto dell'eccitone è la misura della sua mobilità. Quando si applica una forza a un ecci-

tone, la seconda legge del moto di Newton impone che l'eccitone venga accelerato attraverso il cristallo in modo inversamente proporzionale alla sua massa inerziale efficace. Tuttavia, le collisioni tra l'eccitone e gli atomi in vibrazione termica del reticolo cambiano costantemente la direzione di moto dell'eccitone. L'eccitone deve quindi ricominciare ad accelerare tutte le volte che viene diffuso. Per un dato campo di forza, si può dimostrare che esso deve raggiungere una certa velocità media di migrazione lungo il campo il cui modulo è proporzionale all'intensità del campo. La mobilità di una particella a diffusione continua è definita pertanto come il rapporto tra la velocità di migrazione della particella e la forza applicata. Per una particella carica quale l'elettrone la mobilità si misura comunemente in un campo elettrico e la forza per unità di carica agente su tale particella si esprime in volt per centimetro. La mobilità si esprime quindi solitamente in centimetri quadrati per volt per secondo.

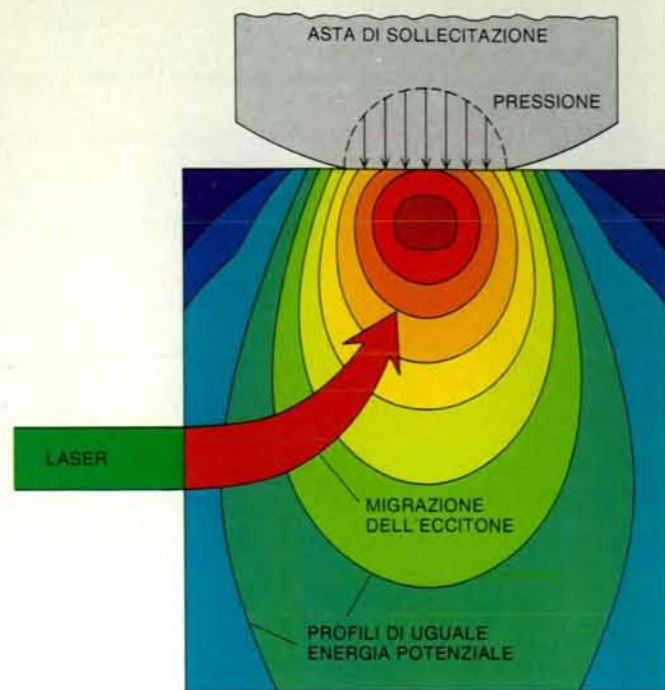
Per poter misurare la mobilità dell'eccitone si deve determinare la sua velocità sotto l'azione di una forza applicata. La forza non può ovviamente essere ricavata da un campo elettrico, dal momento che l'eccitone è elettricamente neutro. Lo stratagemma consiste nell'applicare una sollecitazione esterna al cristallo semiconduttore. La sollecitazione fa diminuire l'energia dell'intervallo proibito del cristallo e riduce quindi l'energia dell'eccitone. Se al cristallo si applica una sollecitazione gradualmente crescente, l'eccitone può diminuire la propria energia spostandosi da una regione di bassa sollecitazione a una di sollecitazione maggiore. Il gradiente di energia, ovvero la variazione di energia dell'eccitone per unità di distanza, è la forza agente sull'eccitone.

Gradienti energetici indotti da sollecitazione sono stati applicati per misurare la mobilità dell'eccitone nel silicio, nel germanio e nell'ossido rameoso. Il cristallo viene compresso al centro di una delle sue superfici levigate dall'estremità arrotondata di un'asta. La sollecitazione nel materiale è massima all'interno del cristallo pochi decimi di millimetro sotto il punto di contatto. Gli eccitoni prodotti dalla luce che colpisce una superficie adiacente del cristallo migrano verso questa regione a elevata sollecitazione e alcuni di essi si ricombinano lungo il percorso. Si può determinare la forza sull'eccitone rappresentando graficamente lo spostamento di frequenza della luce emessa dagli eccitoni in funzione della loro posizione nel cristallo. Lo spostamento spettrale per spostamento unitario equivale a una variazione dell'energia dell'intervallo di banda per quello stesso spostamento, e tale rapporto è la forza agente sull'eccitone.

Un notevole esempio di migrazione degli eccitoni è riportato in una fotografia di un cristallo di silicio (si veda l'illustrazione a pagina 76). Il cristallo è stato raffreddato a circa 10 kelvin, quindi eccitato da un fascio laser continuo. Una telecamera vidicon, sensibile alla radiazione



L'energia dell'eccitone può venire modificata da una regione all'altra di un cristallo esercitando una pressione sulla superficie del cristallo stesso con l'estremità arrotondata di un'asta di sollecitazione. Il gradiente energetico che ne risulta, ovvero la variazione di energia dell'eccitone per percorso unitario nel cristallo, è una forza che tende a spingere l'eccitone verso le regioni di energia minore. Nella fotografia di sinistra, eseguita da Paul L. Gourley e da uno degli autori del presente articolo (Wolfe) all'Università dell'Illinois, l'asta preme sul centro della superficie superiore di un cristallo di silicio. La buca di potenziale, regione di minima energia potenziale per l'eccitone, si trova proprio sotto il punto di sollecitazione, come avviene nell'ossido rameoso; la luminescenza che segnala il flusso e l'accumulo degli eccitoni è, invece, nella



regione infrarossa dello spettro ed è stata perciò registrata da una telecamera sensibile alla radiazione infrarossa e poi convertita in un'immagine visibile sullo schermo di un tubo a raggi catodici. La frequenza di ricombinazione degli eccitoni diminuisce quando questi si spostano verso regioni di minore energia. Il disegno sulla destra mostra profili di energia potenziale costante calcolati da Robert S. Markiewicz dell'Università della California a Berkeley. La successione dei colori, dal violetto, al blu, al verde, all'arancione e a tonalità sempre più scure del rosso (che rappresentano la diminuzione di frequenza della radiazione infrarossa emessa quando in queste regioni si ricombinano coppie di elettroni e di buche), dà le regioni di energia potenziale decrescente degli eccitoni per il cristallo della fotografia.

infrarossa, registrava la luminescenza dell'eccitone a una lunghezza d'onda di 1,2 micrometri, corrispondente a un'energia del fotone pari a circa la metà di quella della luce gialla. La traiettoria di migrazione dell'eccitone è chiaramente visibile perché nel silicio il gradiente di sollecitazione, cioè la variazione della sollecitazione per percorso unitario, può essere molto alto. Per poter eseguire la fotografia, al cristallo di silicio è stata applicata una forza corrispondente a circa 100 chilogrammi su una superficie di un millimetro quadrato; una tale pressione schiaccerebbe molti altri cristalli semiconduttori nei quali si possono produrre eccitoni, quale l'ossido rameoso.

La velocità di migrazione degli eccitoni viene determinata impulsando la luce di eccitazione del laser e osservando poi il picco della distribuzione dell'eccitone in vari istanti mentre esso migra nel cristallo. Le velocità di migrazione degli eccitoni in cristalli di elevata purezza, mantenuti a 1,5 kelvin, variano da 100 a 1000 metri al secondo e quindi le mobilità calcolate risultano dell'ordine di $10^6 - 10^7$ centimetri quadrati per volt per secondo. In confronto, la velocità di migrazione degli elettroni che danno luogo a una corrente elettrica di un ampere in un filo di rame di un millimetro di diametro è inferiore a 0,1 millimetri per secondo. In altri termini, se gli elettro-

ni nel filo potessero migrare con la stessa velocità degli eccitoni in un cristallo, la corrente nel filo sarebbe di quasi 10 milioni di ampere. A temperatura ambiente la mobilità degli elettroni nel filo è di circa 40 centimetri quadrati per volt per secondo, inferiore di circa cinque ordini di grandezza alla mobilità degli eccitoni.

In tutti i semiconduttori studiati finora la mobilità degli eccitoni in un cristallo aumenta drasticamente al diminuire della temperatura del cristallo. Tale relazione tra mobilità e temperatura è un'ulteriore conferma del fatto che gli eccitoni si diffondono dal reticolo cristallino: al diminuire della temperatura del cristallo l'ampiezza delle vibrazioni reticolari decresce e quindi dovrebbe aumentare l'intervallo di tempo tra gli eventi di diffusione. Misurazioni della migrazione possono perciò fornire informazioni microscopiche sul moto degli eccitoni. Per esempio, gli eccitoni dell'ossido rameoso a 1,5 kelvin devono percorrere circa 30 micrometri tra un evento di diffusione e l'altro, una distanza enorme in scala atomica. L'eccitone è chiaramente una particella molto mobile.

Come reagiscono gli eccitoni alla presenza di altri eccitoni? Si può far aumentare la probabilità di interazioni tra gli eccitoni aumentando la potenza

del laser incidente, aumentando così la densità del gas di eccitoni. In questo contesto si rivela ancora utile riprendere l'analogia tra eccitone e atomo di idrogeno. Anche se gli atomi di idrogeno sono elettricamente neutri, in un gas essi hanno una forte tendenza a combinarsi per formare una molecola di idrogeno (H_2). Ciascun atomo si comporta come un dipolo elettrico e due dipoli elettrici si attraggono l'un l'altro esattamente come due dipoli magnetici. A causa di questa attrazione, l'energia della molecola di idrogeno è inferiore a quella di due atomi separati.

Allo stesso modo l'energia di due eccitoni può diminuire se essi formano una molecola eccitonica, o bieccitone. Dal momento che il bieccitone è formato da quattro particelle di massa simile, si prevedeva inizialmente che la molecola fosse dotata solo di una stabilità marginale. Invece, i bieccitoni sono stati rivelati in cristalli quali il silicio, il germanio, il cloruro di rame, il bromuro di argento e il solfuro di cadmio.

Un metodo comune per creare bieccitoni è quello di generare eccitoni con tale abbondanza da farli legare uno all'altro. La presenza di bieccitoni viene spesso segnalata da una nuova riga di luminescenza. L'energia dei fotoni che danno origine alla nuova riga è inferiore a quella

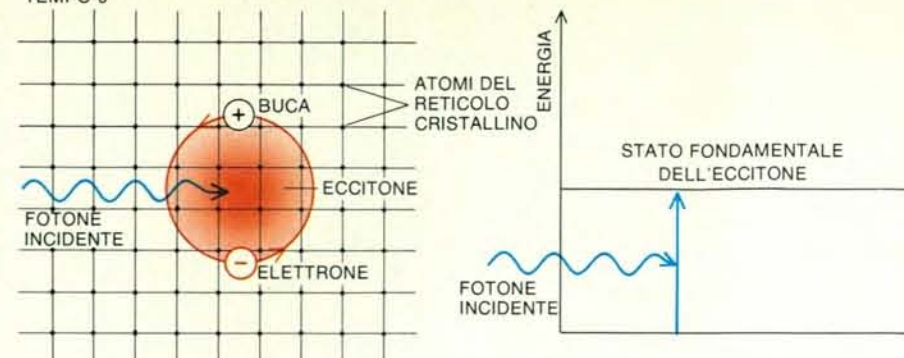
associata alla ricombinazione di un singolo eccitone. Quindi la riga potrebbe nascere quando una coppia elettrone-buca in un bieccitone si ricombina lasciando dietro di sé un eccitone libero. Inoltre, nel silicio e nel germanio l'intensità della nuova riga di luminescenza aumenta con il quadrato dell'intensità di luminescenza dell'eccitone. Tale aumento è esattamente ciò che ci si aspettava per i bieccitoni, perché la probabilità di creare un bieccitone aumenta con il quadrato della densità degli eccitoni.

I bieccitoni vengono spesso prodotti aumentando semplicemente la densità degli eccitoni, ma possono essere prodotti anche per eccitazione ottica diretta. Un metodo richiede il simultaneo assorbimento di due fotoni, ciascuno avente esattamente la metà dell'energia del bieccitone. Poiché il bieccitone è formato da due coppie elettrone-buca legate, l'energia totale dei due fotoni deve essere uguale all'energia di due eccitoni liberi, inferiore all'energia di legame del bieccitone (si veda l'illustrazione in questa pagina). Quando si raggiunge questa condizione di risonanza regolando la frequenza del laser, si osserva un'intensa riga di assorbimento. La riga è del tutto particolare: compare soltanto quando l'intensità della luce incidente è elevata. In un cristallo di cloruro di rame dello spessore di un micrometro è necessario un fascio laser con la potenza di un megawatt per provocare l'assorbimento del 10 per cento dei fotoni. In confronto ad altri processi nei solidi, che comportano l'assorbimento di due fotoni, questo effetto ottico è particolarmente intenso.

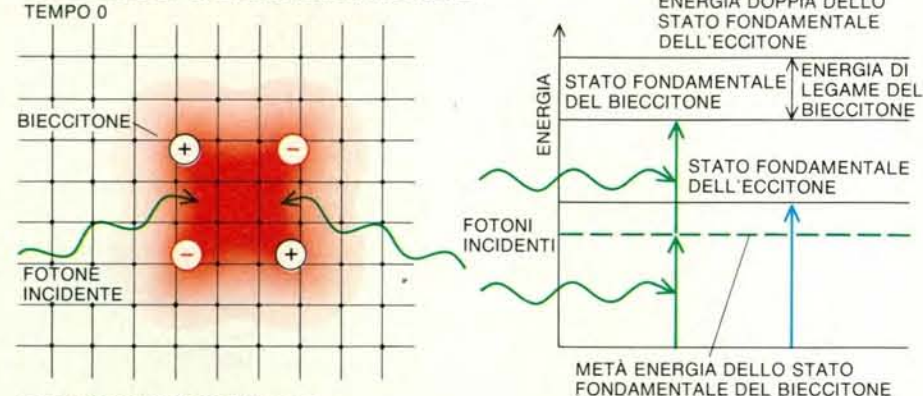
Che cosa si sa sulle altre fasi della materia eccitonica analoghe alle forme liquida e solida dell'idrogeno molecolare, trovate a basse temperature? Finora non sono stati osservati né un liquido molecolare né alcuna fase solida di tale materia. Si è invece osservata in un certo numero di cristalli una nuova fase, non osservata nell'idrogeno. Quando la temperatura di un cristallo del genere viene abbassata a pochi gradi dallo zero assoluto viene emessa un'ampia gamma di frequenze di luminescenza; né la ricombinazione di eccitoni né quella di bieccitoni possono spiegare la gamma di frequenze o la frequenza della luce più intensa all'interno di tale gamma. Inoltre, si è scoperto che la densità delle particelle che danno origine all'emissione è molto più alta che in un gas di eccitoni.

Un modo elegante per dimostrare sperimentalmente l'aumento di densità si basa su un metodo simile a quello applicato per misurare la migrazione degli eccitoni. Ricordiamo che, quando un cristallo viene compresso dalla punta arrotondata di un'asta, si instaura in esso una regione di massima sollecitazione di taglio. Un effetto del genere è ben noto a quanti progettano cuscinetti a sfera: un cuscinetto a sfera sottoposto a una sollecitazione estrema si spezza dall'interno, perché il punto di massimo sforzo si trova entro la sfera. Dato che l'energia dell'intervallo

ECCITAZIONE DI UN ECCITONE IN RISONANZA



ECCITAZIONE DI UN BIECCITONE IN RISONANZA



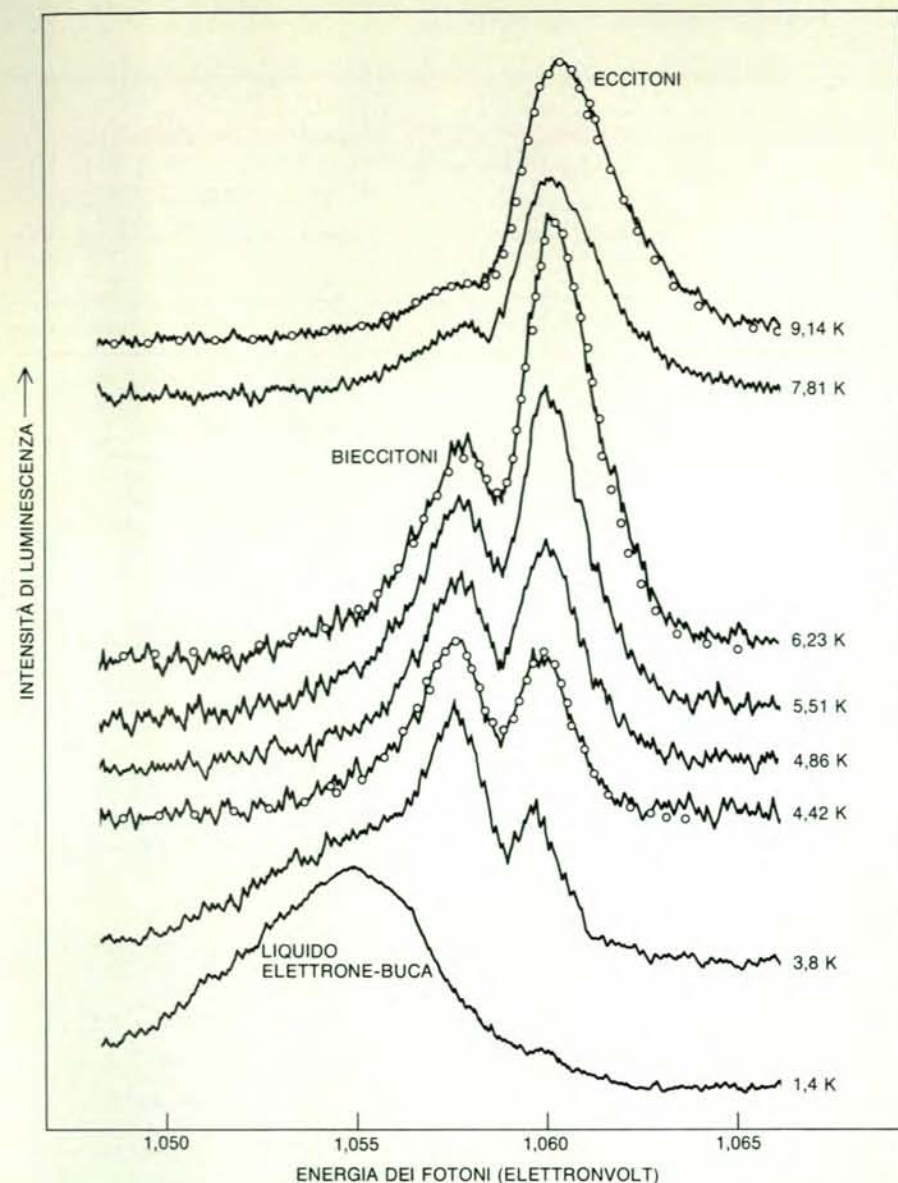
DECADIMENTO DI UN BIECCITONE



RICOMBINAZIONE



Il bieccitone, o molecola eccitonica, viene prodotto quando su una sottile pellicola cristallina convergono due fasci laser che hanno la stessa frequenza. Ciascun fascio laser da solo attraverserebbe il cristallo senza essere praticamente attenuato, mentre l'interazione dei due fasci con il cristallo può portare all'assorbimento a una frequenza uguale alla somma delle singole frequenze dei due fasci. L'assorbimento ha luogo quando si regola la frequenza di ciascuno dei fasci in modo da farla coincidere con la metà dell'energia totale del bieccitone. Il bieccitone è uno stato legato di due elettroni e di due buche che si forma perché la sua energia è inferiore a quella dei due eccitoni liberi. Esso decade quando una coppia elettrone-buca si ricombina e la seconda coppia forma un eccitone ordinario. Questa seconda coppia elettrone-buca si ricombina di lì a poco. Nella rappresentazione sono state usate le stesse convenzioni dello schema di pagina 70.



Un gas di bieccitoni e una fase liquida di elettroni e buche possono essere identificati nello spettro di luminescenza della materia eccitonica confinata in una buca di potenziale all'interno di un cristallo di silicio. È questa la regione di minima energia indotta da sollecitazione nel cristallo della illustrazione di pagina 76. Alla temperatura più alta per la quale è stato tracciato lo spettro, nella buca di potenziale predominano gli eccitoni. Al diminuire della temperatura si osserva un chiaro picco di luminescenza dovuto ai bieccitoni, spostato rispetto all'energia degli eccitoni di una quantità pari all'energia di legame del bieccitone. A temperature ancora inferiori compare un largo picco, che indica la condensazione degli eccitoni e dei bieccitoni in una fase liquida. I cerchietti indicano i valori teorici dell'intensità. I dati sono di Gourley e Wolfe.

proibito diminuisce all'aumentare dello sforzo, nel cristallo si forma una «buca di potenziale», cioè una regione di energia potenziale minima per l'eccitone. La buca di potenziale intrappola la materia eccitonica e la confina. Negli esperimenti effettuati sul silicio è stato trovato che la comparsa del nuovo picco di luminescenza emesso dalla buca di potenziale è accompagnata da una brusca diminuzione del volume della regione della buca interessata. In altri termini, aumenta la densità delle particelle che sono confinate dalla buca di potenziale.

Perché le energie delle particelle nella fase condensata sono distribuite in un intervallo così grande? Per comprendere

questo fenomeno si devono esaminare più da vicino le proprietà statistiche degli eccitoni. A un livello fondamentale tutte le particelle in natura si possono classificare in due gruppi principali a seconda del loro spin. Lo spin di una particella è simile allo spin di una palla, tranne per il fatto che in meccanica quantistica lo spin è definito solo per valori discreti. Una particella con spin intero è detta bosone, mentre una particella con spin semiintero è detta fermione.

Sia l'elettrone che la buca hanno spin $1/2$, quindi sono fermioni. Un eccitone, essendo formato da due particelle con spin $1/2$, ha spin intero ed è quindi un bosone. A densità elevate un insieme di

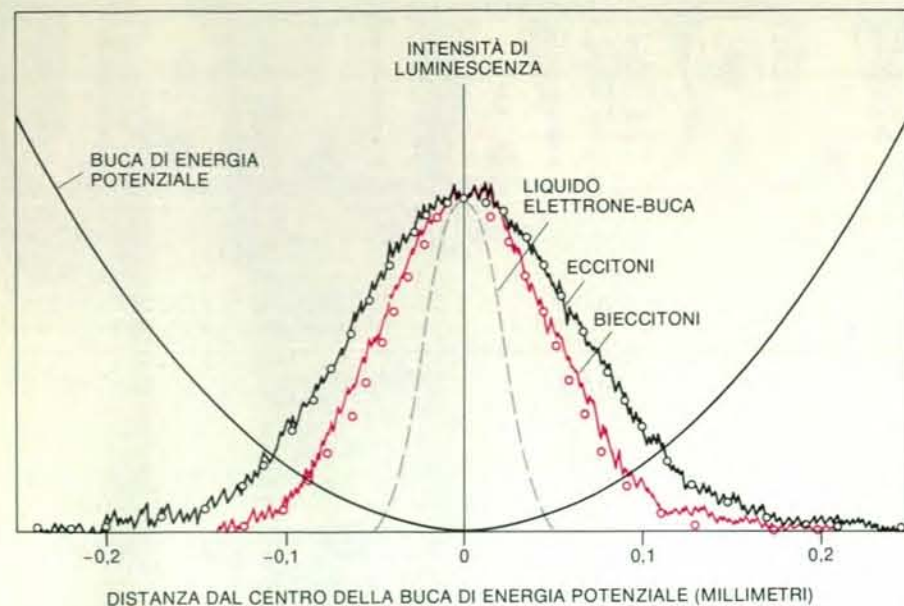
fermioni ha proprietà quantomeccaniche che sono del tutto diverse da quelle di un insieme di bosoni. Secondo il principio di esclusione, formulato da Wolfgang Pauli, solo un fermione alla volta può occupare un dato stato quantomeccanico. Non esiste invece limite alcuno al numero di bosoni che possono occupare un dato stato.

La distinzione tra fermioni e bosoni è particolarmente importante alle basse temperature, alle quali le particelle tendono a occupare gli stati di minima energia disponibili. Se la temperatura di un insieme di fermioni viene abbassata, gli stati di minima energia vengono riempiti senza che si verifichino occupazioni multiple. Quindi anche allo zero assoluto un insieme di fermioni deve presentare un ampio intervallo di energie; lo stato occupato di massima energia dipende dalla densità dei fermioni.

Questa proprietà fermionica degli elettroni e delle buche spiega l'ampio intervallo di luminescenza a bassa temperatura nel silicio. La larghezza dell'intervallo spettrale diminuisce solo lievemente al diminuire della temperatura, il che indica che gli eccitoni si sono dissociati in elettroni e in buche che riempiono tutti gli stati disponibili secondo il principio di esclusione. La misura della larghezza dell'intervallo spettrale fornisce la densità di particelle nella fase condensata; per elettroni e buche confinati sul fondo di una buca di potenziale nel silicio, la densità è di 3×10^{17} particelle al centimetro cubo, pari a un milionesimo della densità atomica del cristallo. Questo fluido di fermioni si chiama liquido elettrone-buca. Nel germanio è stata osservata e studiata con vari metodi una singola grande goccia di un liquido puro elettrone-buca con diametro fino a un millimetro. Dal momento che questo liquido è formato da portatori di carica liberi, ha una conducibilità elettrica paragonabile a quella di un filo di rame.

Nei cristalli non sottoposti a sollecitazioni esterne il liquido elettrone-buca si raccoglie in goccioline molto più piccole, le cui dimensioni sono state stimate nel germanio misurando la diffusione, da parte loro, della radiazione laser infrarossa. Esse hanno un diametro di circa cinque micrometri, che corrisponde a 100 milioni di coppie elettrone-buca per gocciolina. Le goccioline formano comunemente una nube che circonda il punto di eccitazione sulla superficie del cristallo e all'interno del gas di eccitoni.

La nube di goccioline assomiglia a una nebbia di particelle liquide, ma non rimane in quiete. Le goccioline si condensano come le goccioline d'acqua nell'atmosfera; hanno una tensione superficiale debole, ma finita, pari a circa un milionesimo della tensione superficiale dell'acqua. Inoltre, possono essere sospinte attraverso il cristallo da un «vento» di fononi, i quanti di energia vibrazionale del reticolo atomico simili a particelle. Per esempio, un vento di fononi non uniforme si propaga attraverso il cristallo dal punto di eccitazione laser, e il vento spinge la nube di



Il volume della regione della buca di potenziale che emette luce, associato a ogni picco dello spettro di luminescenza della figura di pagina 78, conferma la presenza in essa del gas di eccitoni e del liquido elettrone-buca. Le curve qui sopra mostrano che la luminescenza degli eccitoni è emessa dal volume maggiore. Dal momento che la massa del bieccitone è doppia di quella dell'eccitone, si prevede che la luminescenza proveniente dal decadimento dei bieccitoni (cerchietti in colore) debba essere emessa da una regione il cui diametro è $\sqrt{2/2}$, ovvero circa 0,7 volte quello della regione associata al decadimento degli eccitoni (cerchietti in nero). Le curve mostrano che i dati sono in accordo con la previsione. Il volume della regione che emette la luminescenza tipica del liquido elettrone-buca è ancora più piccolo (curva tratteggiata), il che sta a indicare la condensazione dalla fase di gas a quella liquida, più densa. I dati sono di Gourley e Wolfe.

goccioline elettrone-buca più lontano in certe direzioni che non in altre. Lo studio delle interazioni dei fononi con le goccioline ha portato a una migliore conoscenza della propagazione dei fononi alle basse temperature.

Non tutti gli eccitoni nei semiconduttori si possono condensare in liquidi elettrone-buca. Per esempio, è previsto che gli eccitoni nell'ossido rameoso restino in fase gassosa anche a densità elevate e non sono stati osservati, in esso, né bieccitoni né liquido elettrone-buca. Nel cloruro di rame si possono formare bieccitoni, ma non è stato osservato alcun liquido. Se in questi cristalli gli eccitoni rimangono nella fase gassosa, la larghezza dell'intervallo spettrale che segnala la ricombinazione degli eccitoni dovrebbe diminuire con l'abbassarsi della temperatura, perché gli eccitoni sono bosoni e i bosoni tendono a riempire gli stati di minima energia. In effetti, se la concentrazione di un gas di bosoni supera un valore critico che dipende dalla temperatura e dalla massa del bosone, si prevede che tutte le particelle aggiunte al gas si condensino in uno stato di minima energia quasi privo di moto. L'effetto, che si basa su una teoria formulata nel 1924 dal fisico indiano S. N. Bose per fotoni privi di massa e successivamente generalizzata da Einstein per particelle dotate di massa, è detto condensazione di Bose-Einstein.

Non è ancora possibile affermare con certezza che questa condensazione sia stata osservata. Molti pensano che essa

possa spiegare l'improvvisa diminuzione della viscosità e le altre proprietà di un superfluido, lo stato raggiunto dall'elio liquido a una temperatura prossima allo zero assoluto. Tuttavia, a causa delle forti interazioni tra le particelle di un liquido, è difficile descrivere teoricamente come avvenga nell'elio 4, l'isotopo dell'elio che diventa superfluido a 2,17 kelvin. Molti ricercatori hanno tentato, invece, di osservare l'effetto in un gas rarefatto, per il quale il modello teorico è molto più immediato.

Un condensato di Bose-Einstein, se esiste, rappresenta uno stato della materia del tutto nuovo e quindi le sue proprietà possono essere insolite. Per esempio, è del tutto diverso dagli stati della materia che sono condensati spazialmente, la fase liquida e quella solida; in realtà, il termine condensazione si riferisce solo all'aggregazione dei bosoni negli stati a bassa energia. Il comportamento delle particelle è coerente su distanze macroscopiche, come lo sono i moti coerenti delle coppie di elettroni che trasportano la corrente elettrica senza resistenza in un metallo superconduttore.

Per una data densità in un gas, quanto minore è la massa della particella, tanto maggiore è la temperatura critica per la condensazione di Bose-Einstein. La temperatura critica dell'idrogeno atomico, l'elemento più leggero della tavola periodica, è leggermente inferiore a 0,1 kelvin per una densità del gas di 10^{19} atomi per centimetro cubo. A tale densità l'idrogeno solitamente non può subire

la condensazione di Bose-Einstein, perché i suoi atomi si combinano subito per formare molecole biatomiche. Ciononostante, si stanno effettuando promettenti tentativi per generare un gas di idrogeno atomico ad alta densità allineando magneticamente gli spin degli atomi. Quasi tutti gli altri elementi solidificano prima di raggiungere le condizioni estreme di temperatura e di densità necessarie per la condensazione di Bose-Einstein. Tuttavia, a causa della piccola massa dell'eccitone, è previsto che la condensazione di Bose-Einstein di un gas di eccitoni la cui densità è di 10^{19} particelle per centimetro cubo debba avvenire a circa 100 kelvin, una temperatura molto maggiore di quella prevista per un qualsiasi sistema atomico.

Vi sono oggi numerosi dati sperimentali a favore della condensazione di Bose-Einstein di un gas di eccitoni o di bieccitoni. Se si aumenta gradualmente la densità degli eccitoni nell'ossido rameoso a due kelvin, lo spettro di luminescenza comincia a discostarsi da quello previsto dalla distribuzione classica delle velocità delle particelle. Invece, la distribuzione di velocità ottenuta supponendo che gli eccitoni siano alla soglia per la condensazione di Bose-Einstein è in ottimo accordo con i dati. Ricercatori sovietici riferiscono anche che l'intervallo delle energie dei fotoni emessi che segnalano la ricombinazione di eccitoni nel germanio si restringe all'aumentare della densità degli eccitoni. La scoperta è in accordo con le previsioni del comportamento di un gas di bosoni nei pressi della soglia per la condensazione di Bose-Einstein.

Infine, alle alte densità un gas di bieccitoni in cloruro di rame presenta anomalie nel suo spettro di luminescenza. I bieccitoni vengono creati direttamente in uno stato a bassa energia da eccitazione luminosa in risonanza. A temperature inferiori a 30 kelvin e a densità superiori a 10^{18} particelle per centimetro cubo, nello spettro appare una netta riga di ricombinazione, che indica che la maggior parte dei bieccitoni sta decadendo da uno stato a bassa energia. Se si sonda questo sistema aggiungendovi bieccitoni attraverso eccitazione non risonante, i bieccitoni vengono portati nello stato occupato a bassa energia. Invece, se nel sistema non vi sono già bieccitoni risonanti, i bieccitoni sonda acquistano una distribuzione di velocità non quantomeccanica.

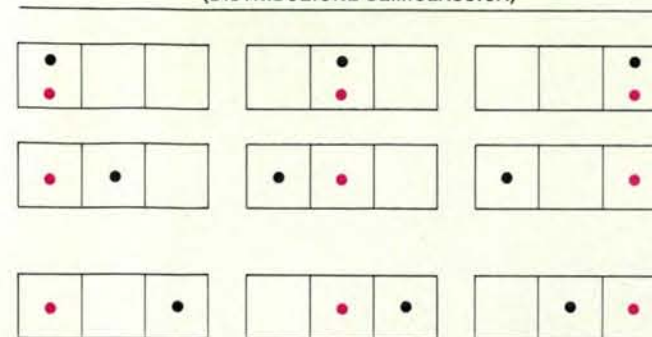
Anche se questo risultato fa pensare seriamente che sia presente nel cristallo un condensato di Bose-Einstein, si può anche concludere che gli effetti si osservano solo perché l'eccitazione in risonanza iniziale vincola tutti i bieccitoni nello stesso stato di bassa energia. Ciò che ci piacerebbe vedere, e che ancora non siamo riusciti a vedere, è la condensazione di Bose-Einstein di un gas che si trova inizialmente in equilibrio termico. Non sorprenderebbe, però, se un gas di eccitoni diventasse il primo sistema fisico capace di mostrare inequivocabilmente la condensazione.

Lo studio delle interazioni degli eccitoni permette di esplorare regioni di temperatura, pressione, intensità di luce e scala dei tempi che sono normalmente ai limiti delle possibilità sperimentali. Inoltre, gli eccitoni si possono osservare in una vasta

varietà di materiali: oltre ai semiconduttori e ai sistemi biologici, abbiamo già detto che essi esistono negli isolanti, nei cristalli organici e nelle strutture periodiche artificiali chiamate super-reticoli. Nuovi strumenti in fase di sviluppo quali i laser capaci di gene-

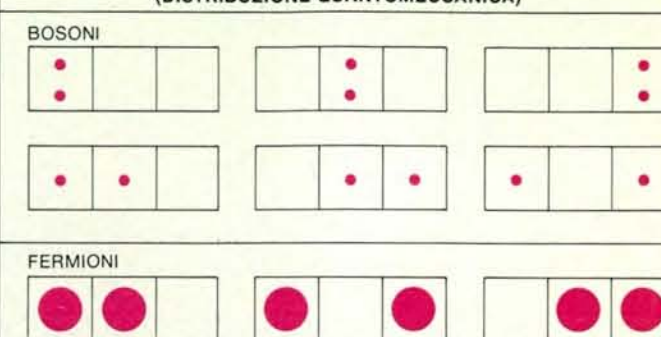
rare impulsi brevi fino alla frazione di picosecondo consentiranno ai ricercatori di esplorare la dinamica degli eccitoni con risoluzione temporale ancora migliore. È pressoché certo che in tali campi verranno scoperti nuovi fenomeni.

PARTICELLE DISTINGUIBILI (DISTRIBUZIONE SEMICLASSICA)

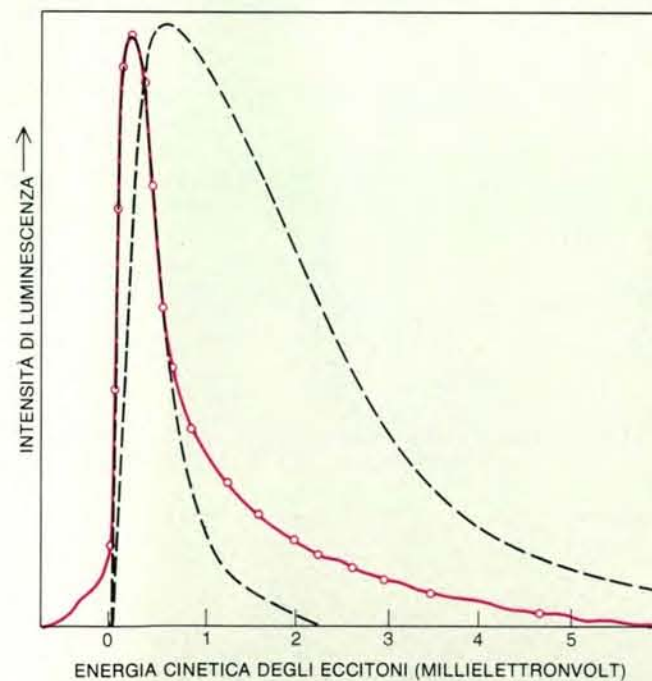


La indistinguibilità delle particelle elementari spiega la differenza tra le distribuzioni quantomeccanica e semiclassica, o non quantomeccanica, delle particelle tra i vari stati energetici disponibili. Nella distribuzione semiclassica ogni particella deve occupare uno dei vari stati energetici quantizzati, ma, poiché le particelle sono distinguibili, esse possono essere distribuite tra gli stati energetici in un maggior numero di modi che non nelle distribuzioni quantistiche. Per esempio, se due particelle distinguibili possono adattarsi entrambe in uno qualsiasi di tre stati, vi sono nove modi nei quali esse possono disporsi. Se in uno stato può adattarsi solo una particella, vi sono sei modi per sistemare le particelle (a sinistra). Nelle distribuzioni quantistiche tutte le particelle di una

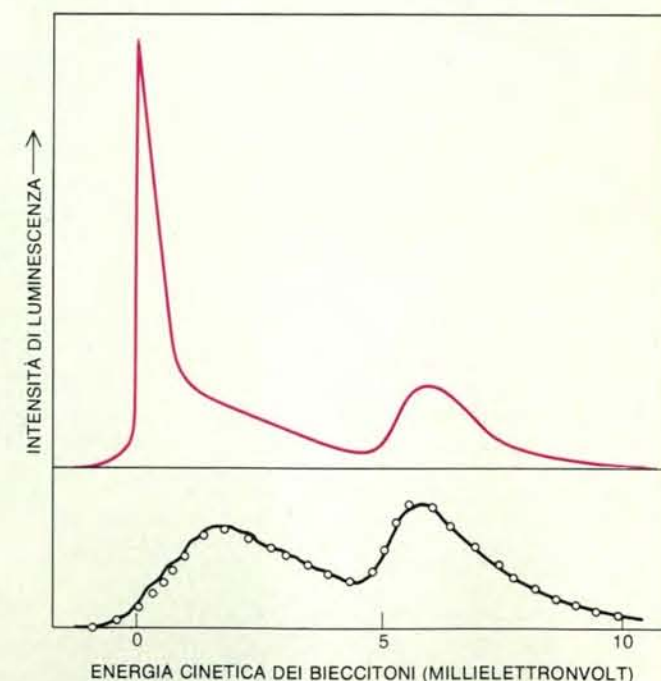
PARTICELLE INDISTINGUIBILI (DISTRIBUZIONE QUANTOMECCANICA)



data specie sono indistinguibili. Per i bosoni, che possono occupare lo stesso stato energetico, tre delle sei possibili disposizioni delle particelle sono formate da stati a occupazione multipla, una proporzione maggiore di tre su nove (in alto a destra). Alle basse temperature, per le quali il numero di particelle è confrontabile al numero degli stati energetici disponibili, questa proprietà della distribuzione quantistica dei bosoni confina efficacemente negli stati di minima energia una frazione di particelle maggiore di quella prevista dalla distribuzione semiclassica. Per i fermioni, due dei quali non possono occupare lo stesso stato energetico, esistono solo tre modi per sistemarli nei tre stati energetici, in luogo dei sei della distribuzione semiclassica (in basso a destra).



La condensazione di Bose-Einstein può spiegare la forma anomala dei due spettri di luminescenza qui riportati. Essa è prevista dalla distribuzione quantomeccanica degli eccitoni e dei bieccitoni, tutti classificati come bosoni, tra gli stati energetici disponibili per essi alle basse temperature. A sinistra è lo spettro di luminescenza per un gas denso di eccitoni in ossido rameoso. La distribuzione delle velocità delle particelle deducibile dallo spettro (in colore) si avvicina molto a quella prevista dalla statistica di Bose-Einstein (cerchietti). Le distribuzioni semiclassiche mostrate per due possibili temperature (in nero) possono adattarsi a parti dello spettro osservato, mentre l'adattamento complessivo non è così preciso come quello che si avrebbe assumendo come valida la statistica di Bose-Einstein. I dati sono stati raccolti da Hulin, Claude



Benoit e Mysyrowicz. Sulla destra è riportato lo spettro di luminescenza di un gas di bieccitoni in cloruro di rame (CuCl). La curva in nero rappresenta lo spettro ottenuto da un gas di bieccitoni relativamente rarefatto, prodotto da un laser di bassa potenza, ed è in buon accordo con la distribuzione di velocità dei bieccitoni calcolata da una distribuzione semiclassica (cerchietti). La curva in colore rappresenta lo spettro ottenuto con un laser di maggiore potenza e che mostra un grande aumento nella luminescenza dei bieccitoni a bassa energia. Il picco netto, il che indica che quasi tutti i bieccitoni hanno la stessa energia cinetica. Uno spettro del genere è previsto per un condensato di Bose-Einstein. I dati sono stati raccolti alla Indiana University da Lloyd L. Chase, Nasser Peyghambarian, Gilbert Grynberg e Mysyrowicz.

L'origine degli ominoidi e degli ominidi

Le scoperte degli ultimi cinque anni hanno consentito di determinare con maggiore sicurezza quando le scimmie antropomorfe si sono separate dalle scimmie del Vecchio Mondo e quando l'uomo si è evoluto da esse

di David Pilbeam

La nostra conoscenza dell'evoluzione umana ha fatto grandi progressi sia nel corso della passata generazione sia durante gli ultimi cinque anni. Per esempio, a partire dagli anni sessanta, la paleoantropologia ha cessato di essere una disciplina riservata esclusivamente agli antropologi fisici e a pochi paleontologi interessati; essa si è trasformata in un ampio impegno di ricerca interdisciplinare a cui collaborano esperti di geologia storica, di ecologia, di biologia molecolare, di zoologia, di etologia e anche di chimica e di fisica.

Nella classificazione linneana del mondo vivente, gli esseri umani appartengono all'ordine dei primati che comprende anche le scimmie antropomorfe, le scimmie e le prosimie (ad esempio, i lemuri); all'interno di quest'ordine vi è la superfamiglia degli ominoidi, costituita unicamente dagli esseri umani e dalle scimmie antropomorfe (antropomorfi o pongidi); di essa fa parte la famiglia degli ominidi, che comprende esclusivamente l'uomo e alcune specie estinte che gli sono affini, molto più affini che alle scimmie antropomorfe.

Mi si consenta di riassumere brevemente in che modo era interpretata cinque anni fa l'evoluzione degli ominoidi e degli ominidi da un eminente studioso di antropologia fisica, Sherwood L. Washburn, dell'Università della California a Berkeley, e di mettere a confronto le conclusioni di allora con quelle a cui si è giunti attualmente. In quel periodo i progressi rivoluzionari della geologia e della biologia molecolare avevano già posto in discussione le risultanze precedenti in materia di geografia e di cronologia dell'evoluzione degli ominidi, come pure le affinità biologiche fra l'uomo e gli altri primati superiori. Per esempio, i dati della biologia molecolare hanno semplificato la ricostruzione dell'origine dei primati, anche se gli antropologi fisici sono stati riluttanti ad accogliere ipotesi secondo le quali le scimmie antropomorfe africane

(scimpanzé e gorilla), a differenza di quelle asiatiche (orangi e gibboni), sarebbero strettamente imparentate con l'uomo. L'ipotesi avanzata dai biologi molecolari, di una possibile differenziazione dell'uomo dalla linea africana delle scimmie antropomorfe risalente a molto meno di 10 milioni di anni o sono, era ritenuta particolarmente ardua.

La sintesi di Washburn ipotizzava anche che *Ramapithecus*, un ominoide del Miocene diffuso in vaste aree del Vecchio Mondo, sarebbe potuto essere il precursore di tutte le scimmie antropomorfe attuali e del genere umano. Inoltre, le scoperte in Etiopia di *Australopithecus afarensis* (noto al grande pubblico per lo scheletro parziale soprannominato «Lucy») e in Tanzania delle famose impronte di Laetoli hanno dimostrato che già più di tre milioni e mezzo di anni or sono si erano evoluti in Africa degli ominidi primitivi, primati con un cervello di piccole dimensioni, che vivevano sul terreno, che deambulavano su due gambe e avevano denti canini relativamente piccoli e non prominenti. Queste scoperte hanno confermato anche che la locomozione bipede è stata il primo e fondamentale adattamento evolutivo degli ominidi. Analogamente agli scimpanzé, i primi ominidi impiegavano come utensili pezzi di legno e altri materiali presenti in natura. Il contrasto fra i piccoli canini degli ominidi e quelli prominenti degli scimpanzé ha fatto ritenere, in realtà, che la funzione offensiva di questi denti fosse già stata soppiantata negli ominidi dall'impiego degli utensili.

Riferendosi sempre al lavoro di Washburn, dopo un altro milione di anni si erano evoluti in Africa ominidi con un cervello più sviluppato, che fabbricavano utensili di pietra e che convivevano con una specie robusta di *Australopithecus* (che si differenziava quindi per le ossa pesanti e i grossi denti), estintasi poi nel corso del successivo milione di anni. Anche prima di allora, diciamo circa un mi-

lione e mezzo di anni fa, aveva fatto la sua comparsa un ominide più progredito: *Homo erectus*. La presenza dei suoi utensili litici frammisti a ossa di animali è stata interpretata come una chiara prova che il comportamento venatorio era divenuto una caratteristica degli ominidi molto prima che il loro cervello si sviluppasse nelle dimensioni attuali.

Da allora le dimensioni cerebrali e la complessità dei manufatti si sono sviluppate molto lentamente fino a circa 100 000 anni or sono quando ha avuto origine un ominide ancora più progredito: *Homo sapiens*. Questi sviluppi costituiscono comunque una prova indiretta della riorganizzazione del cervello degli ominidi. Con l'avvento di *Homo sapiens* si intensifica il ritmo di modificazione tipologica degli utensili litici e le variazioni sul piano anatomico si riscontrano principalmente nel cranio. In seguito, circa 40 000 anni fa, sono comparsi uomini con caratteristiche anatomiche moderne, che si sono diffusi rapidamente per tutto il Vecchio Mondo. Contemporaneamente, il tasso di mutamento comportamentale, come è attestato dai cambiamenti apportati agli utensili litici e ad altri manufatti umani, si è avvicinato a quello che ci si attende oggi dal «normale» comportamento culturale dell'uomo. Washburn ha concluso la sua discussione ipotizzando che il linguaggio umano compiutamente moderno, che probabilmente non esisteva prima della comparsa dell'uomo anatomicamente progredito, abbia catalizzato questo progresso comportamentale, anche se una simile ipotesi non è verificabile con prove anatomiche dirette.

Nel corso dei cinque anni posteriori alla sintesi di Washburn, vi sono state nuove importanti scoperte di fossili, non tutte sul campo, e sono state pubblicate descrizioni più complete dei fossili precedenti. I paleoantropologi possono ora interpretare meglio i loro dati grazie agli apporti degli studiosi del comportamento e sono sempre più consapevoli degli sviluppi che

si hanno nelle altre scienze naturali e che sono pertinenti al loro lavoro. Tendenze del genere hanno indotto questa disciplina a una maggiore integrazione con il resto della comunità biologica.

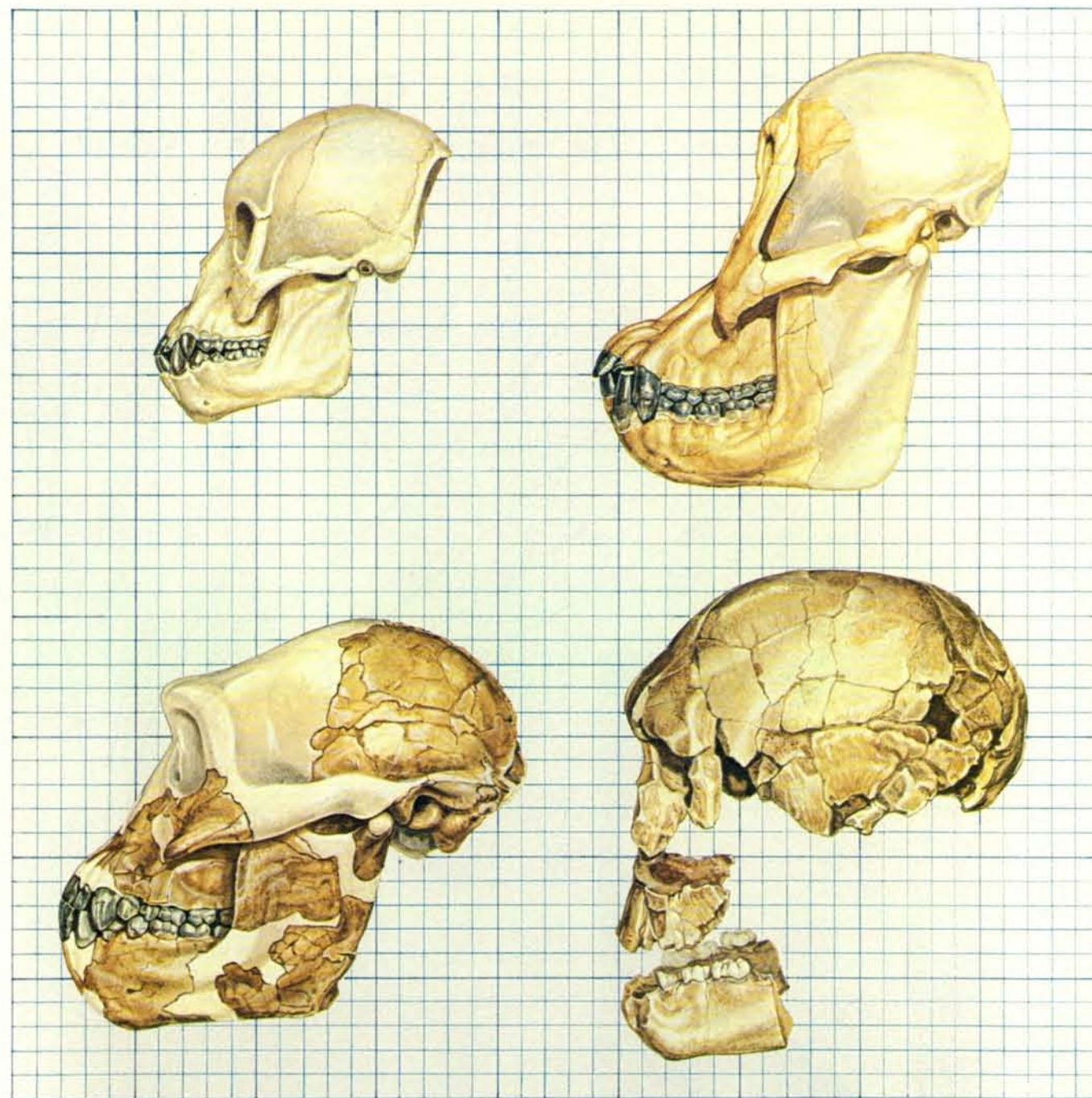
Nel contempo aumentano i dubbi: fino a che punto è possibile «far rivivere» questi racconti dell'evoluzione umana? In che misura queste ricostruzioni possono

essere accurate e attendibili? Ardui problemi sorgono quando gli ominidi del passato vengono interpretati in termini di ominidi del presente. Dopotutto, i primi erano assai diversi dai secondi. Invece, in molti casi, queste differenze sono state ignorate, con il risultato di presentare un'immagine dei primi ominidi troppo simile a quella degli esseri umani attuali.

Vorrei ora passare in rassegna alcuni di questi problemi iniziando con l'evoluzione degli ominoidi, il gruppo da cui hanno avuto origine gli ominidi.

L'evoluzione degli ominoidi

L'annoso dibattito fra antropologi fisici e biologi molecolari sullo schema e sui tempi



I quattro crani fossili, qui di profilo, rappresentano momenti chiave nella divergenza degli ominoidi dal ceppo ancestrale costituito dalle scimmie del Vecchio Mondo e nella successiva divergenza degli ominidi dal ceppo ancestrale degli ominoidi. Quello in alto a sinistra appartiene alla più nota delle scimmie antropomorfe del primo Miocene, *Proconsul africanus*, un primate delle dimensioni di un babuino, probabilmente simile all'antenato comune a tutte le scimmie antropomorfe e all'uomo. In alto a destra è il cranio della scimmia antropomorfa del tardo Miocene *Sivapithecus*, nettamente più grossa e individuata per la prima volta attraverso i denti fossilizzati e i frammenti di mandibola,

trovati nel primo decennio di questo secolo ai piedi dell'Himalaya. Questo primate potrebbe essere l'antenato dell'unico grosso ominoide attuale asiatico, l'orango. In basso a sinistra si vede il cranio del primo ominide del Pliocene scoperto in Africa, *Australopithecus afarensis*, una specie a deambulazione eretta vissuta nel periodo compreso fra quattro e due milioni di anni fa. In basso a destra, infine, è rappresentato il cranio di un ominide con ben più elevata capacità cranica, *Homo habilis*, trovato per la prima volta in Kenya nel 1972. A iniziare da circa due milioni di anni fa, e per qualche centinaio di migliaia di anni, individui di questa specie si diffusero nell'Asia meridionale e orientale.

dell'evoluzione degli ominoidei è ora praticamente composto. Gran parte dei paleontologi (e degli specialisti di anatomia comparata) concorda sul fatto che il quadro molecolare, in base al quale le scimmie antropomorfe africane presentano pochissime differenze genetiche con l'uomo, mentre quelle asiatiche ne conservano in proporzione quasi doppia, ri-

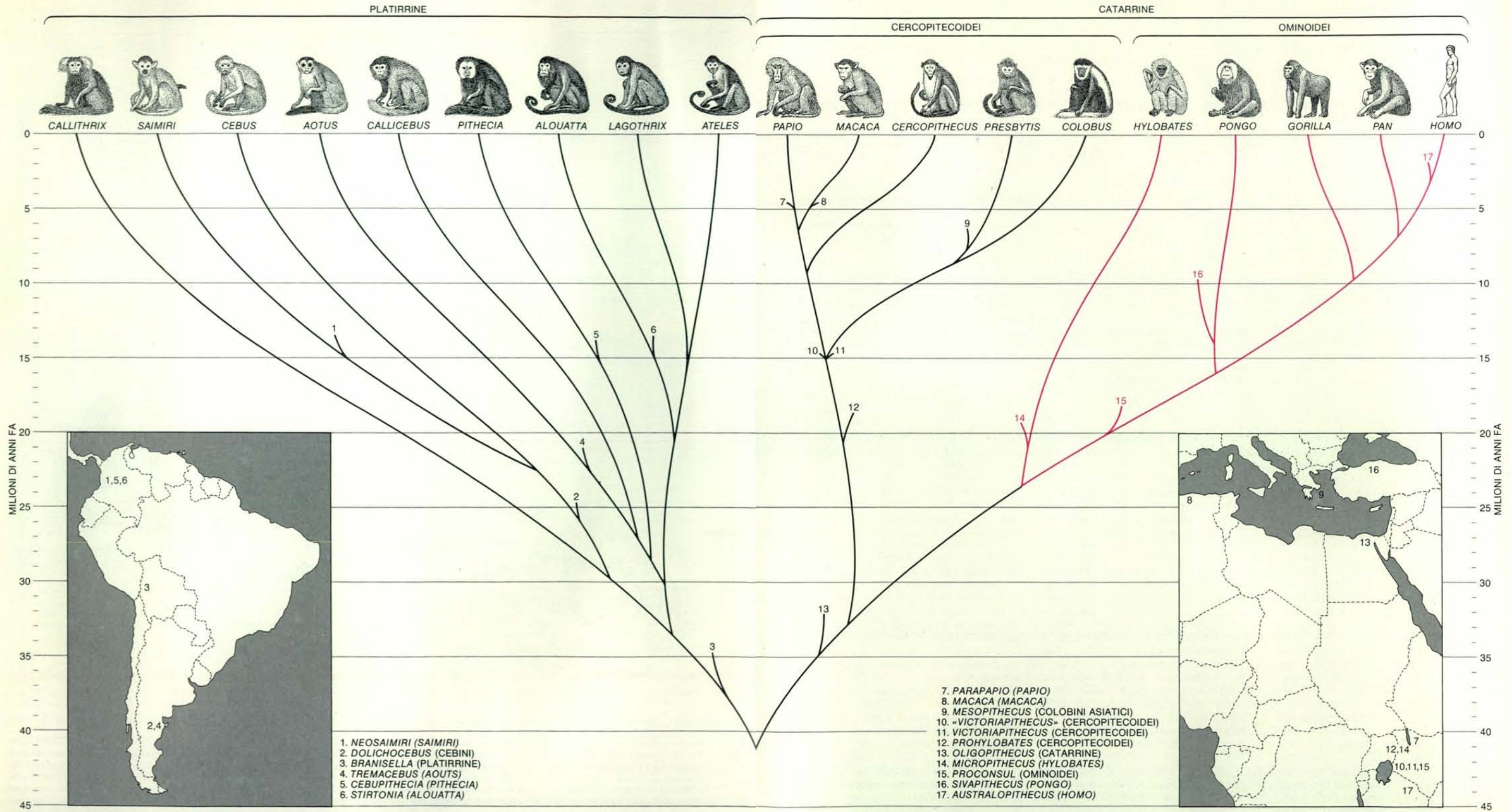
flette la circostanza che l'antenato comune agli ominidi, agli scimpanzé e ai gorilla sia vissuto in un tempo che era appena la metà circa di quello che ci separa dall'ultimo antenato comune degli ominoidei di più grosse dimensioni. Gli studi particolareggiati di anatomia comparata mostrano che il quadro delle somiglianze morfologiche fra ominoidei viventi conferma le scoperte della biologia molecolare. I grandi antropomorfi e gli ominidi non formano un gruppo monofiletico, ossia essi non hanno avuto un precursore comune dopo la divergenza della linea che ha portato all'uomo. Invece le scimmie antropomorfe africane e l'uomo formano un gruppo monofiletico rispetto agli

antropomorfi asiatici (ipotesi già avanzata da alcuni anatomisti del passato). Ciò significa che determinati caratteri comuni a tutti gli antropomorfi, ad esempio l'assenza della coda, devono essere stati ereditati dall'ultimo antenato di tutti i grossi ominoidei oppure essersi evoluti separatamente nelle singole specie. Grazie alle scoperte in biologia molecolare, i

antropomorfi asiatici (ipotesi già avanzata da alcuni anatomisti del passato). Ciò significa che determinati caratteri comuni a tutti gli antropomorfi, ad esempio l'assenza della coda, devono essere stati ereditati dall'ultimo antenato di tutti i grossi ominoidei oppure essersi evoluti separatamente nelle singole specie. Grazie alle scoperte in biologia molecolare, i

morfologi stanno ora riesaminando le caratteristiche anatomiche e le somiglianze fra ominoidei e stanno analizzando più attentamente i processi dello sviluppo embrionale e i loro controlli genetici. Questi «nuovi» schemi di ramificazione a livello molecolare non sono incompatibili con le testimonianze fossili degli ominoidei. Sono passati i tempi in cui

molti paleontologi situavano l'origine degli ominidi a metà del Miocene, 16 o 15 milioni di anni or sono, e molti biologi molecolari sostenevano che l'intera radiazione degli ominoidei si sarebbe svolta nel corso degli ultimi dieci milioni di anni e che gli ominidi si sarebbero separati solamente quattro milioni di anni fa. Oggi si tende piuttosto a collocare la radiazione

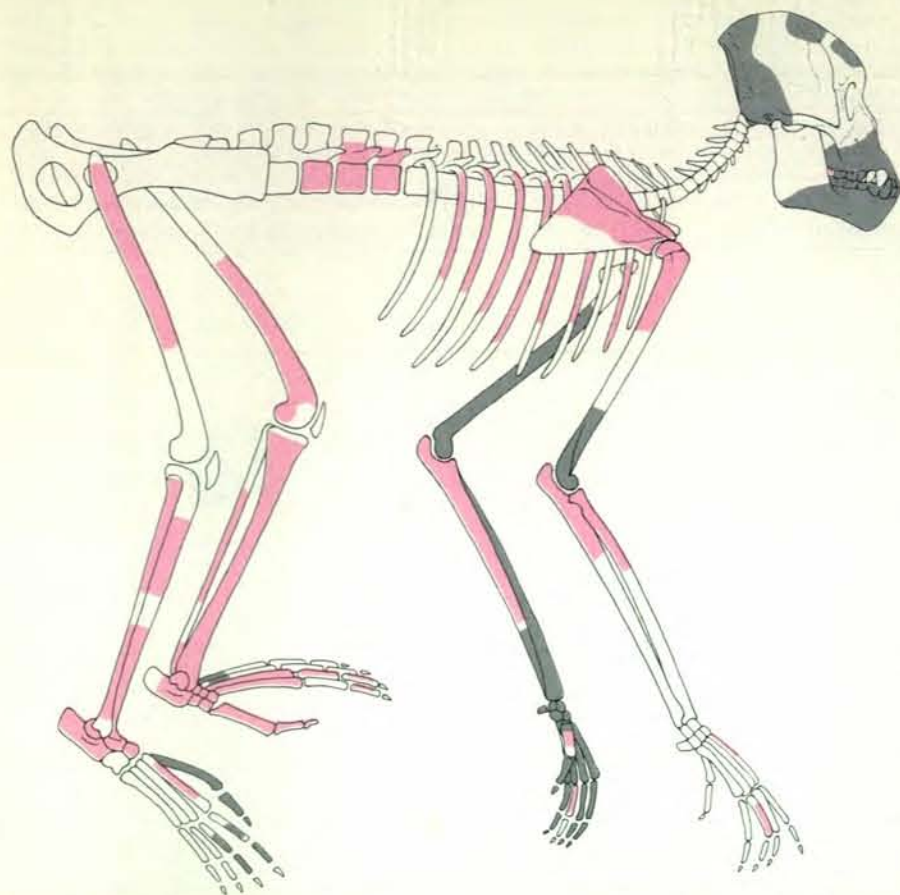


L'albero genealogico dei primati superiori, che si estende da 45 milioni di anni fa a oggi, si può ripartire in due grandi settori: le platirrine, o scimmie del Nuovo Mondo, a sinistra e le catarrine, o scimmie del

Vecchio Mondo, a destra. Le ramificazioni corrispondenti agli ominoidi e agli ominidi appaiono in colore all'estrema destra. Le varie successioni di divergenze si sono potute definire in base a dati biochimici

come quelli relativi a piccole differenze nella sequenza degli amminoacidi di una stessa proteina in specie viventi diverse. La datazione specifica delle diramazioni si basa sull'età dei fossili e su altri tipi di dati

geocronologici. Le cartine a sinistra e a destra indicano il luogo in cui è stato trovato il fossile più antico che rappresenta il genere. Il nome a destra di ciascun genere fossile indica la sua relazione con forme viventi.



In questa ricostruzione di *Proconsul*, un ominoide quadrupede, le parti anatomiche rinvenute nel 1951 da Tom Whitworth sono rappresentate in grigio. Le parti dello scheletro riprodotte in colore sono quelle trovate di recente in raccolte di museo da Alan Walker e Martin Pickford, gli autori di questa ricostruzione. Le parti di cui si nota solo il contorno, ad esempio il bacino, gran parte della colonna vertebrale e parti del cranio e della mandibola, non sono state ritrovate.

di tutti i grandi antropomorfi da un ceppo ancestrale verso la metà del Miocene e la divergenza degli ominidi da una linea di antropomorfi africani verso il tardo Miocene, forse otto o sette milioni di anni fa.

Per quale motivo la documentazione fossile degli ominoidei è stata male interpretata da modesti paleontologi come il sottoscritto? Per una serie di ragioni. Per prima cosa si è prestata eccessiva attenzione alle testimonianze fossili come fonti d'informazione sulle sequenze evolutive che hanno portato alla ramificazione, mentre oggi è chiaro che la documentazione di tipo molecolare può dirci molto di più sullo stesso argomento. (Ad ogni modo le testimonianze fossili rimangono l'unica fonte di conoscenza diretta degli stati evolutivi del passato.) Gli equivoci interpretativi vanno anche imputati al fatto che i reperti fossili degli ominoidei, risalenti a prima degli ultimi quattro milioni di anni dell'evoluzione degli ominidi, sono nel contempo frammentari e insufficientemente datati. Essi consistono quasi interamente di denti e di frammenti di mandibole, parti anatomiche che raramente forniscono informazioni utili ai fini tassonomici. Per esempio, le parti equivalenti del corpo dei mammiferi viventi generalmente non sono molto utili per suddividere gli individui in specie o per

riunire le specie in gruppi tassonomici superiori.

Una terza causa di malintesi sta proprio nella diversità fra ominoidei del medio e del tardo Miocene. I grossi ominoidei viventi non presentano una notevole diversificazione: solo cinque loro generi sono sopravvissuti. Quando si considera l'insieme delle testimonianze fossili degli ominoidei del passato, ne emerge, invece, un gruppo molto eterogeneo; sotto molti aspetti ogni specie è unica. Questi fattori hanno ostacolato la ricostruzione attraverso i fossili delle sequenze che hanno portato alla ramificazione degli ominoidei e agli ominoidei viventi. Risulta persino complicato collegare determinate specie fossili con singole specie viventi.

Nuovi, importanti fossili di ominoidei del Miocene hanno però modificato la situazione precedente. Parti scheletriche - della faccia, del cranio, degli arti - erroneamente interpretate in passato o scoperte solo di recente stanno fornendo un quadro più chiaro e spesso sorprendente sia dei rapporti evolutivi sia della natura dei primi ominoidei. È persino possibile intuire i loro modi di vita e le loro relazioni ecologiche.

Nel contempo, le analisi molecolari più recenti stanno chiarendo e sviluppando nuove idee sul quadro generale e sui tem-

pi dell'evoluzione degli ominoidei: basti ricordare il lavoro di Charles G. Sibley e di Jon Ahlquist della Yale University sull'ibridazione del DNA di specie diverse di primati allo scopo di stabilire le differenze e le somiglianze delle loro sequenze nucleotidiche. La paleoantropologia, dai tempi in cui le testimonianze fossili non portavano alcun contributo alla conoscenza degli schemi filogenetici, si avvia lentamente verso un tempo in cui esse costituiranno una componente della sintesi che rende compatibili fra loro dati di ogni tipo: fossili, anatomici, molecolari e riguardanti lo sviluppo.

La radiazione degli ominoidei

La versione generalmente accettata oggi della radiazione degli ominoidei può essere riassunta come segue. In primo luogo, non è chiaro quando le grandi scimmie antropomorfe si sono diversificate dai loro affini, le scimmie del Vecchio Mondo. (Le scimmie del Nuovo Mondo non hanno avuto alcun ruolo nell'evoluzione degli ominoidei). Sembra però quasi certo che la divergenza si sia manifestata dopo l'Oligocene medio (circa 30 milioni di anni fa) e anteriormente agli inizi del Miocene (circa 20 milioni di anni o sono). Comunque sia, la radiazione degli ominoidei era in corso durante il primo o medio Miocene (da 20 a 15 milioni di anni addietro) e alcune linee di questo gruppo, che ebbe moderatamente successo, non si sono estinte che nel tardo Miocene (da otto a sei milioni di anni fa).

Allo stesso modo, non è chiaro quando i piccoli ominoidei, i gibboni dell'Asia, si sono distaccati dai più grossi. Probabilmente la prova di questa scissione va ricercata in alcuni fossili di primati del primo Miocene, ma si tratta di una questione controversa. Esiste, invece, una migliore documentazione della divergenza delle grosse scimmie antropomorfe in specie asiatiche e africane, verificatasi verso la metà del Miocene, all'incirca 16 milioni di anni fa. L'orango è l'unico e anomalo sopravvissuto di un gruppo di scimmie antropomorfe asiatiche del medio e del tardo Miocene, che in tempi precedenti era più diversificato e più diffuso.

Non esistono purtroppo, in pratica, testimonianze fossili del ramo degli ominoidei africani per la maggior parte del Miocene. Però, in base ai risultati delle indagini molecolari, è possibile collocare nel corso del tardo Miocene la divergenza dei primi gorilla, seguita da quelle degli scimpanzé e degli ominidi (fra 10 e sei milioni di anni fa). È anche noto che, almeno da quattro a tre milioni e mezzo di anni fa, ominidi bipedi erano presenti in Africa orientale. Finché non sarà possibile disporre di testimonianze fossili africane più complete (particolarmente per quanto riguarda le scimmie antropomorfe) si potranno fare solo supposizioni sui tempi e sulla precisa sequenza che ha portato alla ramificazione di queste linee oltre che sulla natura dei loro primi rappresentanti, compresi i primissimi ominidi.

Se si risale all'Oligocene, le testimo-

nianze fossili dei primati africani sono relativamente abbondanti. Elwyn L. Simons della Duke University ha condotto saltuariamente, dal 1961, ricerche di fossili di primati nella regione di el-Fayyūm, in Egitto. A partire dal 1977, lui e i suoi collaboratori, John G. Fleagle della State University di New York a Stony Brook e Richard F. Kay della Duke University, hanno scoperto materiale veramente importante. Sono così disponibili oggi parti considerevoli di alcuni scheletri di *Aegyptopithecus*, una specie finora nota solo per un cranio rinvenuto in precedenza dallo stesso Simons, e di *Apidium*. Anche *Parapithecus* e *Propithecus*, rappresentati molto disorganicamente, si possono ora caratterizzare con relativa facilità.

Questi fossili rappresentano un gruppo diversificato di primati arboricoli non specializzati che avevano già raggiunto lo stato biologico di scimmie durante il medio Oligocene, almeno 30 milioni di anni or sono; non erano però primati primitivi come i lemuri, ma assomigliavano in linea generale alle scimmie attuali del Nuovo Mondo. Erano comunque troppo primitivi e troppo diversi dagli attuali primati superiori del Vecchio Mondo perché sia possibile determinare se siano caratteristici di un periodo anteriore alla divergenza degli ominoidei dalle scimmie del Vecchio Mondo, o posteriore a esso.

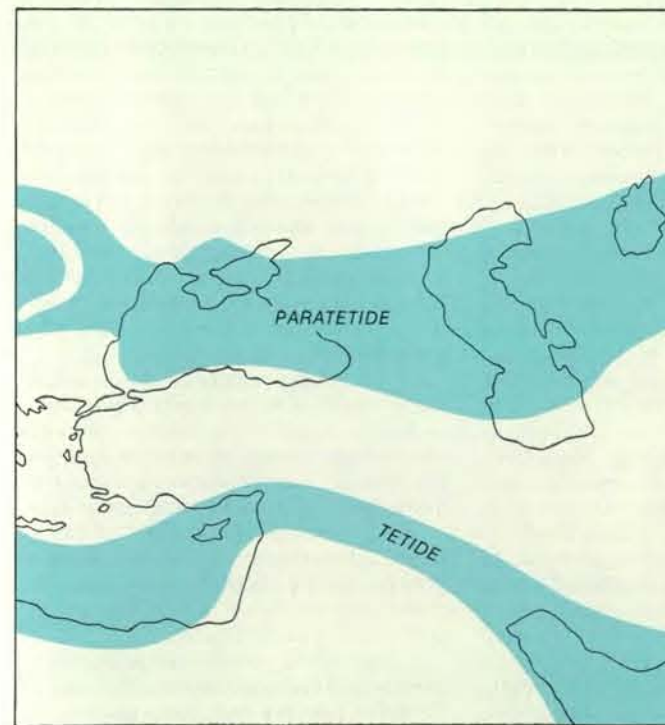
Se la documentazione fossile africana è nel suo complesso scarsa, gli strati del primo Miocene nell'Africa orientale costituiscono una felice eccezione. Gli ultimi 50 anni di ricerche in Kenya e in Uganda hanno portato alla luce circa

1000 fossili di primati superiori, che risalgono a un periodo compreso tra 17 e 12 milioni di anni fa e che quasi certamente rappresentano linee di ominoidei di un periodo posteriore alla loro divergenza dalle scimmie del Vecchio Mondo. La paleoecologia di quel periodo è oggi ben conosciuta, grazie principalmente al lavoro di Peter Andrews della sezione di storia naturale del British Museum. Il paesaggio era costituito da un insieme di foresta, bosco rado e prateria, con prevalenza della foresta che, al pari della boscaglia, era abitata da almeno sei specie di ominoidei. Una di queste, *Proconsul africanus*, è stata per qualche tempo la scimmia antropomorfa meglio conosciuta del Miocene; le conoscenze sul suo conto si sono oggi ancora più approfondite.

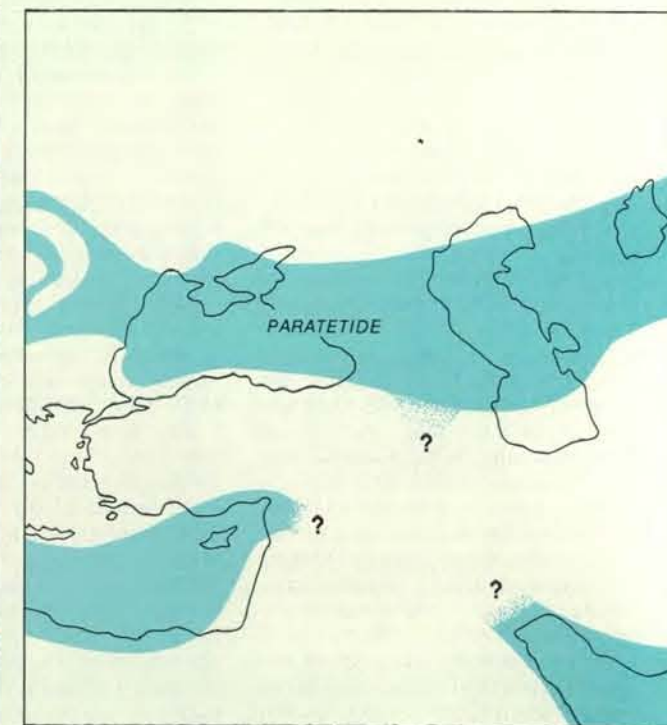
Proconsul, un piccolo animale delle dimensioni di un babbuino, ha avuto una parte preminente nello sviluppo del pensiero sull'evoluzione degli ominoidei. Fino a tempi recenti, di esso si conoscevano unicamente un arto anteriore quasi completo, una mandibola, frammenti del cranio e poche ossa del piede, appartenenti tutte allo stesso individuo, il cranio quasi completo di un altro individuo e resti disparati di arti, denti e mandibole. In un secondo tempo, Alan C. Walker della Johns Hopkins University e Martin Pickford dei Musei nazionali del Kenya hanno scoperto - nelle raccolte dei musei e dei laboratori - molti altri pezzi appartenenti al primo dei suddetti individui, del quale è stato possibile ricostruire gran parte dello scheletro.

Fino a oggi *Proconsul* era stato in-

terpretato in molti modi contraddittori. A un estremo si sosteneva che la sua forma era troppo specializzata perché potesse trattarsi dell'antenato di un qualsiasi ominoide vivente e probabilmente non era nemmeno un ominoide. Altri lo consideravano un ominoide primitivo non specializzato e altri ancora ipotizzavano che potesse essere un antenato dello scimpanzé. La disponibilità di nuovo materiale ha contribuito a chiarire in qualche misura il problema e oggi tutti concordano nel definire *Proconsul* un ominoide, anche se primitivo e dotato di scarsissima specializzazione. Si trattava di una «scimmia antropomorfa in formazione», arboricola, frugivora e con distinta differenziazione dimensionale fra i due sessi. Le sue caratteristiche complessive non hanno alcuna somiglianza con quelle di qualsiasi primate superiore attuale. Le articolazioni del gomito e della spalla, come pure il piede, sono simili a quelle dello scimpanzé, il polso ricorda quello di una scimmia e le vertebre lombari potrebbero appartenere a un gibbono. Molti altri suoi caratteri, come la configurazione generale, sono invece unici. *Proconsul* fornisce una salutare lezione agli studiosi dell'evoluzione: le affinità ipotizzate per questo animale dipendevano di volta in volta dalla parte anatomica che veniva studiata. Quando un animale fossile viene rinvenuto sotto forma di frammenti raccolti nel corso di un determinato arco di tempo, l'ordine stesso dei ritrovamenti influisce sulle interpretazioni filogenetiche, e in modo particolare nel caso di specie «a mosaico» come *Proconsul*.



Il congiungimento dell'Africa con l'Eurasia è posteriore a 20 milioni di anni fa, quando la Tetide (in colore) separò le due masse continentali, come viene rappresentato a sinistra. Nel primo Miocene, cioè circa 17



milioni di anni fa, il mare si ritirò, come è indicato a destra, lasciando un istmo che permise agli animali di migrare fra l'Africa e l'Eurasia. Fra i migratori in direzione dell'Eurasia vi furono alcuni ominoidei africani.

La fine del primo Miocene, circa 17 milioni di anni or sono, è stata contraddistinta da un evento geologico altamente significativo, ossia la deriva dei continenti, che ha congiunto l'Africa e l'Arabia con l'Eurasia, consentendo così la migrazione di ominoidi africani (e di altri mammiferi) nel resto del Vecchio Mondo. Questo congiungimento ha dato origine a importanti processi orogenetici: in Europa si sono formate le Alpi, in Turchia la catena del Tauro e in Iran i monti Zagros; è scomparsa la Tetide e tanto la circolazione oceanica quanto quella atmosferica si sono modificate provocando variazioni di grande scala nel clima e negli habitat.

Durante il medio e tardo Miocene, da 17 a cinque milioni di anni fa, il clima eurasiatico era stagionale, ma con inverni più miti di quelli attuali. Territori che attualmente si presentano come praterie o addirittura come deserti erano ricoperti da boschi e le foreste erano molto più diffuse. La fauna costituita dai mammiferi rifletteva queste condizioni; gli animali da pascolo nelle praterie erano scarsi, mentre erano più abbondanti gli animali che si nutrivano di ramoscelli, germogli e foglie di alberi e arbusti nella boscaglia e nella foresta.

I reperti fossili degli ominoidi di questo periodo in Eurasia consistono principalmente di mascelle o di denti e, di conseguenza, è difficile precisare le relazioni filogenetiche. Comunque, a grandi linee, gli ominoidi possono essere suddivisi in due grandi gruppi, uno dei quali può essere definito in modo approssimativo il gruppo dei driomorfi, rifacendosi alle varie specie europee del genere *Dryopithecus*, un primate con molte caratteristiche primitive da ominoide, mentre l'altro si può chiamare il gruppo dei ramamorfi, con riferimento a *Ramapithecus* e ai generi affini. Per quanto riguarda la distribuzione, i ramamorfi erano principalmente asiatici e differivano dagli altri ominoidi del Miocene per determinate caratteristiche, che ricordano quelle degli ominoidi posteriori.

Edward Lewis, un ricercatore laureatosi a Yale e che ha fatto parte in seguito dell'US Geological Survey, ha dato il nome *Ramapithecus* al frammento fossilizzato di una mascella di primate, che aveva trovato in India nel 1932. Oggi si sa che quel fossile è databile a poco più di sette milioni di anni fa. Negli anni sessanta sono stati trovati denti e altre mascelle di primati in India, in Pakistan e in altre regioni del Vecchio Mondo; essi sono stati attribuiti allo stesso genere. Questo animale appariva diverso dai suoi contemporanei *Dryopithecus* e *Sivapithecus*, un altro fossile di origine indiana, che si presumevano più simili alle scimmie antropomorfe. I suoi vari resti mostrano che aveva mascelle robuste e grossi molari con corone rivestite di uno spesso strato di smalto. Questi e alcuni altri caratteri assomigliano a quelli degli ominidi africani posteriori del genere *Australopithe-*

cus. Tali somiglianze hanno indotto molti, compreso chi scrive, a sostenere che *Ramapithecus* sarebbe, di fatto, uno dei primi ominidi, che questi ultimi si sarebbero distaccati dagli ominoidi africani almeno 15 milioni di anni fa e che, di conseguenza, era assolutamente erroneo far risalire questa divergenza a quattro milioni di anni addietro, come portavano a concludere le indagini condotte a livello molecolare.

Nel 1973 ebbe inizio la mia collaborazione a un progetto di ricerca del Servizio geologico del Pakistan, inteso a ritrovare ulteriori fossili di *Ramapithecus*. Gli strati ricchi di fossili su cui stiamo ancora lavorando sono il prodotto dell'erosione indotta dal sollevamento della catena dell'Himalaya e prendono il nome di «serie dei Siwalik». La loro età varia da uno a 17 milioni di anni ed essi formano uno spesso cuneo che si estende lungo il fianco meridionale della grande catena montuosa tra l'Afghanistan e la Birmania. Nella sezione pakistana di tale formazione abbiamo avuto la fortuna di trovare molti nuovi fossili di ominoidi, compresi esemplari sia di *Ramapithecus* sia del suo affine *Sivapithecus*, oltre a migliaia di altri fossili di mammiferi.

Il nostro numeroso gruppo interdisciplinare ha ora completato una sequenza cronologica delle variazioni faunistiche che si sono verificate durante la sedimentazione protrattasi in questa zona per 16 milioni di anni, e ha fatto considerevoli progressi nella ricostruzione delle comunità animali locali del passato. Lo studio di questo nuovo materiale mi ha convinto che *Ramapithecus* e *Sivapithecus* siano molto più simili tra loro di quanto pensassi prima e potrebbero persino appartenere allo stesso genere.

A ogni modo i nuovi fossili di *Sivapithecus*, costituiti da una porzione di cranio e da parti isolate di arti, presentano una serie di caratteristiche, particolarmente per quanto riguarda la faccia e il palato, che assomigliano a quelle dell'unica grande scimmia antropomorfa asiatica oggi vivente, l'orango (*Pongo pygmaeus*), e non, come ci si aspettava, a quelle di *Australopithecus*. Le somiglianze con la mandibola e i denti di *Australopithecus* esistono, ma non implicano alcun significato filogenetico. Si tratta probabilmente di esempi di evoluzione parallela o di caratteri assunti dall'ultimo antenato comune di tutti gli ominoidi attuali. Questa è perlomeno l'opinione mia e del mio collega Steven Ward dalla Kent State University. Alcuni fossili di ominoidi del Pakistan, che hanno almeno 12 milioni di anni, presentano caratteristiche simili; se la nostra interpretazione è corretta, la separazione fra ominoidi africani e asiatici va sicuramente situata in quel periodo, se non oltre. Forse essa ha coinciso approssimativamente con il congiungimento fra Africa e Asia, prodotto dalla deriva dei continenti e con il conseguente rimescolamento delle faune, avvenuto 17 o 16 milioni di anni fa. Attualmente mi sentirei di collocare questa separazione a circa 16 milioni di anni or sono.

I primi ominidi

I primi autentici ominidi fanno la loro comparsa forse tra 4 e 3,75 milioni di anni fa, in Tanzania e in Etiopia. La loro precedente evoluzione è ancora avvolta nell'oscurità. Le modifiche verificatesi nella fauna e negli habitat nell'arco di tempo compreso fra 7,5 e 4,5 milioni di anni addietro rispecchiano un certo numero di avvenimenti di ordine climatico e geografico, il più spettacolare dei quali è stato il ripetuto prosciugamento del Mediterraneo, circa 5,5 milioni di anni fa. Non è però chiaro il legame fra questi eventi e l'evoluzione degli ominoidi africani. Comunque le impronte di ominidi che camminavano in posizione eretta, scoperte da Mary Leakey e dai suoi collaboratori a Laetoli in Tanzania, sono state sicuramente datate a 3,75 milioni di anni fa e lo scheletro per metà completo di «Lucy» assieme ad altri reperti di *Australopithecus afarensis*, trovati da Donald Johanson e collaboratori a Hadar, in Etiopia, lo è stato ma con minore certezza fra quattro e tre milioni di anni fa. Esistono alcune divergenze sulla tassonomia di Lucy e dei resti di ominidi a essa associati, rinvenuti a Hadar. Rappresentano forse essi una nuova singola specie, *Australopithecus afarensis*? Oppure una sottospecie settentrionale di *A. africanus*? O due o più specie? Lasciando da parte questi interrogativi, considererò una sola versione, pur ammettendo che il problema non è ben definito.

A. afarensis presenta caratteristiche che sono per la maggior parte primitive. Le dimensioni dei maschi adulti superavano dal 50 al 100 per cento quelle delle femmine e i singoli individui pesavano da 25 a 50 chilogrammi o più. Il volume del cervello corrispondeva a quello di una grossa scimmia antropomorfa africana ma, in presenza di una così ampia gamma di pesi per gli adulti, non è possibile stabilire se il volume cerebrale fosse proporzionalmente più piccolo o più grande rispetto al peso corporeo. Per alcune caratteristiche delle ossa facciali e del palato questi primi ominidi assomigliavano agli scimpanzé, ma complessivamente il cranio, così come è stato ricostruito, ha più l'aspetto di quello di una femmina di gorilla: *A. afarensis* ha denti più grossi ed è più robusto di uno scimpanzé.

La dentatura presenta alcune caratteristiche che si possono definire primitive, anche se i canini di entrambi i sessi presentano una corona bassa e non sono più proiettati in avanti come nelle scimmie; i molari sono relativamente grossi e ricoperti da uno smalto spesso, probabilmente per adattamento alla masticazione di grandi quantità di frutti, semi, bacelli, radici e tuberi, alcuni dei quali potevano anche essere molto resistenti.

È impossibile provare che le impronte trovate in Tanzania siano state prodotte da ominidi simili a quelli rappresentati dai fossili di Hadar, ma la morfologia delle articolazioni dell'anca, del ginocchio e della caviglia di Lucy e dei suoi compagni rivela che la popolazione di Hadar era

MILIONI DI ANNI FA	EPOCA	OPINIONI DEL 1978	OPINIONI DEL 1984
0,045	PLEISTOCENE FINALE	Comparsa di un <i>H. sapiens</i> anatomicamente moderno.	Minime variazioni rispetto alle fasi precedenti
0,125	TARDO PLEISTOCENE	Comparsa di <i>H. sapiens</i> .	
0,7	PLEISTOCENE MEDIO	Graduale progresso evolutivo degli ominidi.	Comparsa di un <i>H. sapiens</i> anatomicamente moderno (chiaramente diverso dall' <i>H. sapiens</i> arcaico sotto l'aspetto sia cranico sia subcranico). Abbondanza di utensili di pietra e d'osso e di altri manufatti.
1,6	PRIMO PLEISTOCENE	Comparsa di <i>H. erectus</i> ; estinzione di <i>Australopithecus robustus</i> .	Nuovi tipi di utensili e di tecniche; significativa elaborazione del comportamento.
5	PLIOCENE	Presenza di ominidi gracili e bipedi in Tanzania e in Etiopia (compreso <i>A. afarensis</i>).	Utensili litici di maggiore complessità. Comparsa dell' <i>H. sapiens</i> arcaico (leggermente diverso da <i>H. erectus</i>). Uso del fuoco?
11	TARDO MIOCENE	Divergenza tra le scimmie antropomorfe africane e i protominidi africani.	Utensili litici un poco più complessi.
17	MIOCENE MEDIO	Affermazione di <i>Ramapithecus</i> come protominideo.	Presenza di <i>H. erectus</i> . Estinzione di <i>H. habilis</i> e poi di <i>A. boisei</i> e <i>A. robustus</i> . Primi utensili litici; presenza nell'Africa meridionale di <i>A. africanus</i> e <i>A. robustus</i> . Diversificazione degli ominidi; <i>A. boisei</i> e <i>H. habilis</i> coesistono. Impronte di Laetoli. Presenza nell'Africa orientale di un <i>Australopithecus</i> bipede.
24	PRIMO MIOCENE	Gli ominoidi divergono dalle scimmie del Vecchio Mondo.	7,5-4,5 milioni di anni fa: sostanziali mutamenti climatico-faunistici; diversificazione degli ominoidi africani (deduzione basata su dati molecolari).
	OLIGOCENE		Divergenza fra ominoidi africani e ominoidi asiatici: <i>Sivapithecus</i> (<i>Ramapithecus</i>) imparentato con <i>Pongo</i> ?
			Inizio della radiazione degli ominoidi.
			Inizio della radiazione delle catarrine.

Il mutamento di opinioni sull'evoluzione degli ominoidi e degli ominidi risulta evidente da questa tabella che si estende dall'Oligocene, in basso, al più recente Pleistocene, in alto. I numeri inseriti fra le suddivisioni geologiche indicano gli estremi temporali di ogni suddivisione, in

milioni di anni, prima del presente. A sinistra sono riportate le opinioni generalmente accettate nel 1978 dagli studiosi su questo argomento, unitamente a prove di carattere fossile o archeologico addotte a sostegno delle stesse. A destra sono riportate invece le opinioni attuali.

chiaramente bipede. Le ossa del piede mostrano il medesimo adattamento alla locomozione bipede e un'ulteriore interessante caratteristica: le ossa dell'alluce e i metatarsi sono lunghi e ricurvi. Rispetto agli individui del genere *Homo*, *A. afarensis* aveva braccia lunghe e gambe corte e le sue mani avevano una forte presa. Giudicando dalle proporzioni delle loro ossa e dalla morfologia delle loro articolazioni, esse erano in grado di operare con maggior precisione di quelle degli attuali scimpanzé.

Gli ominidi di Hadar vivevano in una regione di boschi e di savane lontano dalle foreste del Pliocene e facevano parte di una comunità di mammiferi assai simili sotto il profilo strutturale alle successive comunità del Pleistocene. Queste comunità avevano cominciato a svilupparsi nell'Africa orientale nel tardo Miocene ed è possibile che comprendessero ominidi che non ci sono ancora noti. Per analogia con i mammiferi attuali è poco probabile che gli ominidi di Hadar fossero monogami.

A questi ominidi non sono state associate pietre lavorate per la fabbricazione di utensili, ma questo rilievo potrebbe non essere del tutto pertinente. Supponiamo che *A. afarensis* impiegasse lo stesso tipo di utensili degli attuali scimpanzé:

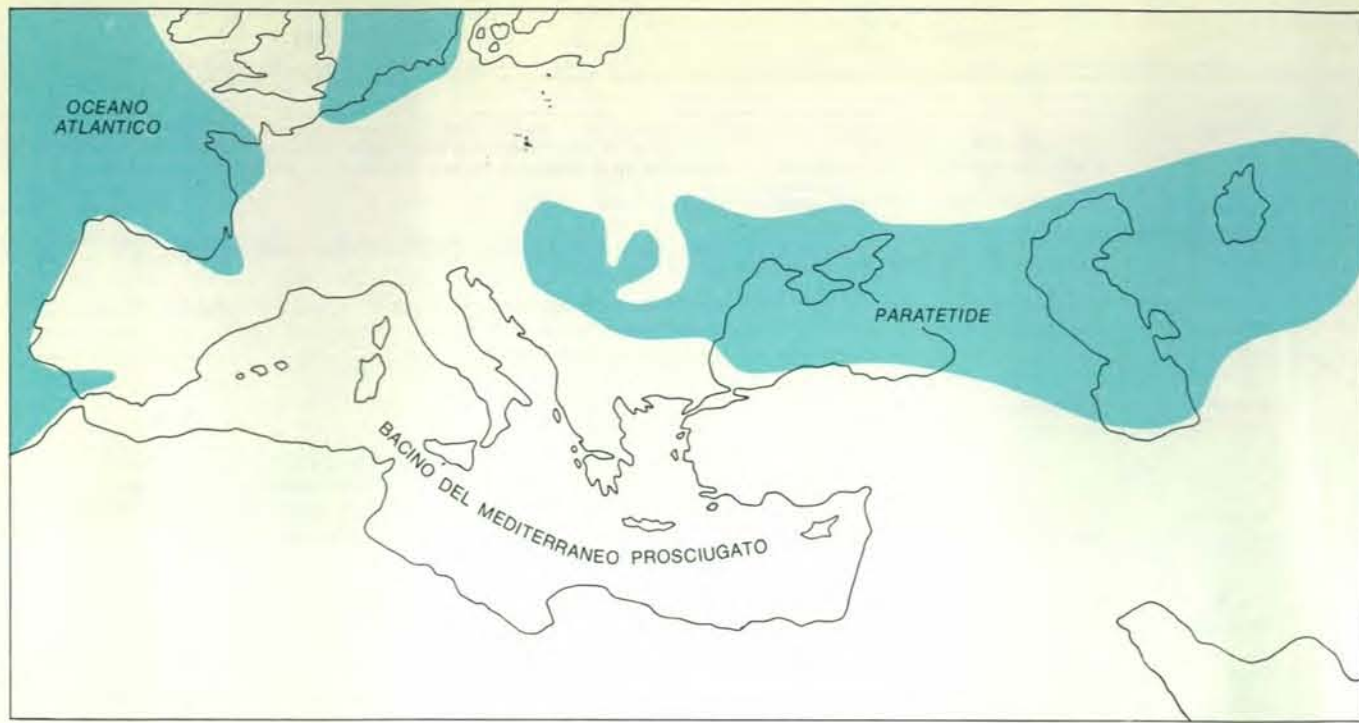
fusti, foglie, pezzi di legno o pietre che si prestavano casualmente a raccogliere e a lavorare il cibo o a comportamenti dimostrativi. E supponiamo anche che questo impiego fosse più intenso di quello praticato oggi dagli scimpanzé. Sarebbe impossibile riconoscere oggi, a quattro milioni di anni di distanza, questi utensili, anche se si fossero conservati per circostanze quasi miracolose.

Un'utilizzazione di questo tipo, se fosse possibile postularlo, potrebbe essere stata una componente importante di quei comportamenti che hanno stimolato la deambulazione bipede e la riduzione dei canini. Concetti di questo genere sono molto dibattuti. Gli ominidi di Hadar erano protocacciatori e assassini, spazzini opportunisti o miti vegetariani? La monogamia si è evoluta con la condizione bipede? I maschi con legame di coppia portavano il cibo alla loro compagna in un sito che fungeva loro da domicilio? Il cibo raccolto, animale o vegetale, veniva condiviso oppure gli ominidi di Laetoli e di Hadar se lo procuravano individualmente? Ciascuno di questi punti di vista, per non parlare di altri, ha i propri sostenitori. Per il momento, comunque, qualsiasi rappresentazione di questi primi ominidi come esseri viventi deve essere rimandata a

quando vi sarà maggiore consenso sulla natura dei loro adattamenti essenziali. Soprattutto dobbiamo resistere alla tentazione di vederli come esseri umani moderni separati da noi solo da un determinato periodo di tempo.

Avvicinandosi all'uomo

In un momento non precisato, fra 2,5 e due milioni di anni fa, gli ominidi africani hanno dato origine a una modesta radiazione adattativa. Circa due milioni di anni fa erano presenti almeno due, probabilmente tre, e forse anche più specie di ominidi, anche se in qualsiasi parte dell'Africa non sono state trovate prove che ne esistessero più di tre. Nell'Africa orientale, *A. boisei*, una specie di *Australopithecus*, robusta e con denti molto grossi, viveva contemporaneamente a un ominide con cervello più sviluppato, il cui cranio fossile tipo è catalogato in Kenya con la sigla ER-1470. Questa specie è oggi universalmente nota sotto la denominazione *Homo habilis*, coniata nel 1964 da L.S.B. Leakey, da Philip V. Tobias e da John R. Napier e applicata a ER-1470 dai suoi scopritori, Richard Leakey e collaboratori. Probabilmente coesisteva nell'Africa orientale una terza specie di ominidi con denti e cervello di



Il prosciugamento del bacino del Mediterraneo dovuto alla sua separazione dall'Atlantico a causa della formazione di montagne deve aver avuto conseguenze catastrofiche sul clima e sull'ecologia dell'Africa.

Questo episodio è avvenuto nel tardo Miocene circa 5,5 milioni di anni fa. Non sono chiari i suoi effetti sui mammiferi, ominoidei compresi, ma gli ominidi sono comparsi forse un milione e mezzo di anni dopo.

piccole dimensioni. Inoltre, nell'Africa meridionale vivevano un'altra specie di *Australopithecus*, *A. robustus*, e forse anche *H. habilis*. Già 1,75 milioni di anni fa, comunque, *H. habilis* era scomparso dall'Africa ed era stato sostituito da *Homo erectus*, un ominide con un cervello di dimensioni ancora maggiori. Si sta così lentamente delineando un'immagine più chiara di questi animali, ma è ovvio che essa non può essere autenticamente definitiva e che bisogna evitare di scriverla troppo da vicino per non correre il rischio che si dissolva in una miriade di punti senza significato.

Si consideri *A. boisei*. Questa specie, più grossa sia di *A. afarensis* sia della specie sudafricana di *Australopithecus*, aveva un cervello piccolo e il maschio era di taglia nettamente più grossa di quella della femmina. Essa è comparsa più di due milioni di anni fa e si è conservata, a quanto sembra con scarse modifiche morfologiche, per un milione di anni. È possibile che *A. boisei* abbia utilizzato rami per scavare e pietre per martellare, ma non vi sono indicazioni che usasse frequentemente utensili.

Pur essendo di taglia superiore rispetto ad altre specie di *Australopithecus*, *A. boisei* potrebbe avere avuto lo stesso comportamento. Gli studi condotti sulle superfici masticatorie dei denti di varie specie di *Australopithecus* da Alan Walker e Frederick Grine della State University di New York a Stony Brook fanno pensare che esse fossero tutte vegetariane e piuttosto simili alle scimmie antropomorfe e che le loro diete, per molti versi simili, richiedessero una masticazione

lunga e ripetitiva. Probabilmente, a differenza delle scimmie antropomorfe che vivevano nelle foreste, si nutrivano più di radici e di tuberi e meno di frutti.

Perché si è evoluto *A. boisei* e da che cosa? La sua evoluzione è stata il risultato di una variazione del clima o dell'habitat o forse di una modificazione nella disponibilità di certi tipi di piante di cui si nutrivano? Perché si è estinto *A. boisei* dopo un milione di anni di apparente stabilità? Non si può ancora rispondere a questi interrogativi; comunque, nell'Africa orientale, la ricerca di fossili è molto intensa e si può sperare in una migliore comprensione futura sia di *A. boisei* sia dei suoi predecessori e delle loro forme di transizione.

Una questione completamente diversa è, invece, capire il processo che, da un predecessore (che potrebbe anche essere lo stesso di *A. boisei*), ha portato a *Homo habilis*. È difficile «immaginarsi» le varie specie di *Australopithecus*, cioè ricostruirle come se fossero animali ancora viventi, poiché oggi non esiste nulla di simile al mondo. Eppure le si possono intuire confusamente come particolari specie di scimmie antropomorfe. Ancor più difficile è formarsi un'idea di *H. habilis*, per vari aspetti alquanto simile ad *Australopithecus* (per esempio nelle ossa facciali e nei denti), ma con un cervello nettamente più grosso: mediamente attorno ai 700 centimetri cubi. Non vi sono precise correlazioni fra i crani di *H. habilis* e le altre ossa che caratterizzano questa specie. Ma le ossa degli arti che si ritiene lo rappresentino, a differenza di quelle di *Australopithecus*, assomigliano a quelle di specie po-

steriori del genere *Homo* (con l'eccezione dell'*Homo sapiens* attuale). È stato ipotizzato che tale rassomiglianza rifletta modificazioni nel modo di camminare e nelle dimensioni della pelvi femminile, dettate dal volume del piccolo al momento della nascita.

Quasi contemporanei alla comparsa in Africa di *A. boisei* e di *H. habilis* sono i loro primi siti archeologici: ammassi di pietre utilizzate o lavorate, spesso portate da luoghi distanti, frammenti a resti di animali. Si è concordi nel ritenere che questi siti indichino un cambiamento nella dieta, con inclusione di una maggiore proporzione di cibo di origine animale. Questo fatto riguarda più *H. habilis* che *A. boisei*. Vi è poi una correlazione causale anche fra il mutamento di dieta e lo sviluppo delle dimensioni cerebrali. Al di là di questo le opinioni divergono. *H. habilis* era forse un cacciatore-raccoglitore che trasportava il cibo verso basi fisse per poi distribuirlo e che praticava una suddivisione dei compiti, per cui la caccia era riservata ai membri di sesso maschile e la raccolta a quelli di sesso femminile? Il significato dell'interrogativo è questo: *H. habilis* aveva un comportamento «moderno» oppure era ancora essenzialmente vegetariano, con eventuali integrazioni alla dieta di carne ottenuta principalmente da carogne?

Le stesse testimonianze archeologiche sono ambigue: le medesime distribuzioni di utensili e di ossa potrebbero significare un'attività opportunistica di sfruttamento delle carogne per consumo individuale, con trasporto e condivisione ridotti al minimo o nulli, oppure potrebbero rap-

presentare basi di insediamento, punti di intensa interazione sociale. Per concludere, l'immagine di *H. habilis* rimane poco chiara. Si rendono necessari ulteriori ritrovamenti di materiali fossili e di siti archeologici, ma è comunque probabile che la transizione da *Australopithecus* a *Homo* rimanga oscura, poiché entrambi gli animali sono troppo diversi da quelli oggi esistenti. Che cosa è accaduto a *Homo habilis*? È sopravvissuto per qualche centinaio di migliaia d'anni per essere poi sostituito da una specie molto più duratura, *H. erectus*. *H. habilis* era il predecessore di *H. erectus*? Anche se molti concordano con questa ipotesi, essa potrebbe non essere corretta.

Il disegno si precisa

Homo erectus è la prima specie di ominidi a vasta diffusione. Per quanto si sa, è comparsa molto presto in Africa e, di fatto, potrebbe avere avuto origine in quella regione 1,6 milioni di anni o sono. Comunque sia, un milione di anni fa essa era già presente nell'Asia sudorientale e orientale, dove è sopravvissuta almeno fino a 300 000 anni fa. In quest'arco di tempo, che supera abbondantemente il milione di anni, la documentazione fisica di *H. erectus* dimostra una sua prolungata stabilità morfologica.

H. erectus assomigliava a specie più tardive di *Homo* (eccezionato il moderno *H. sapiens*) sia per dimensioni corporee sia per robustezza. Aveva un cervello più sviluppato di *H. habilis* (più di 800 centimetri cubi), i suoi incisivi erano grossi come quelli dei primi ominidi, mentre i molari e le ossa facciali erano più piccoli. La documentazione archeologica indica che alcune popolazioni di *H. erectus* fabbricavano utensili litici più grossi, scheggiati simmetricamente: bifacciali o «asce a mano». È inoltre possibile che altre conoscessero l'uso del fuoco. Alcune popolazioni posteriori potrebbero essersi impadronite di tecniche più sofisticate per produrre e modificare utensili litici, ma l'impressione generale è quella di una prolungata stabilità comportamentale e morfologica. Arthur Jelinek dell'Università dell'Arizona ha proposto che la definizione più appropriata di questo comportamento sia quella di comportamento «paleoculturale», per differenziarlo dai rapidi mutamenti che siamo abituati ad associare al comportamento «culturale» degli uomini contemporanei.

I paleoantropologi si collocano in una prospettiva diversa, spesso senza rendersene conto, quando considerano l'evoluzione umana al di là dell'ultimo milione e mezzo di anni. I problemi inerenti all'evoluzione degli ominoidei e ominidi che ho finora esposto riguardano questioni macroevolutive: adattamenti a grandi linee, tendenze evolutive e speciazione. Gli ominidi denominati *H. erectus*, gli *H. sapiens* «arcaici» (compresi gli uomini di Neandertal) e gli *H. sapiens* moderni probabilmente rappresentano un continuum mentre i cambiamenti all'interno di queste linee sono di natura microevolutiva.

Uno dei più interessanti lavori paleoantropologici riguarda nondimeno l'ultimo grande scalino dell'evoluzione umana: il passaggio, verificatosi 45 000 o 40 000 anni fa, dall'*Homo sapiens* arcaico a quello moderno.

Erik Trinkaus, dell'Università del New Mexico, ha dimostrato che i neandertaliani (e probabilmente anche i loro contemporanei di altre zone del Vecchio Mondo) erano diversi da noi non solo per il comportamento, ma anche fisicamente. Gli scheletri degli uomini di Neandertal erano molto più robusti e le inserzioni dei muscoli sulle ossa rivelano che essi erano anche molto più forti di noi. I denti erano più grossi e quelli ritrovati sono molto logorati, probabilmente perché (come è il caso degli attuali eschimesi) venivano usati per vari scopi estranei alla nutrizione (ad esempio per masticare le pelli allo scopo di ammorbidirle). La transizione a *Homo sapiens* è contraddistinta dalla perdita della robustezza neandertaliana nello scheletro, nelle ossa facciali e nei denti. C'erano anche modifiche nella morfologia del bacino della femmina, attribuibili forse al fatto che i parti erano divenuti più difficili che in precedenza, o addirittura che il periodo di gestazione si stava riducendo verso gli attuali nove mesi, dai probabili 11 primitivi (un periodo che si accorda con le conclusioni tratte dalle correlazioni generali fra volume corporeo materno, dimensioni fetali e durata della gestazione nei mammiferi).

Le testimonianze archeologiche registrano anche mutamenti comportamentali importanti, fra cui una proliferazione di utensili litici e ossei più avanzati, cambiamenti negli schemi venatori, nell'impiego e nel controllo del fuoco, nell'uso degli indumenti, nelle modalità d'insediamento, nelle dimensioni delle popolazioni, negli ambiti ecologici, nell'arte e in altre manifestazioni di attività rituale. Tutto questo punta alla comparsa di una specie in possesso di capacità (e di un potenziale) comportamentali moderne, a partire da una specie ancestrale che, almeno in base ai canoni moderni, mancava di alcune caratteristiche umane significative.

Una breve rassegna come questa non può rendere appieno la ricchezza e la complessità delle informazioni disponibili per lo studio di questa recentissima transizione, anche se di tipo microevolutivo. Comunque, fra tutte le transizioni degli ultimi 30 milioni di anni, essa è ancora la meglio documentata e potenzialmente la più comprensibile. Come studioso che lavora in un campo diverso, su un arco di tempo che può anche, in un qualsiasi periodo di 100 000 anni, non contenere fossili di rilievo, provo una certa invidia per coloro che lavorano su questo periodo molto recente, che rappresenta lo smalto finale della storia evolutiva dell'uomo. Bisogna però tenere presente che si tratta solo dello smalto: il corpo sottostante - i molti stadi precedenti di questa storia - richiede ancora un lungo e difficile lavoro d'indagine.

★ ASTRONOMIA DI BASE IN IMMAGINI

276 diapositive, suddivise in 9 sets corredate da didascalie formulate per dare una chiara e completa comprensione dei fenomeni astronomici. Un utile sussidio per l'apprendimento dell'astronomia nelle scuole medie inferiori e superiori; adatta anche per le persone interessate senza una conoscenza specifica della materia.

9 sets sono così suddivisi:
 CMIS 01 - IL SISTEMA SOLARE (36 dia)
 CMIS 02 - IL SOLE (24 dia)
 CMIS 03 - MERCURIO E VENERE (24 dia)
 CMIS 04 - LA TERRA (36 dia)
 CMIS 05 - MARTE (24 dia)
 CMIS 06 - GIOVE (36 dia)
 CMIS 07 - SATURNO (36 dia)
 CMIS 08 - LA NOSTRA GALASSIA (24 dia)
 CMIS 09 - L'UNIVERSO (24 dia)
 Set di 24 dia L. 23.000; Set di 36 dia L. 28.500
 (+ spese postali e di contrassegno)
 La serie completa dei 9 sets è offerta a L. 216.000 con in omaggio la Mappa della Luna Rand McNally.

★ COSTELLAZIONI BOREALI

Elegante e pratica confezione contenente:
 • 30 diapositive raffiguranti 48 costellazioni
 • una guida di 80 pagine con la descrizione dettagliata di ciascuna costellazione
 • un piccolo visore portatile a pile per poter confrontare direttamente le costellazioni osservate in cielo con quelle in diapositiva.
 L. 40.000 (+ spese postali e di contrassegno)
 offerta di lancio per i lettori di Le Scienze L. 37.000 CMIAT 09

★ 3 UTILI MANUALI

per l'appassionato di astronomia
 • Fotografare con il telescopio CMIAT 001
 • Come usare il telescopio CMIAT 002
 • Come montare il telescopio CMIAT 003
 Ricchi di illustrazioni, tabelle, consigli pratici alla portata di tutti.
 L. 7.000 ciascuno
 (+ spese postali e di contrassegno)
 3 volumi L. 19.000
 (+ spese postali e di contrassegno)

★ COSMO CALENDAR

l'astrolabio per il giovane astronomo
 Uno strumento indispensabile per conoscere la posizione e il movimento degli astri nel cielo in ogni giorno dell'anno. Realizzato in cartone stampato in vivaci colori è di uso semplicissimo.
 L. 6.500 (+ spese postali e di contrassegno) CMIAT 03

★ CARTE STELLARI

per giovani astronomi
 10 tavole di grande formato (cm 33x47) stampate a 2 colori e una guida per la consultazione raccolte in un'unica confezione.
 • stelle fino alla sesta magnitudine • stelle variabili • ammassi aperti • galassie • ammassi globulari • nebulose • nebulose planetarie
 L. 12.500 (+ spese postali e di contrassegno) CMIAT 04

★ SCIENZA DELLO SPAZIO

Il nuovo catalogo con tutti i prodotti e le novità disponibili. Sedici pagine a colori.
 • film Super 8 e 16 mm sonori • videocassette per qualsiasi marca di televisore • diapositive • globi • mappe • planetario • manuali • posters • atlanti
 Richiedetelo inviando L. 1.500 in francobolli.
 ordine minimo L. 10.000

Le richieste vanno indirizzate a:
Cosmo-Media
 Via Nino Bixio 1/D 22100 Como
 oppure telefonando (al mattino)
 al numero 031/558444

I suoni muscolari

Generati durante la contrazione, non sono rilevati dall'orecchio umano perché questo non è sensibile alle basse frequenze; oggi sono studiati per la loro possibile applicazione in campo scientifico e medico

di Gerald Oster

In un uomo in stato di veglia i muscoli corporei si contraggono e si rilassano continuamente e, per quanto strano possa sembrare, mentre si contraggono essi generano un suono. Dato che questo suono non viene udito in condizioni normali, potreste anche essere scettici. Se lo siete, provate a mettere entrambi i pollici con delicatezza nelle orecchie e serrate il pugno. Sentirete un debole rombo. Quanto più stringerete, tanto più forte sarà il suono. Quello che ascolterete, sarà il suono generato dai muscoli dell'avambraccio che si contraggono.

I suoni generati dai muscoli non sono normalmente udibili in quanto l'orecchio umano è insensibile ai suoni di bassa frequenza; la loro principale frequenza è infatti di 25 hertz (cicli al secondo) ed è vicina al limite inferiore dell'udito. Benché la loro esistenza sia nota da secoli, i fisiologi e i medici si sono limitati a considerarli un disturbo, tuttavia i recenti sviluppi nell'informatica e nella strumentazione medica hanno permesso di indagare su di essi nei particolari e i risultati sono molto interessanti. La ricerca che ho condotto con le nuove metodiche suggerisce che il suono è generato dalle fibre muscolari mentre il muscolo si contrae. Questi risultati potrebbero portare a una migliore comprensione della fisiologia muscolare, delle patologie muscolari (incluse quelle relative al cuore) e persino di alcuni tipi di comunicazione animale.

La prima citazione di ascolto dei suoni muscolari mettendo i pollici nelle orecchie è quella del gesuita Francesco Maria Grimaldi nel suo libro *Physico-mathesis de lumine, coloribus et iride aliisque adnexi libri duo*, pubblicato postumo nel 1665. Grimaldi è principalmente conosciuto per aver scoperto la diffrazione della luce, ma si è occupato anche di acustica. Egli attribuiva il rombo generato dai muscoli ai «movimenti precipitosi degli spiriti animali». Fin dal tempo di Galeno, medico greco del secondo secolo, c'era stato un generale accordo nel ritenere che i movimenti volontari fossero governati da un fluido - gli spiriti animali - che emanava dal cervello.

Dopo la morte di Grimaldi il problema dei suoni generati dai muscoli venne lar-

gamente ignorato per un secolo e mezzo. Esso fu ripreso dal fisico, chimico e medico inglese William Hyde Wollaston. Nel 1810, nella sua Croonian Lecture alla Royal Society di Londra, Wollaston comparò il suono di cui parlava Grimaldi al rombo lontano delle carrozze che si muovevano al mattino presto sulle strade selciate di Londra. Wollaston aveva realizzato il confronto in un modo eccezionalmente preciso al fine di determinare la frequenza del suono generato dai muscoli: aveva guidato la sua carrozza per le strade a varie velocità finché il rumore si era accordato con il rombo trasmesso attraverso i pollici. Quando la carrozza correva a otto miglia (quasi 13 chilometri) all'ora il suono che produceva assomigliava strettamente a quello derivato dai muscoli degli avambracci.

Il disegno regolare delle strade di Londra fu di notevole aiuto a Wollaston. I cubetti che erano serviti per la pavimentazione avevano tutti un lato di circa 15 centimetri. Otto miglia all'ora corrispondono a circa tre metri e mezzo al secondo, e così le ruote della carrozza venivano a contatto con poco più di 23 cubetti al secondo. Wollaston concluse, allora, che la frequenza del suono muscolare è di circa 23 hertz.

Con un semplice stetoscopio, consistente di un bastone di legno unito a un cuscinetto contro cui egli metteva l'orecchio, Wollaston fece due ulteriori osservazioni. Notò che tutti i muscoli del corpo emettono lo stesso rombo indipendentemente dalle loro dimensioni; inoltre scoprì che il suono non cambiava con la forza esercitata, a meno che questa non fosse molto grande. Con tale forza il suono emesso dal muscolo sembrava lievemente più acuto.

Dopo il 1810 i suoni generati dai muscoli vennero di nuovo ignorati per più di 150 anni, eccetto che in pubblicazioni scientifiche per occasionali riferimenti alla loro presenza come disturbo che interferiva con l'ascolto di altri suoni del corpo. Il mio interesse nei loro confronti si è destato mentre studiavo i difetti congeniti del cuore in ratti giovanissimi le cui madri erano state esposte a determinate sostanze chimiche durante la gestazione.

Per ascoltare i rumori prodotti dal cuore dei ratti risultava comodo usare uno stetoscopio elettronico.

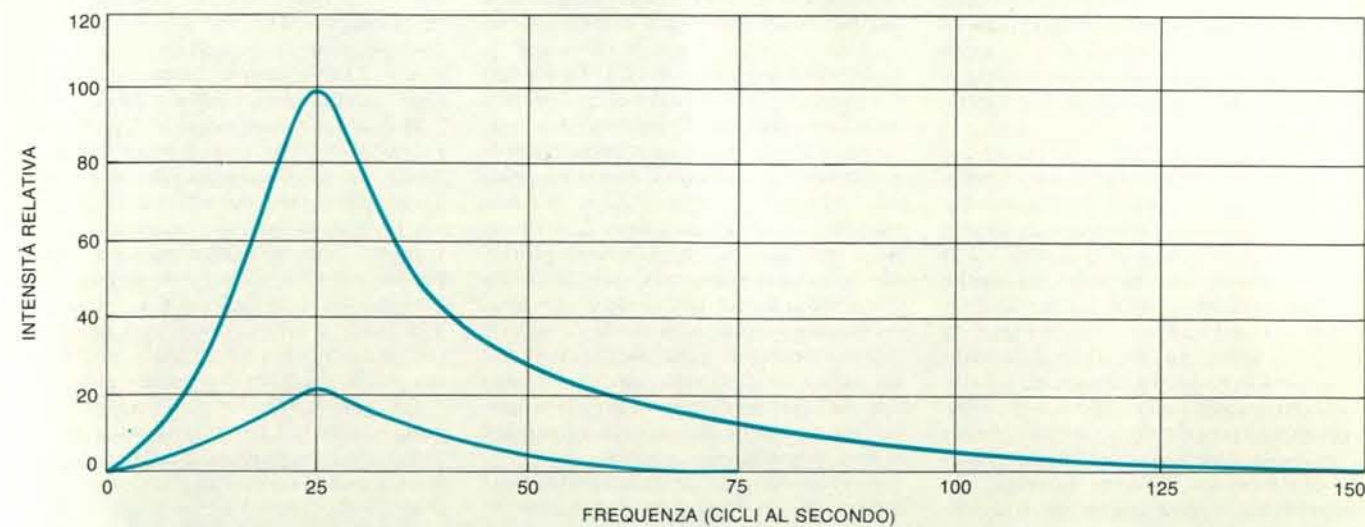
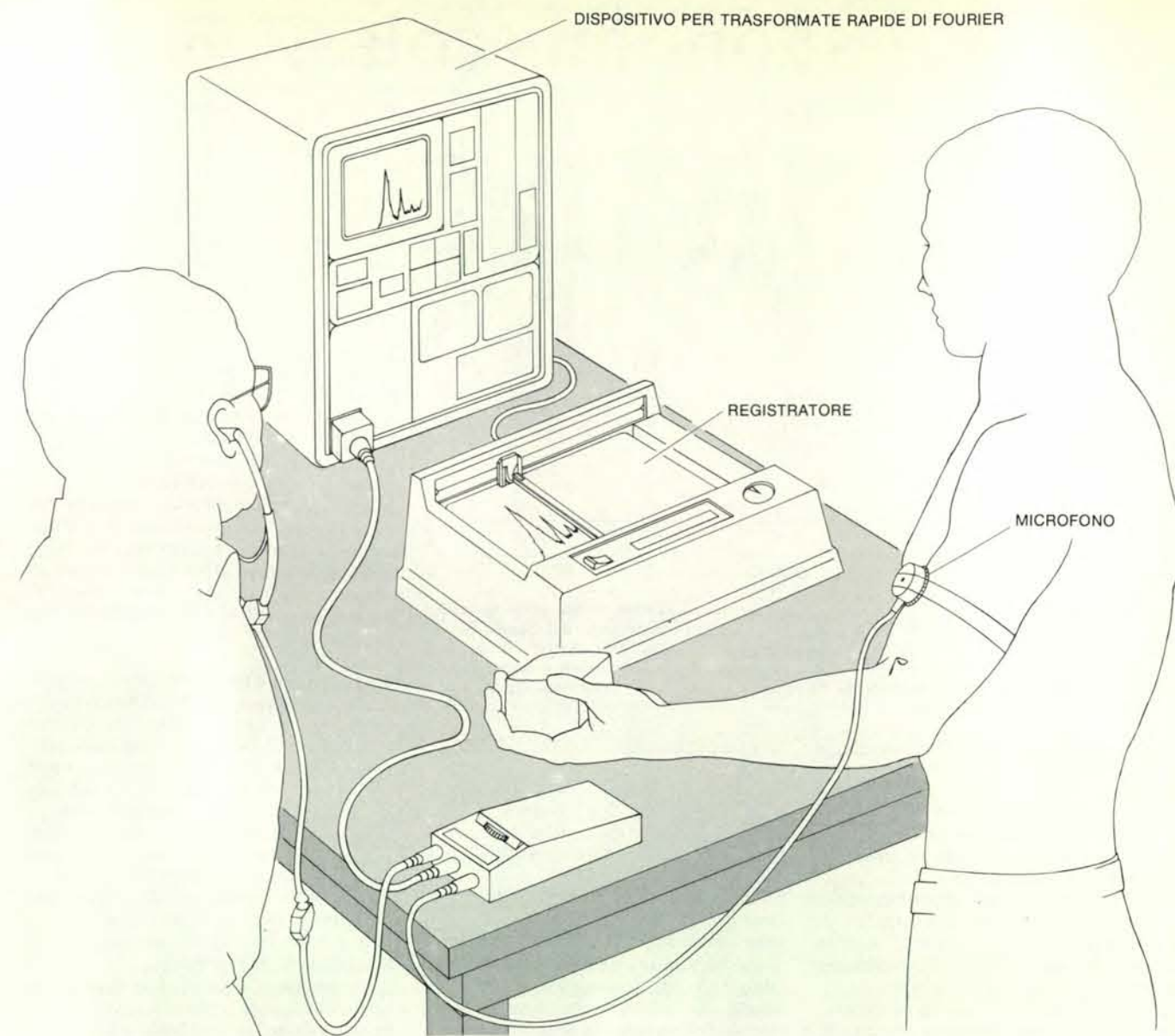
Lo stetoscopio elettronico amplifica i suoni a bassa frequenza. I ratti venivano trattati con leggeri sedativi per renderli più trattabili, ma le loro zampe di tanto in tanto si contraevano e, in tali occasioni, le contrazioni muscolari erano accompagnate da un forte rombo negli auricolari dello strumento.

In un primo tempo ho supposto che i suoni generati dai muscoli fossero già ampiamente conosciuti. Dopo aver studiato la letteratura scientifica e medica e dopo aver trovato che, al contrario, essi erano stati largamente trascurati, ho deciso di analizzarli con strumenti moderni. Ho assunto come collaboratore Joshua S. Jaffe, che stava allora compiendo un tirocinio presso la Mount Sinai School of Medicine di New York.

Non sorprende che i medici avessero prestato una scarsa attenzione ai suoni generati dai muscoli. Lo stetoscopio meccanico, lo strumento più largamente usato per ascoltare i suoni provenienti dal corpo, non serve per i suoni muscolari. Esso è di grande utilità per rilevare suoni di frequenza dai 75 ai 200 hertz, come i soffi cardiaci o i rantoli polmonari, mentre proprio per il modo in cui è costruito esso elimina la maggior parte dei suoni di frequenza inferiore a 50 hertz.

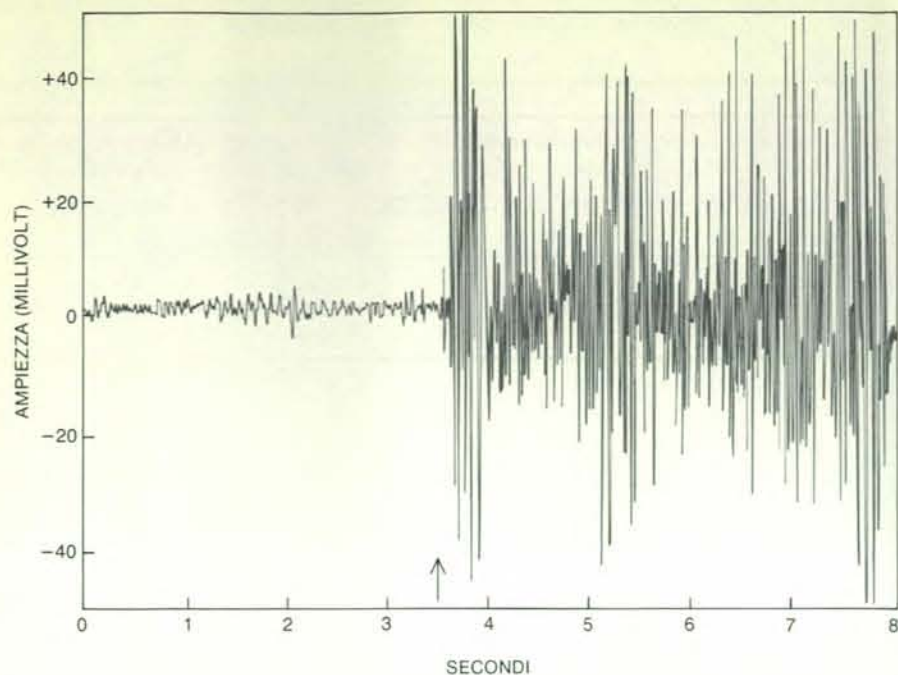
La piccola quantità di suoni di bassa frequenza che passa attraverso lo stetoscopio meccanico è, inoltre, quasi impercettibile perché l'orecchio umano non è sensibile ai suoni di bassa frequenza. Per esempio, l'uomo è circa 10 000 volte meno sensibile a suoni di 25 hertz rispetto a quelli di 250 hertz. La maggior parte degli audiologi considera 20 hertz il limite inferiore dell'udito nell'uomo; al di sotto di circa 30 hertz il suono perde la sua qualità tonale e può essere descritto dall'ascoltatore solo come un rombo profondo.

Lo stetoscopio elettronico è un eccellente mezzo per superare queste difficoltà. Il suo componente chiave è un microfono che funziona come un trasduttore che converte le onde di pressione, quali sono quelle che trasportano i suoni all'o-



L'autore ha utilizzato uno stetoscopio elettronico per rilevare i suoni di bassa frequenza generati dai muscoli durante la contrazione. La sezione superiore di questa figura riproduce l'attrezzatura per un tipico esperimento sui suoni muscolari. Il soggetto (a destra) tiene in mano un peso mentre il microfono dello stetoscopio elettronico è applicato al suo bicipite. L'osservatore (a sinistra) ascolta i suoni muscolari amplificati con gli auricolari dello stetoscopio. Il segnale elettrico provenien-

te dallo stetoscopio è elaborato tramite il programma del calcolatore noto come trasformata rapida di Fourier. Questa dà lo spettro del suono muscolare - densità di energia in funzione della frequenza - che viene registrato su carta. La sezione inferiore della figura mostra lo spettro di due suoni muscolari, uno generato tenendo in mano un peso di cinque chilogrammi (in alto) e l'altro un peso di due chilogrammi (in basso). Il suono ha un'intensità massima a 25 hertz (cicli al secondo).



Il segnale elettrico proveniente dallo stetoscopio elettronico cambia bruscamente quando i muscoli del braccio si contraggono. L'illustrazione riporta una registrazione del segnale ricevuto da un microfono applicato al muscolo flessore radiale del carpo, situato nell'avambraccio. All'inizio del tracciato (a sinistra) il braccio era rilassato. Quando la mano è stata chiusa a pugno e il muscolo si è contratto (freccia), l'ampiezza del segnale è aumentata bruscamente.

recchio, in segnali elettrici. Nella maggior parte degli stetoscopi elettronici i microfoni utilizzano un cristallo piezoelettrico, che può convertire impulsi di pressione direttamente in elettricità.

Un buon trasduttore elettronico può rispondere uniformemente a tutte le frequenze fino a circa 1000 hertz. Il segnale proveniente dal cristallo è modificato tramite circuiti allo stato solido per amplificare i suoni con le frequenze desiderate. Nella ricerca sui suoni generati dai muscoli, quelli compresi, tra 15 e 100 hertz sono stati amplificati in questo modo. Il segnale che ne risulta può essere ritrasformato in suono in un auricolare oppure può essere convertito in modo da dare un'immagine visiva su un oscilloscopio.

Lo stetoscopio elettronico ha chiaramente dei vantaggi rispetto a quello meccanico, ma presenta comunque anche uno svantaggio; amplifica, infatti, non solo i suoni generati dai muscoli, ma anche rumori di fondo prodotti dai condizionatori d'aria, dal traffico stradale e persino dai passi. Questi rumori di fondo a bassa frequenza non sono facilmente eliminabili. La frequenza delle onde sonore è inversamente proporzionale alla lunghezza d'onda e la dimensione del filtro necessario per bloccare il suono aumenta con quest'ultima. I suoni con un'alta frequenza, come la voce umana, il rumore di un aereo a reazione o la sirena della polizia, hanno una lunghezza d'onda corta e sono facilmente assorbiti dai materiali per costruzione. I suoni con una frequenza inferiore a 100 hertz sono, invece, molto più difficili da eliminare. Per esempio, i suoni

con una frequenza di 20 hertz hanno una lunghezza d'onda di 15 metri e, perché non vengano captati da uno stetoscopio elettronico, occorrerebbe un filtro delle dimensioni di un edificio.

Poiché non si può evitare che questi suoni estranei a bassa frequenza raggiungano il microfono, l'emissione dello stetoscopio include una componente casuale - il rumore - in aggiunta ai suoni prodotti dai muscoli. Quindi si è reso necessario un metodo per separare le informazioni utili dal rumore. Mi sono rivolto per questo al metodo matematico della autocorrelazione. L'autocorrelazione delle vibrazioni casuali è stata formulata da G. I. Taylor dell'Università di Cambridge per analizzare la turbolenza dei fluidi. I particolari matematici del metodo non hanno importanza in questa sede: in sostanza il metodo consiste nel confronto tra un segnale e lo stesso segnale dislocato nel tempo. L'informazione utile in questo segnale ha una forma regolare che si ripete nel tempo, come avviene per la forma delle onde di pressione del suono generato dai muscoli. I rumori di fondo, invece, non mostrano in genere tale regolarità. Ne risulta che, quando il segnale viene spostato e confrontato con se stesso, l'informazione utile appare rafforzata e la non informazione cancellata.

Il procedimento dell'autocorrelazione mostra che la frequenza dominante del suono muscolare è di 25 hertz. Essa mostra anche che il suono muscolare non consiste di una frequenza singola, ma include una gamma di frequenze. Lo spettro di un suono che include parecchie frequenze può essere definito come densità di energia in funzione della frequenza. Lo spettro ap-

rossimativo del suono prodotto dai muscoli raggiunge un massimo a 25 hertz più o meno 2,5 hertz. Tale suono è equivalente a un rumore casuale che passa attraverso un filtro che blocca i suoni di frequenza inferiore a 22,5 hertz o superiore a 27,5 hertz.

Nasce naturalmente l'interrogativo sul perché lo spettro del suono generato dai muscoli sia stato determinato tramite la procedura di autocorrelazione piuttosto che direttamente tramite i dati rilevati con lo stetoscopio elettronico. La ragione è che il calcolo dello spettro, a partire da dati vibrazionali come quelli che derivano dalle contrazioni muscolari, sarebbe stato proibitivo, fino a tempi recenti, come dispendio sia di tempo sia di denaro, anche con i calcolatori digitali più veloci.

Il calcolo di un tale spettro è stato invece molto semplificato oggi grazie a un programma per calcolatore, scritto da James W. Cooley dell'IBM Thomas J. Watson Research Center e da John W. Tukey dei Bell Laboratories e della Princeton University. L'algoritmo su cui è basato si chiama trasformata rapida di Fourier: con esso un milione di calcoli può essere ridotto a 5000, impiegando un tempo 200 volte minore.

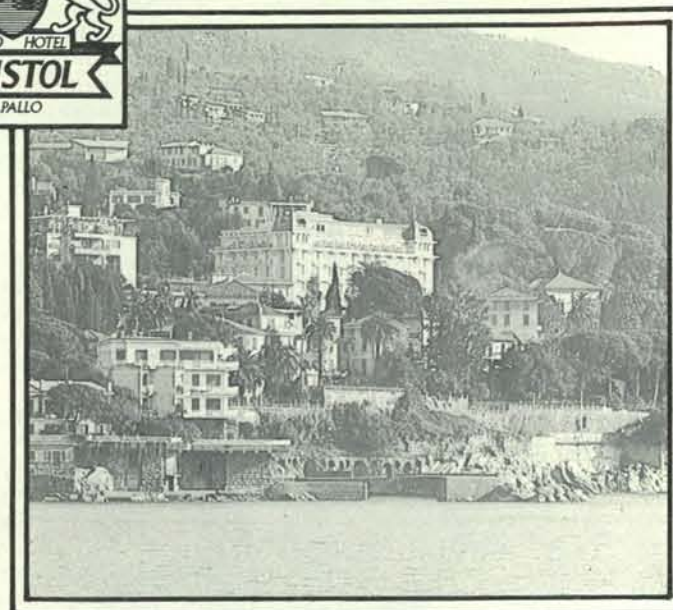
La trasformata rapida di Fourier ha avuto un effetto molto incisivo sulla ricerca in molti campi. La disponibilità del programma di Cooley e Tukey nello studio dei suoni muscolari significa che si può ottenere lo spettro esatto di tali suoni, senza doversi basare sullo spettro approssimativo dato dall'autocorrelazione. Nei nostri esperimenti un calcolatore capace di produrre trasformate rapide di Fourier è stato collegato allo stetoscopio elettronico: esso ha fornito lo spettro del suono muscolare quasi simultaneamente alla contrazione muscolare.

In un tipico esperimento il soggetto (generalmente Jaffe) tiene in mano, con il palmo rivolto verso l'alto, un peso di piombo. Il microfono può essere attaccato al bicipite. Quando il peso è tenuto fisso per mantenere una contrazione costante, l'ampiezza dei suoni muscolari è direttamente proporzionale a esso.

Il fatto che l'ampiezza del suono sia proporzionale al carico implica che la misura dei suoni muscolari può essere utilizzata per determinare l'intensità con cui sta lavorando un dato muscolo. Se un soggetto tiene un peso in mano, il suono proveniente dal bicipite è minore quando l'angolo tra avambraccio e braccio è di 115 gradi. I sollevatori di pesi sanno da tempo che questo è l'angolo a cui il braccio può sopportare il peso maggiore.

Quando il corpo è in posizione prona, il gastrocnemio, il più grande muscolo del polpaccio, non produce quasi nessun suono. Quando il corpo è in posizione eretta, il suono che deriva da esso invece è abbastanza forte; quando il peso del corpo è sostenuto dalle dita dei piedi, il suono è intenso. Quando Jaffe stava sulle punte dei piedi, l'intensità del suono proveniente dal gastrocnemio era uguale all'intensità del suono proveniente dall'avambraccio quando questo sosteneva un peso di

SRL



Torna a vivere a Rapallo il Grand Hotel Bristol

Il Grand Hotel Bristol, dopo anni di accurato lavoro dedicati al suo completo rinnovamento, è tornato.

Lo stile, il comfort, l'efficienza, che sempre hanno caratterizzato il Bristol, sono stati da guida alla realizzazione del nuovo Grand Hotel, situato al centro di uno scenario di incantevole bellezza, nella stupenda riviera ligure. Nulla è stato tralasciato per rendere più piacevole il soggiorno al Grand Hotel Bristol, tutto è stato appositamente ricreato per la Vostra meravigliosa villa sul mare.

Camere con bagno privato, suites, aria condizionata, frigo bar, TV, Rai, filo-diffusione. Ben 5 ristoranti, roof garden, bars, piano bar, solarium, piscina acqua mare riscaldata, sauna, massaggi e parrucchiere per signore e signori. Per gli uomini d'affari la possibilità di 6 sale meetings indipendenti dalla vita dell'Hotel, con capacità da 20 a 150 partecipanti, dotate di tutti i più moderni servizi ed attrezzature. Possibilità per: tennis, equitazione, golf, ormeggio privato; all'interno: post congressi, serate di gala, spettacoli e sfilate.

Grand Hotel Bristol SpA - Via Aurelia Orientale 369 - 16035 Rapallo
Tel. 0185/273313 - Tgr Bristol Rapallo - Tlx 270688 GHBRIS I

6,75 chilogrammi, il massimo peso che poteva tenere saldamente in mano.

Quando abbiamo misurato il suono generato dai muscoli di una ballerina professionista, abbiamo trovato che non riusciva a tenere in mano un peso pari a quello di Jaffe. Quando, invece, reggeva il peso del proprio corpo sulle dita dei piedi, il suono derivato dal gastrocnemio era molto più intenso di quello generato da Jaffe quando stava sulle punte dei piedi. L'allenamento della ballerina a stare sulle punte l'aveva chiaramente indotta ad

adozzare una posizione differente da quella presa da Jaffe. Questo suggerisce che la misurazione del suono muscolare può rivelare quali muscoli sono attivi in un particolare esercizio fisico.

Il soleo è un muscolo della gamba situato vicino al gastrocnemio. L'ho studiato con un microfono in miniatura la cui membrana aveva un diametro di solo mezzo centimetro. Quando il corpo era in una posizione eretta normale, il suono proveniente dal soleo era quasi 10 volte più intenso di quello proveniente dal ga-

strocnemio. L'intensità del suono generato dal soleo è correlata alla fisiologia e anche al ruolo che questo muscolo ha nella deambulazione. Il soleo è responsabile del mantenimento dell'angolo tra gamba e piede e quindi dà un maggior contributo alla posizione eretta nell'uomo.

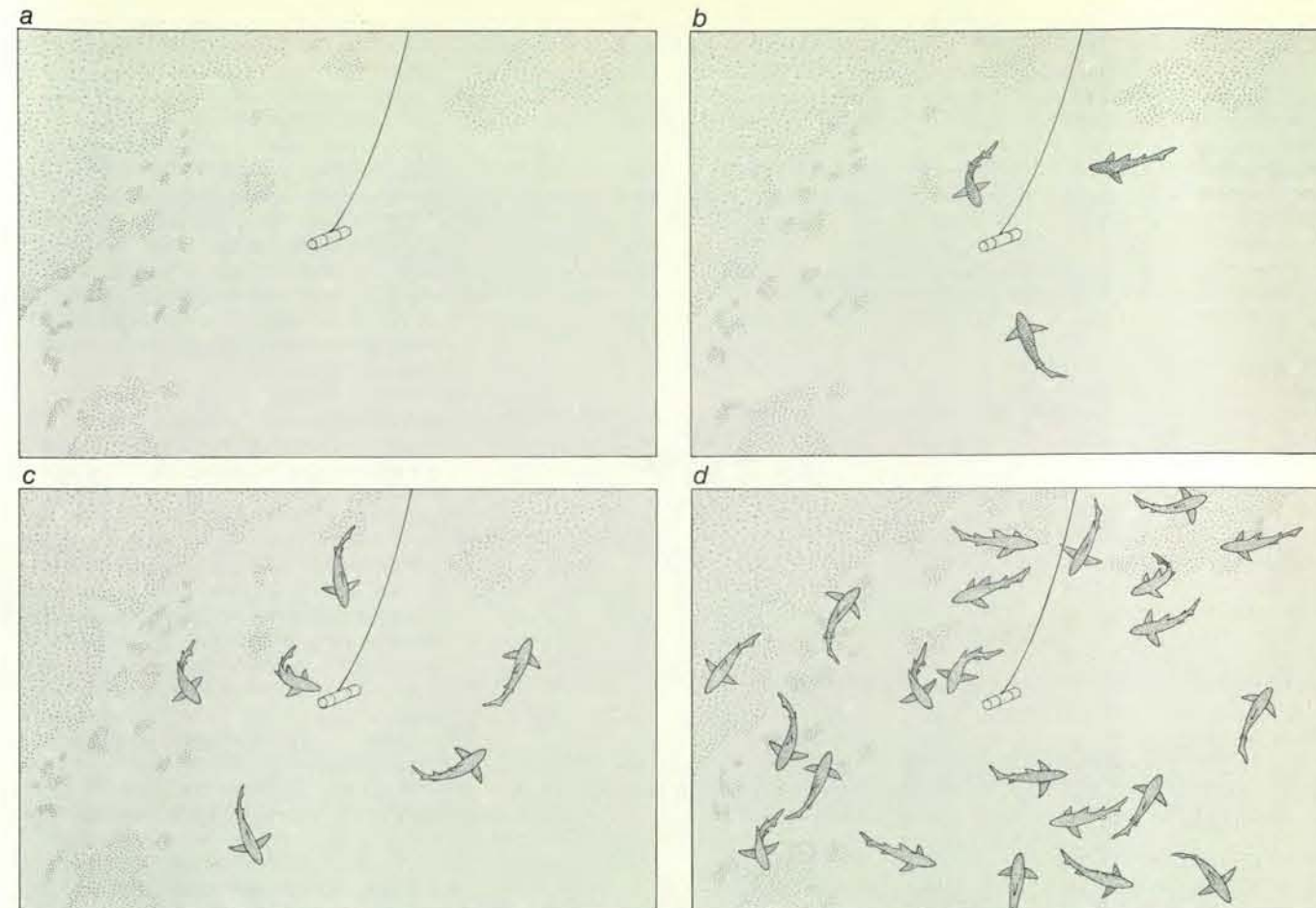
Nella stazione eretta, nella marcia o nella corsa, il soleo si contrae molte volte e per un lungo periodo senza fatica. Gli altri muscoli scheletrici si affaticano più facilmente ed è evidente che la velocità di trasformazione dell'energia nel soleo è considerevolmente inferiore a quella in altri muscoli. La resistenza del soleo deriva dal fatto che esso è abbondantemente provvisto di mitocondri, gli organelli cellulari trasduttori di energia. Nei mitocondri gli enzimi ossidativi producono adenosintrifosfato (ATP), che è la fonte universale di energia nella cellula.

Le fibre muscolari che contengono molti mitocondri sono fibre a contrazione lenta; le fibre con un minor numero di mitocondri possono, invece, contrarsi rapidamente. È ovvio che le prime sono importanti nelle attività che richiedono resistenza e le seconde in quelle che richiedono velocità. Il soleo, che ha una prevalenza di fibre a contrazione lenta, è più sviluppato nei maratoneti di quanto non lo sia negli scattisti, perché uno scatto di breve durata non richiede la prolungata contrazione del soleo. Può darsi che la misurazione quotidiana dell'intensità del suono generato dal soleo comparata con l'intensità del suono generato, per esempio, dal gastrocnemio serva come metodo per valutare i progressi nell'allenamento di un maratoneta.

I suoni muscolari potrebbero essere di notevole interesse per gli studiosi del comportamento animale. Sott'acqua i suoni in genere possono essere uditi senza difficoltà. Se si tiene la testa sott'acqua e si stringe la mano a pugno in prossimità dell'orecchio, si riesce a sentire il suono generato dai muscoli. La ragione per cui il suono può essere udito meglio attraverso l'acqua che attraverso l'aria è che questa è 1000 volte meno densa dei tessuti umani (e proprio per questa differenza il suono viene trasmesso male da essa all'altro mezzo) mentre l'acqua è molto più densa dell'aria.

Uno dei pochi pesci che può essere udito al di sopra della superficie dell'acqua è il gourami brontolone (*Trichopsis vittatus*). Parecchie altre specie di pesci generano suoni facendo vibrare la loro vescica natatoria, l'organo gonfiabile che li aiuta a mantenere il galleggiamento. I gourami emettono suoni solo quando si muovono in modo violento e rapido durante l'accoppiamento. Dato che questi suoni non derivano apparentemente né dalla vescica natatoria né dalla turbolenza dell'acqua, può darsi che la loro fonte sia la contrazione muscolare.

Alcuni squali sono attratti dai suoni di bassa frequenza. Nel 1963 Donald R. Nelson e Samuel H. Gruber della Scripps Institution of Oceanography hanno dimostrato che lo squalo *Negaprion brevirostris* si muove verso la sorgente di un suono se questo ha una frequenza tra 20 e



Gli squali delle scogliere del Pacifico sono attratti da registrazioni di suoni di bassa frequenza riprodotte da un altoparlante sott'acqua. I quattro quadri che compongono l'illustrazione sono ricavati da fotogrammi di un film girato da Donald R. Nelson e da Richard H. Johnson della California State University a Long Beach sull'atollo Eniwetok

che fa parte delle isole Marshall. L'altoparlante era stato posto a 15 metri sotto la superficie. Quando si è cominciato a trasmettere la registrazione, non c'erano squali nella zona (a). Entro 30 secondi, si sono raccolti attorno all'altoparlante alcuni squali (b, c). Dopo due minuti oltre una ventina di squali giravano attorno all'altoparlante (d).

50 hertz. Se la frequenza è al di fuori di questo intervallo, non ne è attratto.

Nelson, che attualmente lavora alla California State University a Long Beach, ha proseguito nella dimostrazione che i suoni intermittenti a bassa frequenza provenienti da un altoparlante posto sott'acqua producono una notevole attrazione per gli squali (*Hexanchus griseus*) che popolano le scogliere del Pacifico. Questi animali si muovono verso l'altoparlante quando esso emette la registrazione di un pesce preso all'amo o quella di un pesce in bocca a uno squalo ed è verosimile che siano attratti dagli intensi suoni muscolari emessi dal pesce che cerca di liberarsi.

I suoni muscolari possono servire come via di comunicazione persino tra embrioni. Margaret A. Vince, che lavora in Inghilterra all'Agricultural Research Council, ha rilevato suoni a bassa frequenza emessi da uova di quaglie giapponesi. Utilizzando un trasduttore ideato specificamente per l'ascolto dei suoni provenienti dalle uova, la Vince ha notato che questi cominciano circa quattro giorni prima che le uova si schiudano. Per osservare l'embrione, ha praticato un'apertura nel guscio e l'ha coperta con una plastica traspa-

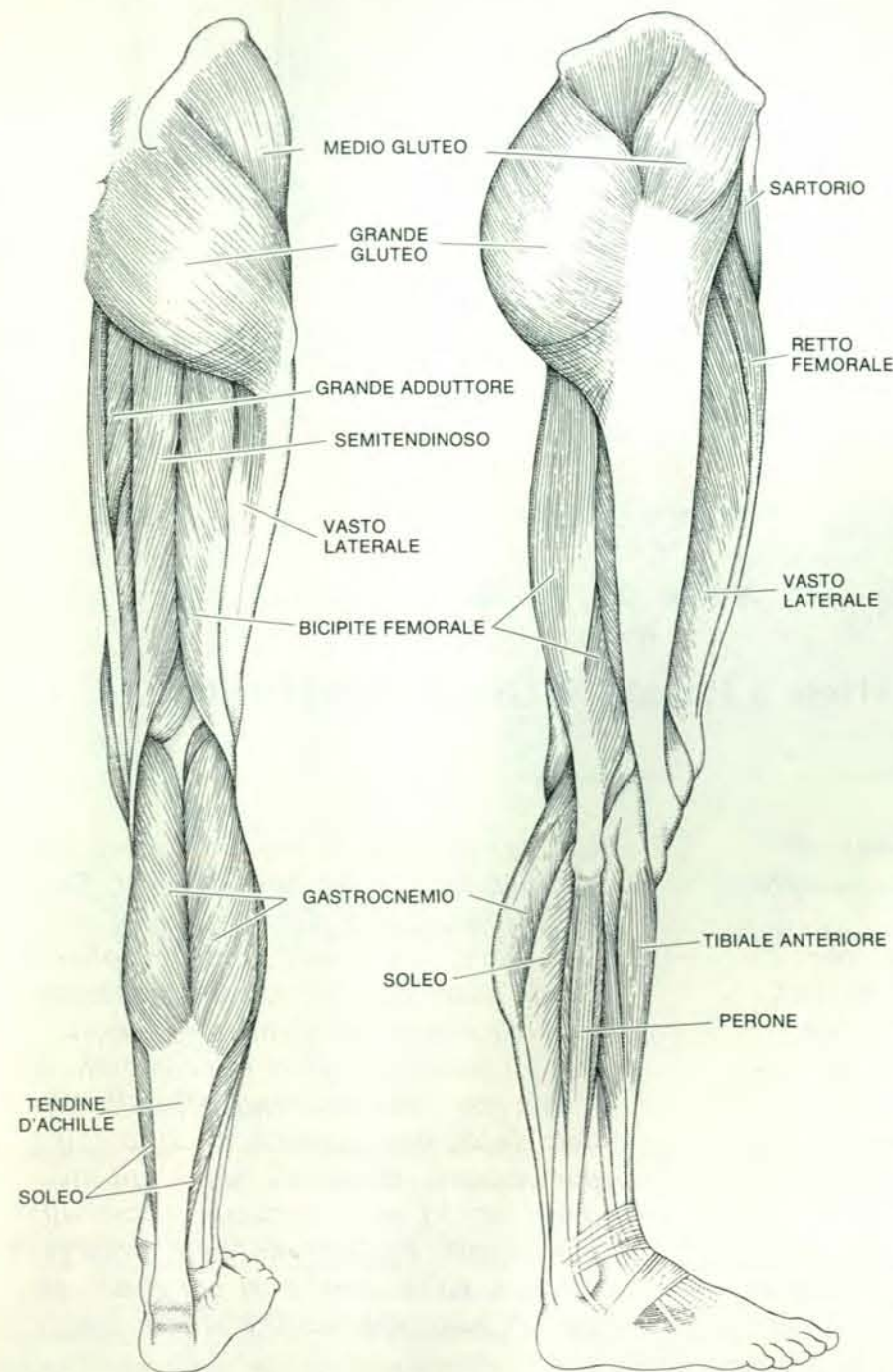
rente; attraverso questa finestra ha poi osservato i movimenti muscolari durante l'emissione di rombi. Altri fatti suggeriscono che i suoni muscolari provenienti da un embrione possano stimolare altri embrioni a cominciare i loro sforzi per uscire dal guscio. Secondo un'ipotesi che la Vince ha formulato, i suoni muscolari spiegherebbero il fatto che le uova a contatto con un altro uovo generalmente si schiudono prima delle uova isolate.

Qual è la fonte dei suoni muscolari? Sembra che il rombo origini dall'attività delle singole fibre muscolari, in particolare di quelle a rapida contrazione. A scopo sperimentale fibre isolate a rapida contrazione possono essere montate a ponte su supporti meccanici. Quando vengono stirate a una determinata lunghezza e quindi stimulate elettricamente, ogni singola fibra esercita una forza sui supporti. Essa non può accorciarsi, ma misurando la forza esercitata sui supporti si può esaminare la risposta. In tali condizioni le fibre a rapida contrazione rispondono in circa 40 millisecondi, cioè in un venticinquesimo di secondo. È questo grosso modo il tempo richiesto per il *turn-*

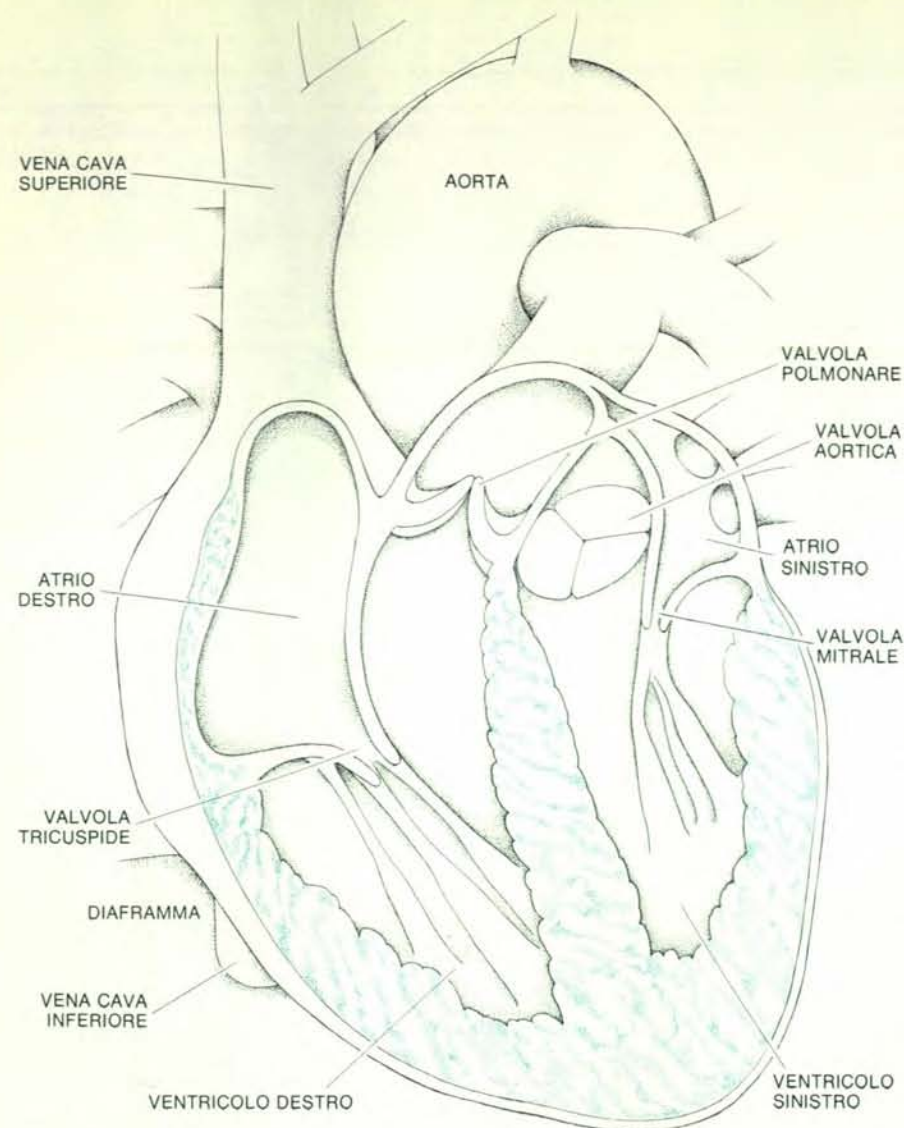
over - cioè l'utilizzazione e la rigenerazione - dell'ATP. Si ritiene che il turnover dell'ATP controlli le reazioni chimiche che sono alla base della contrazione muscolare. Va anche sottolineato che un venticinquesimo di secondo è il periodo necessario per un ciclo completo di un suono con frequenza di 25 hertz.

Molti ricercatori hanno dimostrato che le fibre muscolari vibrano quando il carico sul muscolo viene leggermente cambiato. Clara Franzini-Armstrong, Andrew F. Huxley e F. J. Julian dello University College di Londra, lavorando con fibre isolate da muscolo di rana sotto tensione e con un carico che poteva essere variato con considerevole precisione, hanno dimostrato che, aumentando o diminuendo il carico, le fibre vibrano lungo il loro asse con una frequenza di 25 hertz. La variazione in lunghezza è soltanto dell'1 per cento circa rispetto alla lunghezza normale della fibra e quindi può essere rilevata solo con strumenti sensibili.

G. Goldspink, R. E. Larson e R. E. Davies dell'Università della Pennsylvania hanno lavorato con il muscolo latissimo anteriore del dorso di pulcini dell'età di 24 giorni. Il latissimo anteriore del dorso è un



L'anatomia dell'arto inferiore dell'uomo è riportata in queste proiezioni posteriore e laterale. I due muscoli principali del polpaccio sono il soleo e il gastrocnemio. Nella normale stazione eretta il suono dal soleo è quasi 10 volte più intenso del suono emesso dal gastrocnemio. Il soleo è il muscolo responsabile del mantenimento dell'angolo tra gamba e piede. Esso è costituito in prevalenza da fibre a lenta contrazione, che possono contrarsi molte volte senza fatica.



Il cuore umano genera parecchi suoni in ogni ciclo completo della sua funzione di pompa. Il primo tono cardiaco segna l'inizio della fase sistolica (di contrazione) del battito cardiaco: è stato ipotizzato che sia causato dalla chiusura della valvola mitrale o dal brusco cambiamento di pressione all'interno del ventricolo sinistro, la più potente delle due cavità cardiache con funzione di pompa. Potrebbe però anche essere il suono emesso dal ventricolo sinistro mentre si contrae.

muscolo a contrazione rapida, come tutti i muscoli della carne bianca dei polli. (La carne scura è costituita da muscoli a contrazione lenta). Con una tecnica completamente diversa da quella dei ricercatori di Londra, questi studiosi hanno trovato una vibrazione con una frequenza di circa 25 hertz nelle fibre del latissimo anteriore del dorso quando cambia la tensione.

Gli esperimenti sopra descritti suggeriscono che sono le vibrazioni delle fibre a rapida contrazione a generare i suoni muscolari. Vi è però un altro tipo di vibrazione dei muscoli che si potrebbe ritenere responsabile della produzione dei suoni: si tratta del tremore fisiologico, che può essere osservato, ad esempio puntando il dito su una parola di questa pagina. Come risultato di complessi fenomeni di *feedback* neurologico, il dito mostrerà un'oscillazione di 10 hertz appena visibile.

Jaffe e io abbiamo deciso di scoprire se il tremore fisiologico contribuisce o no al

suono muscolare. Sono state eseguite misurazioni del suono proveniente da un braccio che sostiene un peso in una vasca di acqua gelata e da un braccio in cui il flusso sanguigno era stato fortemente ridotto tramite un bracciale gonfiabile. Tali accorgimenti eliminano o diminuiscono di molto il tremore, ma non hanno effetto sul suono muscolare.

Si deve quindi escludere qualsiasi partecipazione del tremore fisiologico ed è logico concludere che il suono di 25 hertz proviene soltanto dalle fibre a rapida contrazione. Le numerose fibre nel fascio che costituisce un muscolo possono essere considerate come un insieme di oscillatori. Tutti gli oscillatori hanno la stessa frequenza ma non devono necessariamente essere in fase. Così, in un particolare muscolo, alcune fibre potrebbero essere in fase le une con le altre, altre potrebbero essere in opposizione di fase rispetto a esse e altre ancora

potrebbero situarsi in una posizione intermedia. La tensione del muscolo provoca l'accorciamento del fascio e il suo aumento in larghezza, il che dà il via alle vibrazioni delle singole fibre. A sua volta questo movimento dà origine alle onde di pressione che causano un suono di 25 hertz.

Quando il muscolo si contrae, il rombo continuo non sempre comincia immediatamente. Se il muscolo è sottoposto solo a un leggero esercizio, a volte si ascoltano dei *click*, o rapide emissioni di suono. Se la forza aumenta, l'intervallo tra i *click* diminuisce fino a che essi sfumano in un basso rombo.

I *click* provengono dall'attività delle subunità muscolari note come unità motorie. Ogni unità motoria è costituita da un gruppo di fibre attivate da un singolo nervo. Le sue dimensioni e la sua modalità d'azione possono variare ampiamente in base alla funzione del muscolo.

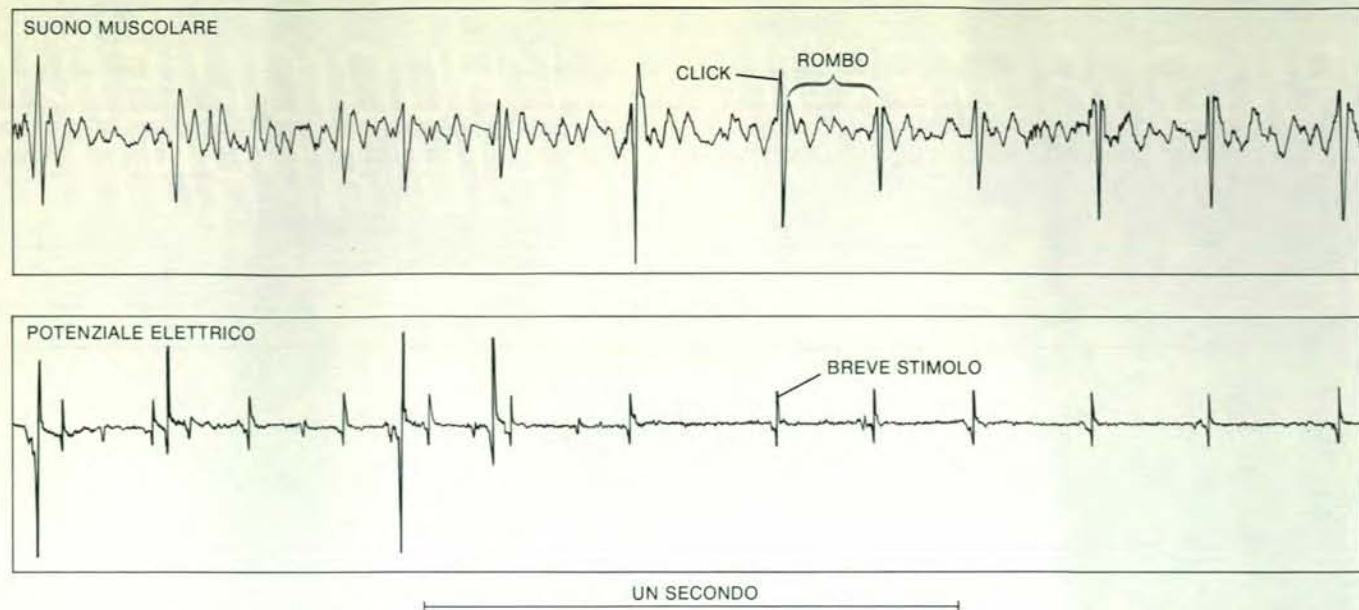
Nel gastrocnemio ogni unità motoria è costituita da circa 2000 fibre. Nella marcia o nella corsa le unità motorie non sono tutte attive simultaneamente e la forza esercitata dal muscolo aumenta lentamente a mano a mano che vengono reclutate nuove unità motorie. Al tempo stesso, le unità in cui subentra la fatica cessano di funzionare per poter recuperare. Quest'alternanza continua per tutta la durata della contrazione muscolare.

Nell'orbicolare dell'occhio, il muscolo della palpebra, ogni unità motoria è costituita soltanto da una ventina di fibre. Nell'ammicciamento, la palpebra risponde in circa 50 millisecondi, cioè in un ventesimo di secondo. La risposta è del tipo tutto o nulla, con tutte le unità motorie che si eccitano contemporaneamente. Per la velocità e la precisione dell'operazione le unità motorie devono essere piccole.

I *click* che precedono il rombo sono stati notati per la prima volta nell'orbicolare dell'occhio; l'osservazione è stata descritta nel 1948 da G. F. Gordon e A. H. S. Holbourn dell'Università di Oxford, che hanno usato un piccolo microfono per studiare il movimento della palpebra. L'attività delle unità motorie nella palpebra è così definita che ogni unità motoria può far sentire la sua attività indipendentemente dalle altre. Se si appoggia leggermente il dito medio alla palpebra, si sente ogni tanto una pulsazione, che è la contrazione di un'unità motoria. È questa pulsazione che produce il *click* (o schiocco) che si trasmette alla colonna d'aria del microfono.

Delbert D. Thiessen dell'Università del Texas ad Austin ha ipotizzato che i gerbilli comunicano tramite *click* prodotti dalle unità motorie delle palpebre. Se alcuni esemplari vengono tenuti in gabbie separate, cominciano spesso a ripulirsi simultaneamente. La pulizia è preceduta da un rapido sbattere delle palpebre. Vari meccanismi potenziali di segnalazione, incluso l'odore, sono stati esclusi. Sembra, invece, che la segnalazione possa essere effettuata da brusche emissioni di suoni dall'orbicolare dell'occhio.

La connessione tra attivazione dell'unità motoria e *click* è stata confermata dall'esame elettromiografico. In questo pro-



Da un piccolo stimolo nervoso deriva una brusca emissione di suono dal muscolo, un click. Il tracciato in alto è la registrazione di un suono proveniente dal vasto mediale, un grande muscolo della coscia. Il tracciato in basso mostra il potenziale elettrico dello stesso muscolo. Al soggetto è stato chiesto di pensare alla contrazione dei muscoli della co-

scia senza realmente muoverli. Ne è risultato una serie di piccoli, brevi impulsi nervosi che appaiono come picchi sul tracciato in basso, ognuno accompagnato da un click sul tracciato in alto. Il click è il suono prodotto dalle fibre di una unità motoria, il gruppo di fibre attivate da un singolo nervo. Tra i click il tracciato in alto mostra il rombo di 25 hertz.

cedimento piccoli elettrodi a forma di ago vengono inseriti in un muscolo vicino a una data unità motoria; per il loro tramite si registra il potenziale d'azione che accompagna la contrazione delle fibre. Il potenziale d'azione è un breve incremento del potenziale elettrico di circa un centesimo di volt.

Gordon e Holbourn hanno osservato che nell'elettromiogramma della palpebra ci sono picchi rapidi che coincidono con i click. Essi non sono riusciti a dimostrare, invece, che anche i muscoli degli arti emettono click. Vi sono riusciti però, di recente, D. T. Barry e S. R. Geiringer dell'Università del Michigan i quali hanno registrato click emessi dai muscoli delle gambe; più precisamente hanno rilevato click generati da un muscolo della coscia in un soggetto in stato di relativa quiete. Come nelle osservazioni di Gordon e Holbourn i click sono comparsi contemporaneamente ai picchi dell'elettromiogramma.

Lo studio dei suoni provenienti dalla muscolatura cardiaca potrebbe essere di grande utilità pratica. L'azione di pompa del cuore dà origine a parecchi suoni. I due più importanti sono detti primo tono e secondo tono. Il primo tono segna l'inizio della fase sistolica (di contrazione) del battito cardiaco; il secondo l'inizio della fase diastolica (di rilassamento). Ho rilevato che lo spettro del primo tono ha un massimo attorno ai 22 hertz e che il secondo tono ha un'altezza nel complesso leggermente maggiore.

L'origine dei toni cardiaci è oggetto di controversie, ma le due interpretazioni più ampiamente accettate del primo tono sono l'ipotesi dello schiocco valvolare e quella del colpo d'ariete. La prima è stata proposta nel 1830 da Pierre Rouanet, un

medico di New Orleans. Egli ha attribuito il primo tono alle ripercussioni generate dalla brusca chiusura della valvola mitrale, situata tra atrio sinistro e il ventricolo sinistro. Studi ecocardiografici hanno, invece, dimostrato che il primo tono cardiaco comincia in realtà appena prima della chiusura della valvola.

Secondo l'ipotesi del colpo d'ariete, proposta da Robert F. Rushmer dell'Università di Washington, il primo tono cardiaco risulta da un meccanismo simile a quello che dà origine a un forte scoppio in un condotto di un sistema idraulico costruito male, quando viene chiuso troppo velocemente un rubinetto. (In un sistema costruito bene c'è, invece, una camera d'aria che previene il colpo d'ariete.) Quando la valvola mitrale si chiude, la pressione del sangue sale rapidamente. Così il ventricolo sinistro si distende, provocando una brusca ripercussione, che potrebbe generare il primo tono.

Nelle ipotesi suddette, il muscolo cardiaco in sé è relegato a un ruolo minore. Tuttavia, il tono cardiaco è più intenso se la forza di contrazione è maggiore, il che fa pensare che il primo tono cardiaco sia dovuto alla contrazione del muscolo cardiaco senza che vi contribuisca molto la dinamica della valvola mitrale o del ventricolo sinistro.

Quest'ipotesi trova conferma in esperimenti condotti con rane. Ho posto un microfono in miniatura a contatto diretto con il ventricolo di un cuore pulsante di rana: nonostante la grande differenza di massa tra il cuore della rana e il cuore dell'uomo, i suoni muscolari provenienti dai due sono molto simili. L'ipotesi del colpo d'ariete prevederebbe, invece, una frequenza molto più alta per quelli prodotti dal muscolo cardiaco della rana, perché, quando si distende un ventricolo

di piccole dimensioni, la ripercussione dovrebbe generare un suono di altezza maggiore rispetto a quello generato da un ventricolo di grandi dimensioni. L'ipotesi del colpo d'ariete è indebolita ulteriormente dall'osservazione che, quando viene ostacolato l'afflusso di sangue al cuore, i toni cardiaci non ne sono influenzati. Sembrerebbe quindi che la principale sorgente del primo tono cardiaco sia la contrazione muscolare.

Poiché più di un terzo della mortalità totale negli Stati Uniti è dovuta a degenerazione del muscolo cardiaco, i suoni che esso genera potrebbero diventare un prezioso strumento diagnostico. L'analisi dello spettro particolareggiato del suono proveniente dal muscolo potrebbe rivelare lo stato del tessuto e quindi diagnosticare un infarto miocardico: morte del tessuto cardiaco a causa di un'interruzione dell'apporto ematico. Persino con lo stetoscopio meccanico il primo tono cardiaco ha un'altezza anormalmente bassa in soggetti con infarto avanzato; per ottenere una registrazione più oggettiva si dovrebbe applicare a esso la trasformata rapida di Fourier.

Oltre a queste possibilità cliniche, lo studio dei suoni muscolari potrebbe dare un contributo alla comprensione dei meccanismi che sono alla base della contrazione muscolare. Per esempio, F. V. Brozovich e Gerald H. Pollack dell'Università di Washington hanno scoperto che un muscolo di rana produce una sequenza di brusche emissioni sonore, il che suggerisce che, quando un muscolo si contrae, lo fa a piccole fasi successive e distinte. Se ciò è vero, è in disaccordo con l'immagine che generalmente si dà della contrazione muscolare come processo continuo che opera a livello molecolare.

Florence Nightingale

Salvò la vita di migliaia di soldati durante la guerra di Crimea, fu uno dei fondatori dell'assistenza medica moderna e inoltre fu tra i primi a servirsi di statistiche sociali e a rappresentarle graficamente

di I. Bernard Cohen

Florence Nightingale è ricordata oggi come uno dei precursori nel campo dell'assistenza infermieristica e come una riformatrice degli ospedali. Essa vide però la sua missione nella prospettiva più ampia di servire l'umanità attraverso la prevenzione di malattie e decessi evitabili. Per gran parte della sua lunga vita (1820-1910) perseguì questa missione con un'inflexibile risolutezza, che conferì a ogni sua azione una singolare coerenza. I suoi massimi contributi vanno individuati senza dubbio negli sforzi da lei compiuti per riformare il sistema dell'assistenza sanitaria militare britannica e per fare dell'attività infermieristica, attraverso l'istituzione di programmi di addestramento e la definizione di sicure norme professionali, una professione rispettata. Gran parte di ciò che oggi ci appare ovvio nella moderna assistenza sanitaria può essere ricondotto a battaglie campali combattute da Florence Nightingale nell'Ottocento. Molto meno noto, essendo stato trascurato dai suoi biografi, è il suo uso ugualmente innovatore, in quelle battaglie, delle nuove tecniche perfezionate di analisi statistica.

La Nightingale imparò per esperienza diretta nella sua qualità di capoinfermiera durante la guerra di Crimea (1854-1856), che il miglioramento delle condizioni sanitarie in ospedali militari e in caserme poteva ridurre drasticamente l'indice di mortalità e salvare migliaia di vite. La sua battaglia fu rivolta a convincere gli uomini scettici che detenevano il potere. In un periodo in cui la raccolta e l'analisi di statistiche sociali era ancora insolita, la Nightingale riconobbe che dati attendibili sull'incidenza dei decessi evitabili fra i militari conferivano una grande forza alle argomentazioni a favore di riforme. Così, oltre a proporre la causa della riforma medica in sé, essa contribuì a introdurre quella nozione davvero rivoluzionaria che i fenomeni sociali potevano essere misurati obiettivamente e sottoposti ad analisi matematica.

I risultati ottenuti dalla Nightingale sono tanto più sorprendenti quando vengono valutati di contro allo sfondo delle

restrizioni sociali di cui le donne erano oggetto nell'Inghilterra vittoriana. Suo padre, William Edward Nightingale, era un proprietario terriero estremamente facoltoso e la famiglia era introdotta negli ambienti più elevati della società inglese. A quei tempi, le donne appartenenti alla classe sociale dei Nightingale non frequentavano le università e non intraprendevano carriere professionali; il loro scopo nella vita era sposarsi e far figli. La Nightingale fu fortunata: suo padre pensava che le donne dovessero ricevere un'istruzione, e le insegnò personalmente italiano, latino, greco, filosofia, storia e - cosa del tutto eccezionale per le donne del tempo - scrittura e matematica. Quando, poco dopo i vent'anni, Florence esprime un interesse per il lavoro di infermiera, il padre prese la cosa abbastanza sul serio da interpellare dei medici circa l'opportunità di una tale carriera.

Se intraprendere una carriera era un passo rivoluzionario per una donna della classe sociale della Nightingale, la decisione di fare l'infermiera parve inopportuna anche in una famiglia illuminata come la sua. Non era tanto «la parte fisicamente rivoltante di un ospedale» a colpire William Nightingale quanto quelle che sembravano prove schiaccianti delle abitudini dissolute delle infermiere. A quei tempi le infermiere non avevano una preparazione specifica; erano quasi sempre donne rozze e ignoranti, dedite alla promiscuità e all'ubriachezza. La stessa Nightingale raccontò in seguito a suo padre che la capoinfermiera di un ospedale di Londra le aveva detto di «non aver mai conosciuto un'infermiera che non fosse ubriaca» e che la maggior parte delle infermiere teneva una «condotta immorale» con i pazienti nelle corsie. Non sorprende quindi che i genitori di Florence sperassero che la figlia rinunciasse alla sua insolita ambizione, si sposasse e si acquietasse.

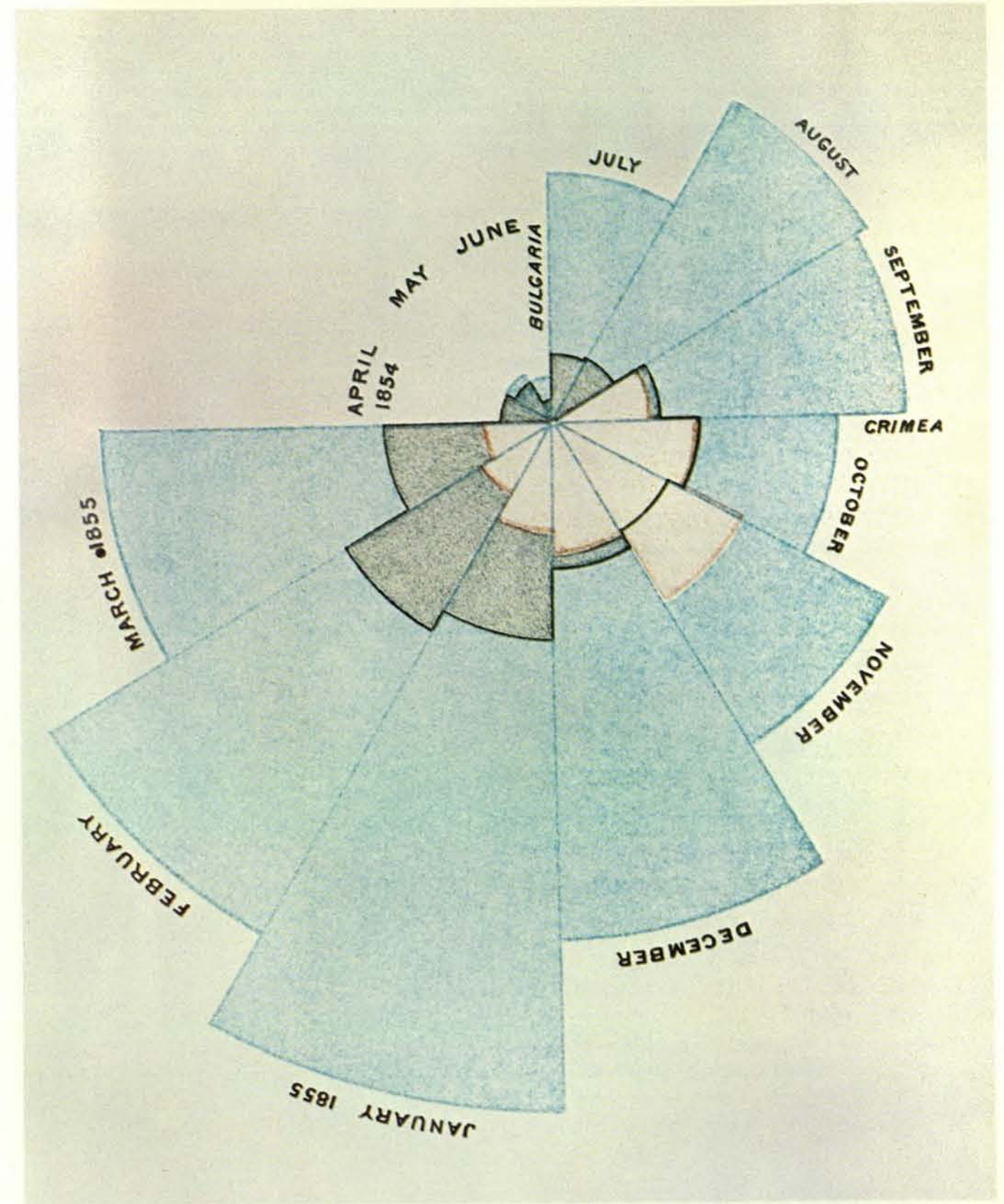
Florence Nightingale era sotto ogni aspetto una giovane attraente e non era certo per mancanza di occasioni che rifiutava il matrimonio. Una volta, in ef-

fetti, fu tentata di accettare un corteggiatore, ma dopo un lungo fidanzamento concluse con riluttanza che avrebbe tradito la sua natura «morale» e «attiva» se avesse «trascorso una vita con lui occupandosi di intrattenimenti e di problemi domestici». Il matrimonio convenzionale, scrisse nel suo diario, significava «essere costretta a prolungare e a portare all'esasperazione la vita che stava conducendo», prospettiva che le sembrava «simile al suicidio». Dio, decise, aveva preso in considerazione per lei un destino diverso. Essa era una di quelle donne per le quali Dio «aveva stabilito chiaramente qualcosa di unico».

Quando i suoi genitori le proibirono di fare l'infermiera, Florence si rivolse per conforto alla religione, che sarebbe rimasta un elemento guida nella sua vita. I suoi sentimenti religiosi ruotavano però attorno alla convinzione che il modo migliore per servire Dio fosse quello di dedicarsi al servizio dell'umanità. Così, nella sua difficile giovinezza fino all'età di trent'anni, non rinunciò mai all'ambizione di intraprendere una carriera; lesse voracemente libri di medicina e sull'assistenza sanitaria, dedicò qualche tempo alla visita degli ospedali a Londra e lavorò in forma privata, assistendo i bambini degli *slums*, che amava definire i suoi «ladruncoli di Westminster». Ciò nonostante, si sentiva frustrata.

Finalmente, nel 1851, la Nightingale riuscì ad allontanarsi da casa; trascorse tre mesi vicino a Düsseldorf, in Germania, presso un ospedale e un orfanotrofio gestito da un ordine protestante di «diaconesse». In seguito, malgrado il dissenso della sua famiglia, fece il proprio tirocinio presso un altro ospedale, diretto dalle Sorelle della Misericordia, a St. Germain, vicino a Parigi.

Di ritorno a Londra nel 1853, la Nightingale ottenne però presto il suo primo «impiego» (non retribuito) come sovrintendente di un «istituto per signore malate». Il suo compito consisteva nel sorvegliare le infermiere, nel vigilare sul funzionamento delle apparecchiature e nel garantire la purezza dei medicinali.



I diagrammi ad aree polari vennero inventati da Florence Nightingale per evidenziare il numero di decessi evitabili avvenuti negli ospedali militari britannici durante la guerra di Crimea (1854-1856). La Nightingale chiamò tali diagrammi *coxcombs* (berretti da giullare) per i loro colori. L'area di ogni cuneo in colore, misurata a partire dal centro, è proporzionale al dato statistico rappresentato. I cunei azzurri rappresentano i decessi da malattie «zootiche prevenibili o mitigabili» (malattie contagiose come il colera e il tifo), i cunei rosa i decessi da ferite e i cunei grigi i decessi conseguenti a qualsiasi altra causa. L'indice di

mortalità negli ospedali britannici raggiunse il valore massimo nel gennaio del 1855, quando 2761 soldati morirono in seguito a malattie contagiose, 83 per ferite e 324 per altre cause, per un totale di 3168 decessi. Basandosi sull'organico medio dell'esercito britannico in Crimea per quel mese, la Nightingale calcolò un indice annuale di mortalità di 1174 per mille. Il diagramma è tratto dal libro della Nightingale *Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army* (1858); la metà di esso, che rappresenta il periodo dall'aprile 1855 al marzo 1856, non compare in figura.

Pur riuscendo a creare un'istituzione modello rispetto ai livelli del tempo, aperta a pazienti di tutte le classi e di tutte le religioni, fu delusa per il fatto di non poter compiere quel che già allora era giunta a considerare il suo obiettivo primario: la fondazione di una scuola per la formazione delle infermiere.

La Nightingale si dedicò a questo primo lavoro solo per un anno perché maggiori opportunità l'attendevano. Nel settembre 1854, le truppe britanniche e francesi invasero la Crimea, sulla costa settentrionale del Mar Nero, inviate in aiuto della Turchia in conflitto con la Russia. (La Russia aveva da tempo ambizioni territo-

riali in Turchia, particolarmente in relazione a Costantinopoli, la città santa degli ortodossi; una delle cause più dirette della guerra di Crimea era stata la richiesta da parte della Russia che le venisse riconosciuto un protettorato sui sudditi ortodossi del sultano.)

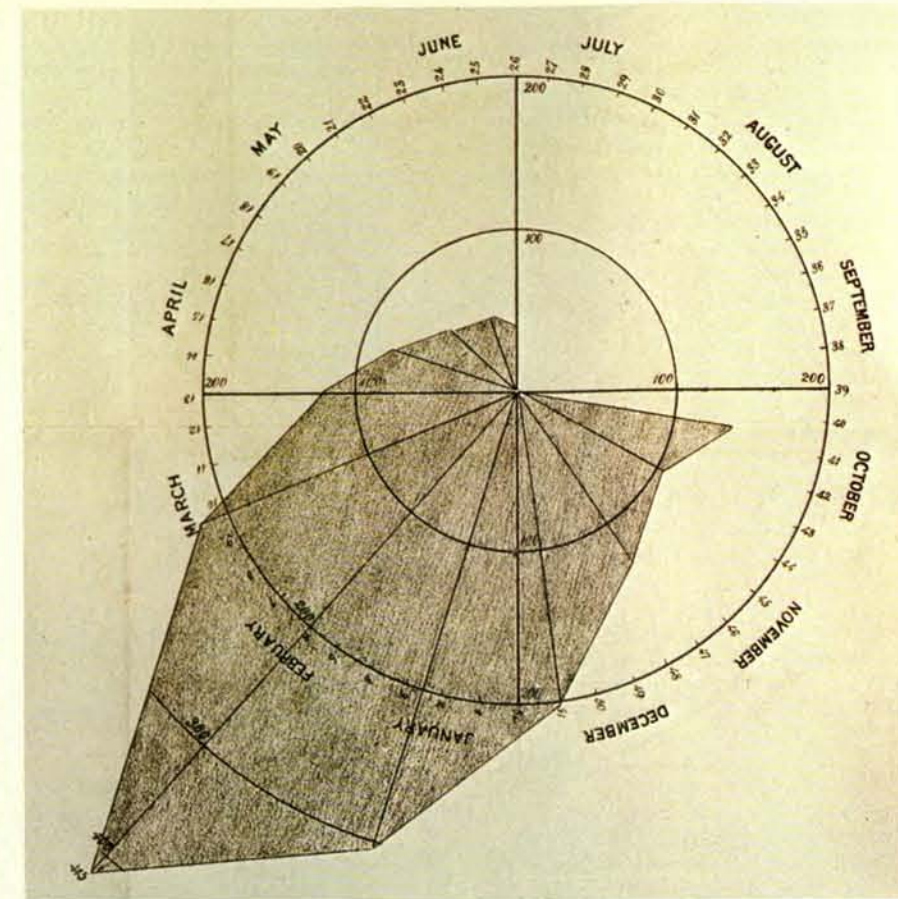
Le forze alleate colsero un rapido suc-

cesso nella battaglia del fiume Alma il 20 settembre, dopo di che iniziarono un assedio alla base navale russa di Sebastopoli. Le manifestazioni di pubblico giubilo in Gran Bretagna lasciarono ben presto il posto alla costernazione quando il corrispondente dalla Crimea del «Times», William Howard Russell, riferì che soldati britannici malati e feriti venivano lasciati morire senza ricevere alcuna assistenza medica. Non solo c'erano troppo pochi chirurghi e mancavano «persino tele di lino per le fasciature», ma nell'ospedale militare britannico di Scutari (nei pressi di Costantinopoli) non c'era neppure un'infermiera qualificata. I francesi avevano invece inviato in Crimea cinquanta Sorelle della Misericordia.

Per l'ambiziosa Nightingale era un'opportunità meravigliosa. Essa scrisse immediatamente a un amico di vecchia data, Sidney Herbert, il «Ministro della guerra», offrendo i suoi servizi. Indipendentemente dalla sua richiesta, Herbert le aveva già spedito una lettera nella quale le chiedeva di reclutare un corpo di infermiere ben addestrate e di guidarle a Scutari. Quando la Nightingale partì per la Turchia, il 21 ottobre 1854, accompagnata da 38 infermiere, aveva l'appoggio ufficiale del governo (anche se non dell'esercito) e, fatto forse più importante, il sostegno finanziario privato di un fondo speciale raccolto dal «Times». Oltre a fare di lei un'eroina nota a livello internazionale, il suo lavoro in Crimea e le condizioni che ebbe modo di conoscere le avrebbero dato un'impronta alla sua missione per tutto il resto della sua vita.

La situazione che la Nightingale e il suo gruppo trovarono al loro arrivo a Scutari, il 5 novembre, giorno dell'importante battaglia di Inkerman, era terribile. Le baracche dell'ospedale erano infestate da pulci e da ratti. Sotto gli edifici, come riferì in seguito una commissione di inchiesta, «c'erano fognature... colme di immondizia... attraverso le quali il vento soffiava esalazioni maleodoranti su per le condutture di numerose latrine aperte nei corridoi e nelle corsie dove i malati giacevano» su pagliericci, in uno stato di sovraffollamento che dopo Inkerman peggiorò ulteriormente. Le lenzuola di tela, secondo la Nightingale, erano «così ruvide che i feriti pregavano di essere lasciati nelle coperte»; la biancheria veniva inoltre lavata in acqua fredda e molti capi che venivano restituiti come puliti erano così «infestati da parassiti» da dover essere distrutti. Mancavano forniture chirurgiche e mediche essenziali, oppure la loro distribuzione era bloccata dalla burocrazia militare.

Queste erano le condizioni che trovavano i pazienti arrivando a Scutari dopo un lento viaggio per mare attraverso il Mar Nero e il Bosforo, deboli ed emaciati, sofferenti di assideramento e di dissenteria oltre che per le ferite. Furono in effetti le epidemie di colera e di tifo conseguenti a queste condizioni, e non le ferite, a mietere il maggior numero di vittime a Scutari. Nel febbraio 1855 l'indice



L'indice di mortalità a Scutari, nel principale ospedale britannico durante la guerra di Crimea, subì una brusca flessione dopo i miglioramenti sanitari introdotti sotto l'influenza della Nightingale. Nell'inverno 1854-1855 l'esercito britannico che assediava la piazzaforte russa di Sebastopoli fu falciato dalla malnutrizione, dall'assideramento e da malattie quali la dissenteria, il colera, il tifo e lo scorbuto. L'indice di mortalità in quella sede, calcolato dalla Nightingale su base annuale come frazione della popolazione di pazienti, raggiunse in febbraio il 415 per cento. Le riforme sanitarie ebbero inizio in marzo. Il diagramma è tratto dalla relazione della Royal Commission, istituita subito dopo la fine della guerra per esaminare le condizioni sanitarie nell'esercito.

di mortalità all'ospedale era pari al 42,7 per cento dei casi trattati.

Nei suoi sforzi per creare un ospedale efficiente in Turchia, la Nightingale dimostrò notevoli doti amministrative. A ogni passo doveva però far fronte agli ostacoli posti dalle autorità militari, le quali opponevano resistenza a ogni mutamento che potesse sembrare un'ammissione dei propri errori e della propria incompetenza. I militari non sopportavano che la Nightingale non fosse soggetta all'autorità dell'esercito e che quei compiti fossero stati affidati a un civile e - peggio ancora - a una donna.

L'ostilità nei confronti della missione della Nightingale fu così forte che in principio le sue infermiere non furono ammesse nelle corsie. Anche dopo aver ottenuto una maggiore considerazione, essa dovette lottare contro ufficiali gretti, come l'incaricato ai vettovagliamenti che si rifiutò di distribuire camicie di cui pure c'era notevole bisogno finché l'intero invio di 27 000 camicie non fu ispezionato da un ufficiale del Comitato di controllo.

Nonostante tali impedimenti, furono

proprio l'indipendenza della Nightingale dai militari e, soprattutto, le sue fonti di finanziamento private a consentirle di ottenere risultati così prestigiosi a Scutari. Essa aprì una propria lavanderia, nella quale si usavano caldaie per far bollire l'acqua; installò in ospedale altre cucine; divenne, infine, la fornitrice dell'intero ospedale, «una sorta di commerciante in qualsiasi genere, in calzini, camicie, coltelli e forchette, cucchiai di legno, vasche da bagno di stagno, tavoli e panche, cavoli e carote, tavoli operatori, asciugamani e sapone, piccoli pettini, un precipitato per distruggere i pidocchi, forbici, padelle da letto e cuscini». Il denaro per questi approvvigionamenti e per il personale reclutato dalla Nightingale fu fornito non solo dal «Times», ma anche da filantropi; la Nightingale attinse inoltre ai propri fondi privati.

Mentre svolgeva i propri compiti amministrativi, la Nightingale trovò anche il tempo di assistere i malati, compito al quale dedicava le ore notturne, compiendo quei giri interminabili nelle corsie che diedero origine alla leggenda



Florence Nightingale è qui fotografata negli anni della maturità. Subito dopo il ritorno dalla Crimea nel 1856, sino alla sua morte, avvenuta nel 1910 all'età di novant'anni, essa visse come un'invalida, confinata

per lo più nella sua camera da letto. Può darsi che la malattia non fosse di natura organica, ma non le impedì di esercitare la sua influenza, ricevendo frequenti visite e mantenendo un'estesa corrispondenza.

dell'«angelo soccorrevole» della Crimea. Di notte allontanava tutte le altre donne dalle corsie (era stata costretta a rimandare in patria alcune infermiere per comportamento scorretto) e camminava, secondo il commissario responsabile del fondo messo a disposizione dal «Times», «da sola, tenendo in mano una piccola lampada», attraverso «quei chilometri e chilometri di uomini sofferenti». Longfellow immortalò quest'immagine della «dama con una lampada» in una poesia del 1857 («Lo! in that house of misery / A lady with a lamp I see»: «Ecco! in quella casa di dolore / Una dama con una lampada io vedo»). C'è però una misura più significativa dei risultati ottenuti dalla Nightingale, e da lei stessa sottolineata: nella primavera del 1855, sei mesi dopo il suo arrivo a Scutari, la mortalità nell'ospedale era diminuita dal 42,7 per cento a soltanto il 2,2 per cento.

La Nightingale tornò in Inghilterra nel luglio 1856, quattro mesi dopo la fine della guerra. A quell'epoca, trentaseienne, era una figura famosa in tutto il mondo e circondata da sentimenti di riverenza. Essa eluse nondimeno ogni tentativo di onorarla pubblicamente e decise, invece, che il modo migliore per riconoscere l'utilità dei servizi da lei svolti sarebbe stato quello di creare una commissione incaricata di investigare sull'assistenza medica militare. In Crimea, scrisse, circa 9000 soldati erano sepolti «nelle loro tombe dimenticate», ed erano morti «per cause che avrebbero potuto essere prevenute». La tragedia dei decessi che si sarebbero potuti evitare stava continuando in ogni caserma e in ogni ospedale militare, anche in tempo di pace. Si poteva mettere fine solo istituendo, nell'intero Servizio medico dell'esercito, quelle medesime riforme sanitarie che avevano salvato un così grande numero di vite umane a Scutari. Questo fu il compito che la Nightingale si prefisse.

Ma in che modo si poteva convincere la gente della necessità della riforma? La Nightingale vide che l'argomento che imponeva maggior rispetto poteva esserle offerto dalla statistica. L'idea di utilizzare dati statistici per un fine del genere - per analizzare le condizioni sociali e l'efficacia delle misure pubbliche - è oggi qualcosa di ovvio, ma non lo era a quel tempo. Nel promuovere la causa della riforma medica la Nightingale divenne anche una promotrice della nascente scienza della statistica sociale.

Considerata semplicemente come raccolta di dati numerici, la statistica ha una lunga storia (che risale perlomeno sino al libro dei *Numeri*, nell'Antico Testamento), ma l'analisi di tali dati ebbe inizio solo con la rivoluzione scientifica del Seicento. I primi tentativi di analizzare dati concernenti fenomeni sociali furono ostacolati da insufficienze sia nei dati stessi sia negli strumenti matematici di analisi. Secondo la storica della statistica Helen M. Walker, la nascita della statistica moderna nell'Ottocento ebbe tre distinte radici: lo sviluppo della teoria matematica delle proba-

bilità, l'emergere dello stato moderno, con i suoi enti per la raccolta di informazioni riguardanti i cittadini e le loro attività, e l'interesse teorico degli economisti politici sulle cause dei comportamenti sociali umani. Questi «tre movimenti», scrisse la Walker, confluirono nella figura dell'astronomo-statistico belga Lambert-Adolphe-Jacques-Quételet, considerato verso la metà dell'Ottocento il fondatore della moderna statistica sociale. Nel 1841 Quételet organizzò in Belgio l'ufficio centrale di statistica, che divenne un modello per enti analoghi in altri paesi; la sua autorità a livello internazionale nell'ambito della ricerca statistica durò per tutta la sua esistenza (Quételet morì nel 1874).

Gli studiosi dell'Ottocento che tentarono di fare dello studio del comportamento umano una scienza si trovarono di fronte a un dilemma: la scienza modello del tempo era la fisica classica, con le sue leggi deterministiche che descrivevano fenomeni naturali, ma il comportamento umano sembrava individuale e indeterminato. La soluzione di questo problema fornita da Quételet aggirò la questione dell'individuo introducendo il concetto di «uomo medio». Quételet mostrò che, pur non essendovi leggi che determinano il comportamento individuale, vi sono regolarità negli attributi e nel comportamento di gruppi, e che queste regolarità potevano essere descritte matematicamente dalle leggi della probabilità. Quételet era convinto che anche i caratteri mentali e morali avrebbero seguito leggi regolari di distribuzione statistica, se solo fosse stato possibile misurarli con precisione.

L'opera più originale e più sorprendente di Quételet fu la sua analisi dell'influenza di fattori come il sesso, l'età, l'istruzione, il clima e la stagione sul tasso di criminalità in Francia (1831). I dati non consentivano di predire chi avrebbe commesso un certo tipo di crimine, ma secondo Quételet manifestavano regolarità che avrebbero permesso a uno scienziato di «predire quanti individui si sarebbero macchiati le mani del sangue dei loro simili, quanti sarebbero stati falsari, quanti avvelenatori». La scoperta di queste regolarità condusse Quételet alla conclusione radicale che «è la società a preparare in qualche modo questi crimini, e il criminale è soltanto lo strumento che li esegue».

Benché l'opera di Quételet abbia goduto di grande considerazione presso molti studiosi, fu aborrita da altri. Il determinismo della sua «fisica sociale» era anatema per persone che accettavano le dottrine prevalenti del libero arbitrio e della responsabilità individuale. John Stuart Mill, per esempio, scrisse molto contro la probabilità in generale e contro la sua applicazione alla scienza sociale in particolare. Un altro vocante oppositore alla concezione statistica dell'uomo e della società fu Charles Dickens. Nel romanzo *Tempi difficili* si propone di mettere in satira quelle persone che, com'egli ebbe occa-

sione di dire in seguito, non riuscivano a vedere altro che «numeri e medie», quelle «teste guaste» che avrebbero usato la temperatura annuale media della Crimea «come una ragione per far indossare a un soldato pantaloni di tela in una notte in cui sarebbe morto congelato in pelliccia». Dickens era avverso alla concezione statistica perché la considerava disumanizzante e in *Tempi difficili* descrisse come un «orologio statistico mortale» le regolarità trovate dagli statistici nell'indice di malattia mentale, crimine, suicidio e prostituzione.

La Nightingale, invece, ammirava molto l'opera di Quételet, e rivelò assai presto una predilezione per la raccolta e l'analisi di dati. A Scutari, a prescindere dalle riforme sanitarie di grande importanza da lei istituite, rese sistematiche le pratiche di registrazione, che erano nel caos più totale; fino ad allora non si conosceva con precisione neppure il numero dei morti. Quando tornò in Inghilterra, nel 1856, conobbe William Farr, medico e statistico di professione. Sotto la sua guida, riconobbe ben presto il potenziale dei dati statistici raccolti a Scutari, e della statistica medica in generale, come strumento per migliorare l'assistenza medica in ospedali militari e civili.

In tutta la storia militare sino al XX secolo, la causa principale dei decessi in guerra furono le malattie più che le ferite subite in battaglia, e la guerra di Crimea non fece eccezione. I dati raccolti dalla Nightingale conservano ancor oggi tutta la loro eloquenza. Durante i primi mesi della campagna di Crimea, vi fu «fra le truppe una mortalità del 60 per cento all'anno per sola *malattia*», un indice superiore a quello della Grande Peste del 1665 a Londra e più alto anche «della mortalità da colera» fra coloro che avevano contratto la malattia. Nel gennaio 1855 la mortalità in tutti gli ospedali britannici in Turchia e in Crimea, calcolata in relazione al totale dei militari presenti in Crimea, ma senza tener conto degli uomini uccisi in azione, raggiunse l'indice annuale massimo del 1174 per mille. Di questo dato, il 1023 per mille dei decessi era attribuibile a malattie «zimotiche» (categoria introdotta da Farr e comprendente malattie epidemiche, endemiche e contagiose). Ciò significa che, se l'indice di mortalità riscontrato in gennaio si fosse mantenuto identico durante un anno intero, e se i soldati morti non fossero stati rimpiazzati, le sole malattie avrebbero spazzato via l'intero contingente britannico in Crimea.

I vari metodi utilizzati dalla Nightingale per calcolare la mortalità misero in grande evidenza sia l'impatto delle malattie sia gli effetti delle migliorate condizioni sanitarie. Calcolato su una base annuale come percentuale della popolazione dei pazienti, l'indice di mortalità all'ospedale di Scutari raggiunse nel febbraio 1855 il valore incredibile del 415 per cento. In marzo, però, cominciarono a essere realizzate le riforme sanitarie della Nightingale e la mortalità dei

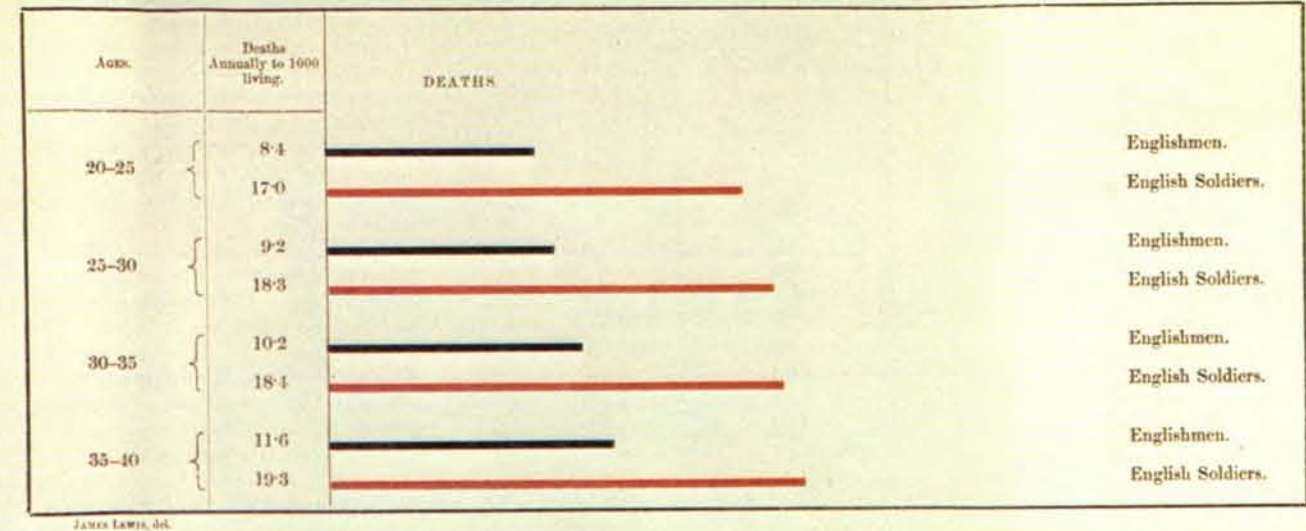
pazienti presentò un brusco calo. Alla fine della guerra, secondo la Nightingale, l'indice di mortalità fra i soldati britannici malati in Turchia non era «molto maggiore» di quello riscontrabile fra soldati sani in Inghilterra; fatto ancor più considerevole, la mortalità fra le truppe britanniche in Crimea fu «solo di due terzi

rispetto a quella esistente fra le nostre truppe in patria.»

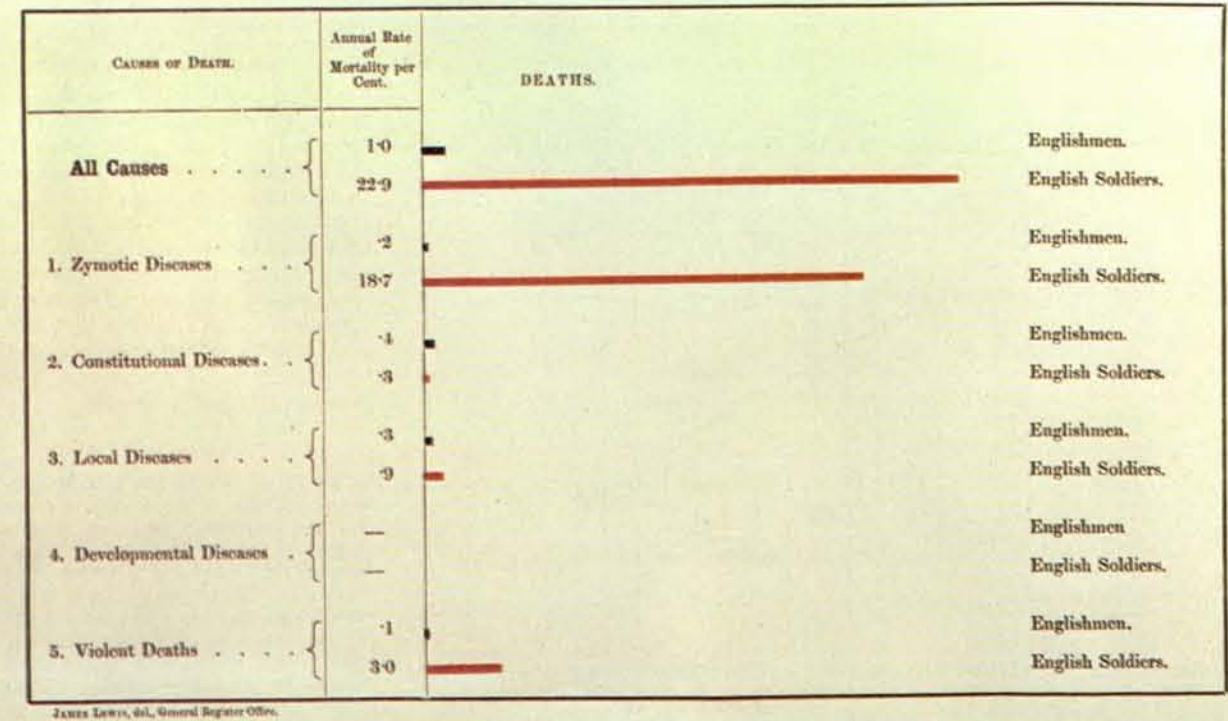
Il confronto suggeriva che i soldati in patria vivevano nelle caserme in condizioni malsane. Dopo che Farr ebbe spiegato alla Nightingale il significato delle tabelle di mortalità, essa pensò subito a confrontare la mortalità fra i civili con

quella fra i militari. Essa trovò che in tempo di pace, in Inghilterra, i soldati di età compresa fra 20 e 35 anni presentavano un indice di mortalità quasi doppio di quello tra i civili. È altrettanto criminale, scrisse nel 1857, «avere una mortalità del 17, 19 e 20 per mille in fanteria, in artiglieria e nella guardia in Inghilterra, men-

Representing the Relative Mortality of the Army at Home and of the English Male Population at corresponding Ages.

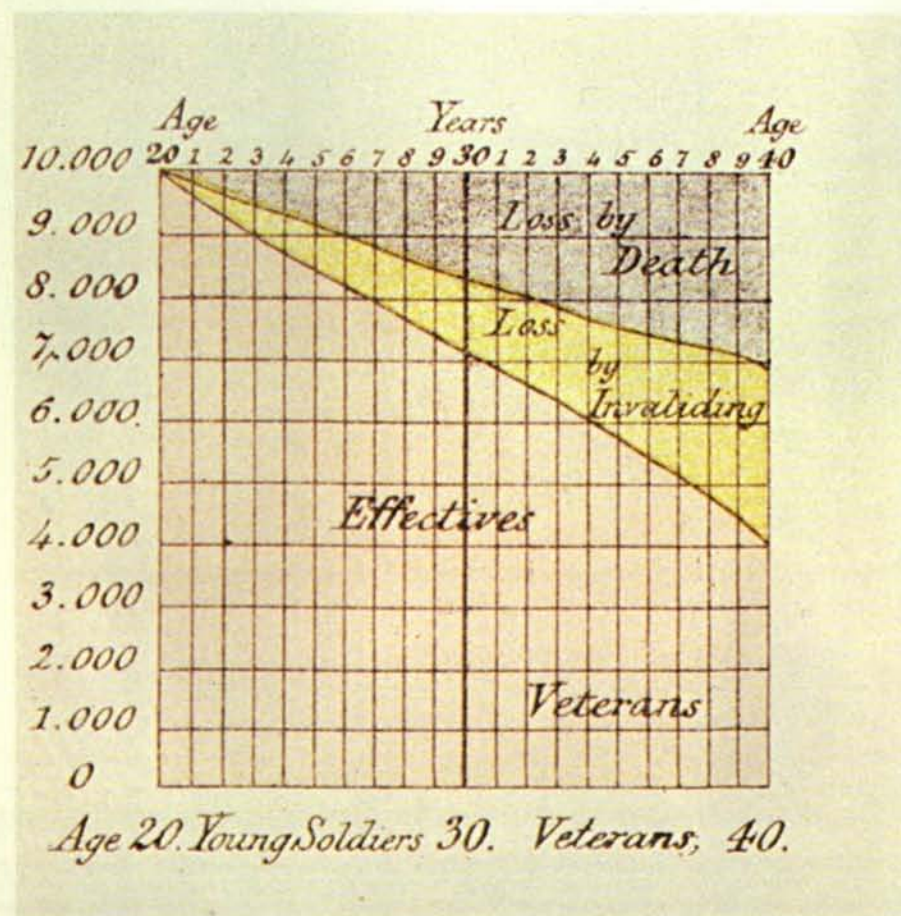
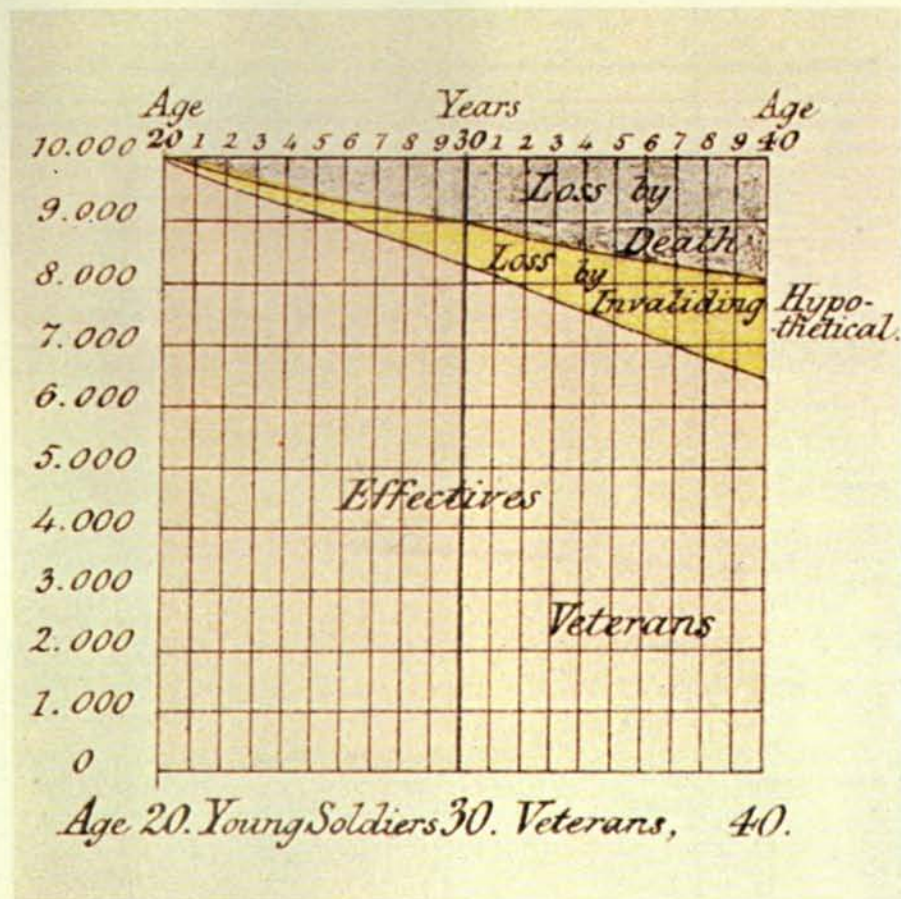


Representing the Relative Mortality, from different Causes, of the Army in the East in Hospital and of the English Male Population aged 15-45.



Questi diagrammi, tratti dalla relazione della Royal Commission, mettono a confronto le condizioni nell'esercito e nella vita civile. La mortalità fra i soldati in tempo di pace in Gran Bretagna era press'a poco doppia che fra i civili di sesso maschile di età corrispondenti (*in alto*). Nella guerra di Crimea, tra i soldati ricoverati in ospedale, le malattie «zimotiche» furono, tra le varie cause di morte (malattie costituzionali, malattie locali, malattie dello sviluppo, atti di violen-

za), quelle che incisero maggiormente, più che non tra i civili di sesso maschile, di età tra i 15 e i 45 anni, in Inghilterra (*in basso*). Le cifre nella seconda colonna del diagramma in alto indicano il numero di decessi per mille; quelle nel diagramma in basso sono percentuali. La relazione portò all'adozione di un codice sanitario per l'esercito e a miglioramenti in edifici militari. È questo un esempio dell'uso innovativo fatto dalla Nightingale di rappresentazioni grafiche in statistica.



tre quella della vita civile è solo dell'11 per mille, quanto lo sarebbe portare 1100 uomini all'anno nella Salisbury Plain e fucilarli». (I 1100 uomini ai quali si riferiva la Nightingale rappresentavano il 20 per mille di una forza arruolata di 55 000 uomini.) È chiaro che il bisogno di miglioramenti sanitari fra i militari non era limitato agli ospedali da campo. Sostenendo la sua causa con questi dati statistici, la Nightingale richiamò infine su di sé l'attenzione della regina Vittoria e del principe Alberto, oltre che del primo ministro, Lord Palmerston. Nonostante la resistenza passiva del Ministero della guerra, il suo desiderio che venisse compiuto uno studio ufficiale sull'assistenza sanitaria nelle forze armate fu realizzato nel maggio 1857, con la fondazione di una apposita commissione, la Royal Commission, per lo studio della sanità nell'esercito.

A quel tempo non sarebbe stato possibile per una donna far parte di una tale commissione. La Nightingale esercitò nondimeno una forte influenza sulla sua attività, sia perché alcuni membri (fra cui Sidney Herbert, il ministro che l'aveva mandata in Crimea) erano suoi amici, sia perché fu lei a fornire alla commissione la maggior parte delle informazioni di cui essa poté disporre. Come espressione delle proprie idee, la Nightingale scrisse e fece stampare privatamente un libro di 800 pagine intitolato *Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army* (Note su questioni concernenti la sanità, l'efficienza e l'amministrazione ospedaliera dell'esercito britannico), comprendente una sezione di statistiche accompagnate da diagrammi. Farr definì il libro «la [cosa] migliore che sia mai stata scritta» sia sui diagrammi statistici sia sull'esercito.

La Nightingale fu una vera anticipatrice nel campo della rappresentazione grafica in statistica. Essa inventò i diagrammi ad aree polari, in cui i dati statistici vengono rappresentati in modo proporzionale all'area di un cuneo in un diagramma circolare e utilizzò questi diagrammi, da lei chiamati *coxcombs* (berretti da giulla-

La perdita di uomini nell'esercito britannico in conseguenza di un eccesso di mortalità e di cause invalidanti viene qui illustrata da diagrammi tratti dalla relazione della Royal Commission. Alla base di entrambi i grafici c'è l'assunto che 10 000 reclute di vent'anni di età vengano aggiunte ogni anno agli effettivi e che la carriera di un soldato sano duri vent'anni. Ogni rettangolino rappresenta mille uomini. In condizioni malsane (*in basso*) i decessi (*area in marrone*) e le invalidità (*area in giallo*) riducono gli effettivi dell'esercito (*area in beige*) a 141 764 unità a partire da un valore massimo di 200 000, con una perdita del 29 per cento. Se la mortalità fosse bassa come nella popolazione civile e il rapporto fra mortalità e indice di invalidità restasse lo stesso, concludeva la relazione, la forza del contingente britannico sarebbe cresciuta in misura significativa fino a 166 910 unità (*in alto*).

re) a causa dei loro colori vivaci, per sottolineare in modo drammatico in quale misura i decessi nella campagna di Crimea avrebbero potuto essere prevenuti. Farr fu impressionato dalle sue *Notes* e accolse gran parte dell'opera della Nightingale nei diagrammi statistici che preparò per la relazione finale della Royal Commission. Nell'ambito della sua «marcia di fiancheggiamento» contro le forze della resistenza alla riforma medica, la Nightingale fece stampare la sezione statistica della relazione come opuscolo a sé e gli diede ampia diffusione in Parlamento, nel governo e nell'esercito. Essa fece addirittura incorniciare alcune copie dei diagrammi per farne dono a esponenti del Ministero della guerra e del Dipartimento medico dell'esercito.

Gli sforzi della Nightingale non furono vani. Furono create quattro sottocommissioni per realizzare le riforme raccomandate nella relazione della Royal Commission. La prima si occupò dei miglioramenti fisici alle caserme e agli ospedali militari: miglioramenti nella ventilazione, nel riscaldamento, nell'eliminazione delle acque di scolo, nell'approvvigionamento d'acqua e nelle cucine. Altre sottocommissioni elaborarono un codice sanitario per l'esercito, fondarono una scuola medica militare e riorganizzarono i procedimenti in uso nell'esercito per la raccolta di dati statistici di interesse medico.

La Nightingale rivolse poi la sua attenzione alla salute dei militari britannici in India. Insieme a Farr cominciò a studiare i registri dell'India Office, sulle malattie e sulla mortalità, e inviò moduli alle varie sedi di guarnigioni militari in India per avere dati sulle condizioni sanitarie vigenti in quelle località. Nel 1858 e nel 1859 fece molte pressioni per far istituire un'altra commissione, che si occupasse della questione indiana, ottenendo infine il successo. Due anni dopo sottopose alla commissione un rapporto, basato sui questionari da lei inviati alle sedi di guarnigioni in India, sulle condizioni che stavano causando fra le truppe un indice di mortalità sei volte maggiore di quello dei civili in Inghilterra: sistemi di fognature insufficienti, sovraffollamento nelle caserme, mancanza di esercizio fisico e ospedali inadeguati, fra varie altre cose. La commissione completò il suo studio nel 1863. Dopo 10 anni di riforma sanitaria, nel 1873 la Nightingale riferì che la mortalità fra i soldati in India era diminuita dal 69 al 18 per mille.

Le statistiche, come la Nightingale dimostrò in modo così efficace, offrono un modo organizzato di apprendimento dall'esperienza, e le statistiche mediche possono insegnarci assai più del semplice fatto che le condizioni malsane uccidono. Statistiche ospedaliere uniformi e accurate, scrisse la Nightingale, avrebbero «consentito di valutare statisticamente il valore di particolari metodi di trattamento e di speciali operazioni»; in breve, le statistiche avrebbero condotto a un miglioramento nella pratica medica e chirurgica. Il problema consisteva nel fatto che le stati-

stiche tenute dagli ospedali al tempo della Nightingale non erano né uniformi né sempre accurate. Per ovviare a questo inconveniente essa sviluppò, con l'aiuto di Farr e di altri medici, un formulario statistico modello per ospedali. Questo venne approvato dal Congresso internazionale di statistica che si tenne a Londra nell'estate del 1860.

Il nuovo schema fissò le categorie base di dati che gli ospedali dovevano raccogliere: il numero di pazienti ricoverati in ospedale all'inizio e alla fine di un anno e il numero di pazienti ammessi durante l'anno, il numero di pazienti che erano guariti o che erano stati dimessi come incurabili o a loro richiesta, il numero di pazienti che erano deceduti e la durata media della degenza in ospedale. Eppure, benché l'ideale di raccogliere statistiche ospedaliere uniformi fosse chiaramente buono, e molto in anticipo rispetto ai tempi, il nuovo schema non entrò mai in una pratica generalizzata. Il formulario proposto era eccessivamente complesso e comprendeva un sistema molto personale, escogitato da Farr per la classificazione delle malattie, nei confronti del quale molti patologi dissentivano fortemente. Nella scienza medica, purtroppo, la Nightingale non manifestò la stessa lungimiranza che l'aveva condotta a riconoscere il valore delle statistiche mediche; per esempio, essa non manifestò alcun interesse per la nuova teoria dei germi come agenti patogeni e per le implicazioni che tale teoria aveva nella cura delle malattie contagiose.

L'interesse della Nightingale per le statistiche trascese il suo interesse per la riforma dell'assistenza sanitaria e fu strettamente connesso alle sue convinzioni religiose. Per lei, le leggi che governavano i fenomeni sociali, «le leggi del nostro progresso morale», erano leggi di Dio e dovevano essere rivelate dalla statistica. La scienza di Quételet, essa spiegava, era «essenziale all'intera amministrazione politica e sociale»; eppure la maggior parte dei leader politici non aveva alcuna preparazione per interpretare le statistiche. Il risultato di tale ignoranza, nel pensiero della Nightingale, era una legislazione «non progressista ma altalenante», scritta da funzionari che «legiferano senza conoscere ciò che stanno facendo». Ecco perché la Nightingale sperimentò rappresentazioni grafiche, che tutti potevano capire, e si batté per introdurre lo studio della statistica nell'istruzione superiore, anche se il suo sogno di istituire una cattedra universitaria di statistica non divenne realtà se non dopo la sua morte. Ancor oggi la società non ha accettato pienamente il punto di vista della Nightingale, come risulta evidente dal fatto che la statistica non è ancora diventata una parte obbligatoria della pubblica istruzione.

Qualcosa del fervore religioso che la Nightingale provava per la statistica traspare da una nota sulla sua copia personale del libro di Quételet *Physique sociale*. Sul frontespizio, essa incluse il titolo in una formulazione del suo proprio credo:

«Il senso di potere infinito
Le assicurazioni di una
concreta certezza
La prospettiva senza fine
di un miglioramento
in base ai principi della
PHYSIQUE SOCIALE
se solo fosse possibile applicarli
in occasioni in cui
ne è tanto grande il bisogno»

Per la Nightingale, Quételet era il fondatore della «scienza più importante nel mondo intero», perché «ne dipende l'applicazione pratica di ogni altra [scienza]». A giudicare dalla loro corrispondenza, pare che il rispetto sia stato reciproco.

Benché la statistica fosse un campo così importante per la Nightingale, negli anni seguenti, durante i quali la sua «influenza» andò aumentando sempre più, crebbe in lei anche il desiderio di tornare alla professione di infermiera, la prima «vocazione» che aveva avuto «da Dio». Essa non poté però realizzare questo desiderio perché dopo il ritorno dalla Crimea dovette trascorrere buona parte della sua vita come un'invalida.

La cattiva salute della Nightingale è stata messa in relazione con uno stato febbrile da lei contratto in Crimea; tuttavia, alcuni hanno suggerito che non si sia trattato affatto di una malattia organica, ma che la sua invalidità sia stata nevrotica o addirittura intenzionale. In ogni caso il confinamento nella sua camera da letto, dove essa ricevette un flusso continuo di visitatori, non diminuì per nulla la sua influenza né le impedì di fondare lo status professionale dell'assistenza infermieristica moderna. Con il denaro del fondo Nightingale (quasi 50 000 sterline, raccolte con una sottoscrizione pubblica in onore dell'«Eroina popolare»), essa poté realizzare un antico sogno, fondando nel 1860 una scuola per infermiere, la Nightingale Training School for Nurses. Non ebbe la possibilità come aveva sperato, di dirigerla, ma in essa vennero seguiti i suoi principi: «1) che le infermiere debbano avere il loro addestramento tecnico in ospedali specificamente organizzati a tale fine; 2) che debbano vivere in una casa idonea a formare la loro vita e disciplina morale.»

Entrambi i principi erano rivoluzionari a quel tempo. Che essi siano accettati oggi come ovvi è una testimonianza di ciò che Florence Nightingale ha fatto per la professione infermieristica, cosa che contribuì non meno di qualsiasi progresso scientifico al miglioramento della qualità generale dell'assistenza medica. In riferimento all'altra sua passione, è giusto che un altro eloquente indice dei risultati conseguiti con la sua opera sia statistico: nel 1861 il censimento britannico accertò la presenza in Gran Bretagna di 27 618 infermiere, che furono elencate nelle tabelle delle occupazioni alla rubrica «domestici»; nel 1901 il numero delle infermiere era aumentato a 64 214, ed era elencato sotto la rubrica «medicina».

(RI)CREAZIONI AL CALCOLATORE

di Brian Hayes

*L'automata cellulare offre un modello
del mondo e un mondo in se stesso*

È davvero sorprendente che le molecole d'acqua «sappiano» come formare le elaborate simmetrie di un fiocco di neve. Non c'è architetto a dirigerne le combinazioni e le molecole non hanno uno «stampo» per la forma cristallina. La configurazione su larga scala emerge completamente dalle interazioni a breve raggio di molte unità identiche. Ogni molecola risponde solo all'influenza dei suoi immediati adiacenti, tuttavia una disposizione coerente si mantiene in una struttura fatta forse di 10^{20} molecole.

Un modo per avvicinarsi alla comprensione di questo processo consiste nell'immaginare che ogni sito in cui può prender posto una molecola sia governato da un rudimentale calcolatore. Col crescere del cristallo, ogni calcolatore passa in rassegna i siti circostanti e, a seconda di ciò che trova, stabilisce attraverso qualche regola prefissata se il proprio sito debba essere occupato o rimanere vuoto. Lo stesso calcolo viene fatto in tutti i siti secondo la stessa regola.

Il modello computazionale della crescita di un fiocco di neve è un automata cellulare: una schiera uniforme di numerose celle identiche, in cui ogni cella può assumere solo pochi stati e interagisce solo con poche celle adiacenti. I componenti del sistema (le celle e la regola per calcolare lo stato successivo di una cella) possono essere davvero semplici e nondimeno dar luogo a un'evoluzione notevolmente complessa.

L'idea dell'automata cellulare è vecchia approssimativamente quanto il calcolatore elettronico digitale. Le prime ricerche furono condotte da John von Neumann (con un importante contributo di Stanislaw Ulam) nei primi anni cinquantenni. L'obiettivo principale di von Neumann era escogitare un semplice sistema capace di riprodursi alla maniera di un organismo vivente. Anche il più noto automata cellulare, il gioco «Vita» inventato nel 1970 da John Horton Conway, ha un aspetto biologico, come suggerisce il nome; le cellule nascono, vivono o muoiono a seconda della densità della popolazione locale.

Nelle ricerche più recenti sugli automi cellulari, il centro di interesse si è leggermente spostato. Le schiere di celle a inte-

razione locale sono viste come modelli potenzialmente utili di sistemi fisici, dai fiocchi di neve ai ferromagneti alle galassie. Inoltre possono trovare applicazioni in problemi della scienza dei calcolatori, sia pratica (come si dovrebbe organizzare una rete di molti calcolatori interagenti?) sia teorica (qual è il limite ultimo alla potenza di una macchina per il calcolo?). La cosa forse più interessante è che l'automata cellulare può essere visto come un «universo digitale» che val la pena di esplorare per se stesso, al di là della sua utilità come modello del mondo reale.

Il risorgere dell'interesse per gli automi cellulari è stato evidenziato da un seminario tenuto sull'argomento nel 1983 al Los Alamos National Laboratory. Gli atti (circa 20 articoli) sono stati in seguito pubblicati in «Physica D» e, in forma di libro, dalla North-Holland Publishing Company. Il contenuto del presente articolo si basa quasi tutto su quanto discusso nell'incontro di Los Alamos.

Quattro proprietà caratterizzano un automa cellulare. La prima è la geometria della matrice di celle. Per un modello della crescita di un fiocco di neve sarebbe adeguata una matrice esagonale a due dimensioni, ma nella maggior parte dei contesti si sceglie un reticolo rettangolare fatto di quadrati identici. È facile costruire matrici in tre o più dimensioni, ma non è facile visualizzarle. Recentemente sono state fatte sorprendenti scoperte con l'ancor più semplice matrice unidimensionale: una semplice fila di celle.

All'interno di una matrice data è necessario specificare l'intorno che ogni cella esamina nel calcolo del suo stato successivo. Nella matrice rettangolare bidimensionale si è prestata molta attenzione a due intorni. Von Neumann limitava l'attenzione di ogni cella alle quattro adiacenti più vicine, quelle a nord, sud, est e ovest; questo insieme di celle è chiamato ora intorno di von Neumann. L'intorno che comprende queste quattro celle e le quattro diagonalmente adiacenti è detto intorno di Moore, dal nome di Edward F. Moore. Ovviamente gli intorni si sovrappongono e una data cella è inclusa simultaneamente negli intorni di parecchie celle adiacenti. In qualche caso la cella di centro - la cella che effettua un calcolo - è considerata membro del suo stesso intorno.

Il terzo fattore da prendere in conside-

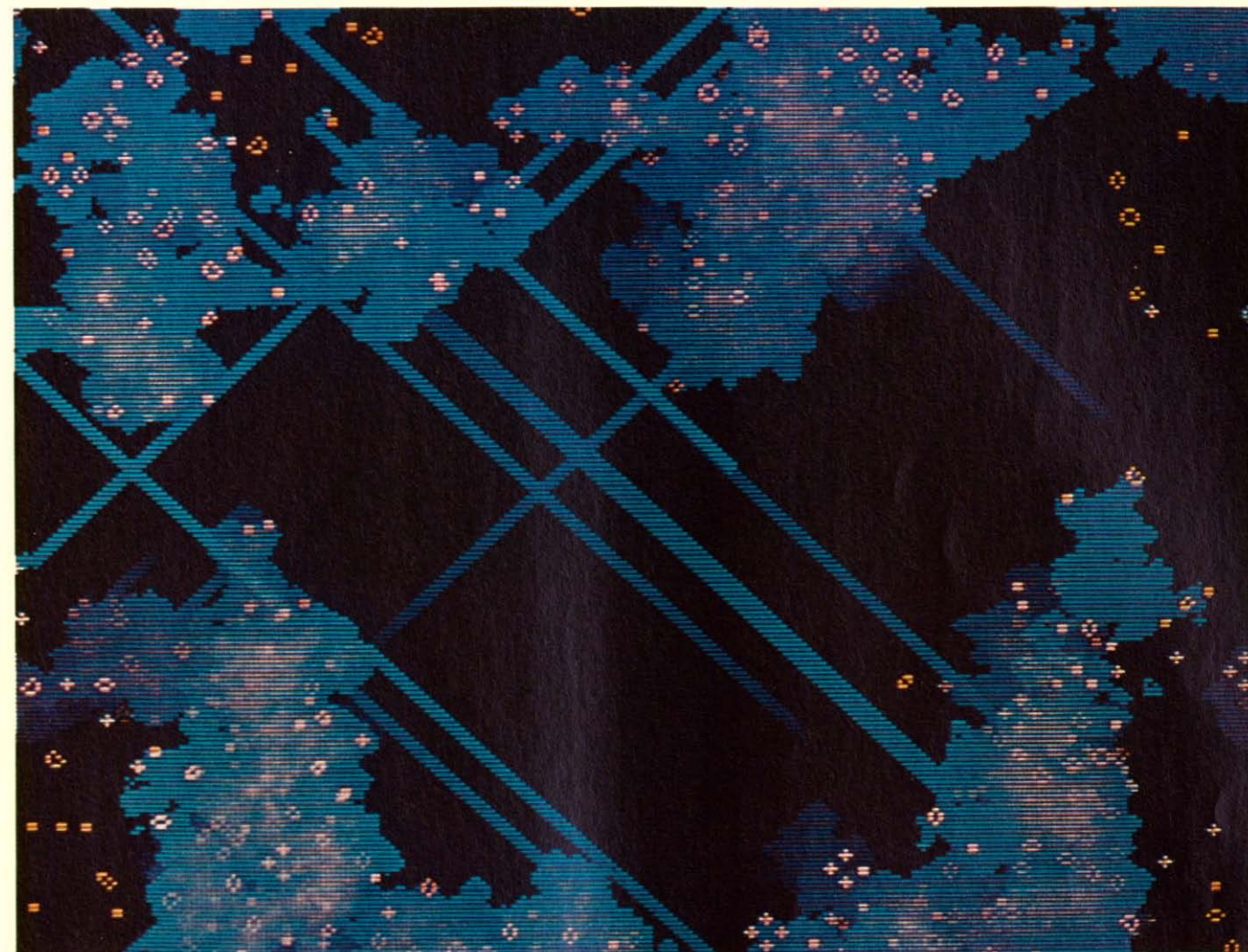
razione per descrivere un automa cellulare è il numero di stati per cella. Von Neumann aveva scoperto una configurazione autoreplicante fatta di celle con 29 stati possibili, ma la maggior parte degli automi sono molto più semplici. C'è, in effetti, un'ampia possibilità di variazione anche tra gli automi binari, quelli con due stati per cella; gli stati possono essere rappresentati come 1 o 0, vero o falso, acceso o spento, vivo o morto.

La principale fonte di varietà nell'universo degli automi cellulari è l'enorme numero delle regole possibili, per stabilire il futuro stato di una cella, basate sull'attuale configurazione del suo intorno. Se k è il numero di stati per cella e n è il numero di celle incluse nell'intorno, vi sono k^n possibili regole. Per un automa binario, quindi, nell'intorno di von Neumann (dove n è 4) ci sono più di 65 000 possibili regole; nell'intorno di Moore (dove n è 8) ve ne sono 10^{77} . Solo una frazione insignificante di tali regole è stata esaminata.

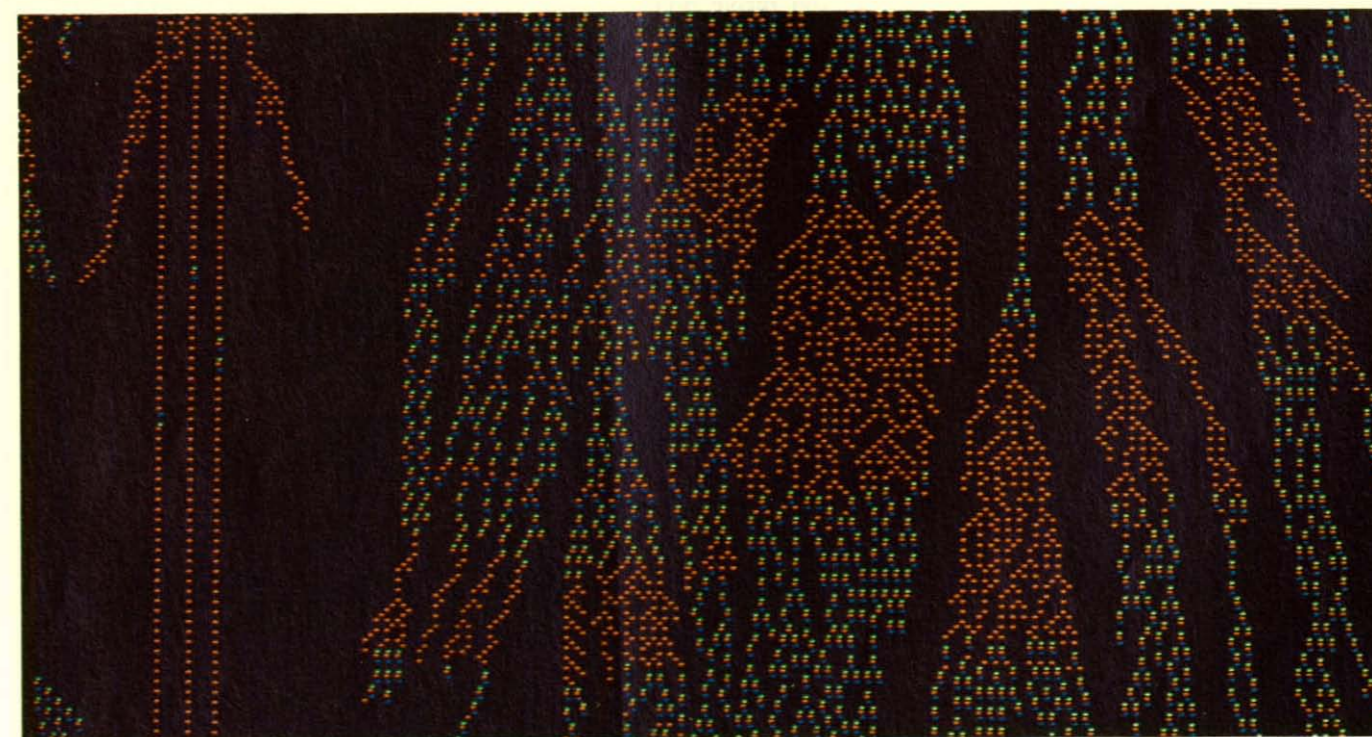
Il gioco Vita si effettua con celle a due stati su un reticolo rettangolare nell'intorno di Moore, con la complicazione che la cella centrale è significativa. In altri termini, a ogni passo dell'evoluzione del sistema ogni cella controlla sia lo stato delle otto celle circostanti sia il proprio. Secondo la regola definita da Conway, se la cella centrale è viva continuerà a vivere nella generazione successiva se due o tre delle otto celle dell'intorno sono anch'esse vive. Se ci sono tre celle vive nell'intorno, la cella centrale è viva nella generazione successiva indipendentemente dal suo stato attuale. In tutti gli altri casi, la cella centrale muore o rimane morta.

Il fascino del gioco Vita sta nella sua imprevedibilità. Alcune configurazioni muoiono completamente; molte di più cadono in una configurazione stabile o in una ciclica con un periodo di poche generazioni. Nel corso degli anni, comunque, si sono scoperti un certo numero di stati iniziali più interessanti, come il «cannone ad alianti», che lancia un fiume interminabile di proiettili. L'esplorazione dei meandri della vita continua. Alcuni recenti sviluppi sono descritti da Martin Gardner in *Wheels, Life, and Other Mathematical Amusements* (pubblicato da Freeman & Co., New York, nel 1983). Vorrei ora passare ad altri automi cellulari con proprietà che solo da poco cominciano a essere chiarite.

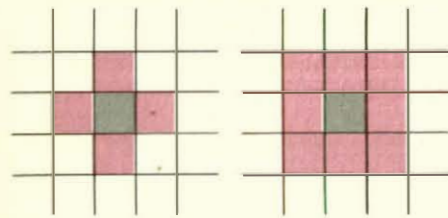
Tra la moltitudine di possibili regole di transizione, molte sono di scarso interesse intrinseco. Per esempio, una regola che stabilisce che una cella sarà «accesa» se e solo se la cella alla sua sinistra è accesa determina un'evoluzione decisamente facile da prevedere: una qualsiasi configurazione iniziale conserva la sua forma ma si sposta a destra di una cella a ogni passo nel tempo. Una sottoclasse di regole dette regole di conto o regole totalistiche sembra includere esempi di quasi tutte le varietà osservate di automi cellulari. Con regole di questo tipo, il nuovo stato di una cella dipende solo dal numero di



Il gioco «Vita» evolve sullo schermo della macchina automa cellulare di Tommaso Toffoli



Una configurazione di dendriti è creata da un automa cellulare con una regola di transizione asimmetrica



Gli intorni di von Neumann e di Moore

adiacenti in un dato stato, non dalla loro posizione. Molti automi basati su queste regole sono stati studiati da membri dell'Information Mechanics Group del Laboratory for Computer Science al Massachusetts Institute of Technology. Il gruppo è formato da Edward Fredkin, Norman Margolus, Tommaso Toffoli e Gérard Y. Vichniac.

Una delle più semplici regole di conto è la regola della parità, che assegna il valore 1 a una cella se un numero dispari di celle adiacenti hanno valore 1; in caso contrario le assegna il valore 0. L'evoluzione di questo sistema, quando la regola è applicata nell'intorno di von Neumann, è stata descritta su queste colonne nel fascicolo di dicembre 1983. Qualsiasi configurazione di partenza è replicata quattro volte; le quattro copie sono poi replicate a loro volta, e così via.

Un'altra classe di regole di conto è costituita dalle regole «di votazione», che danno il valore 1 alla cella centrale ogniqualvolta il numero degli 1 nell'intorno supera una certa soglia. Vichniac, in un saggio presentato all'incontro di Los Alamos, sottolinea come regole di questo tipo diano modelli della percolazione e della nucleazione, fenomeni importanti nella fisica dello stato solido e in altri campi. Percolazione è il termine adottato per la formazione di un percorso ininterrotto attraverso qualche spazio; per esempio, quando un metallo si disperde in una matrice isolante, la conducibilità del composito dipende dalla probabilità di formare una catena continua di atomi di metallo; proprio come la trasmissione di una malattia infettiva è possibile attraverso una sequenza ininterrotta di individui adatti. La nucleazione è il processo che dà inizio alla crescita di un cristallo, all'ebol-

lizione di un liquido e a eventi analoghi.

Una regola di transizione che dà origine alla percolazione assegna il valore 1 alla cella centrale se vi sono degli 1 in almeno tre delle cinque celle dell'intorno di von Neumann più la cella centrale. L'avvio della percolazione è estremamente sensibile alla concentrazione iniziale di 1. Se la concentrazione è inferiore a un mezzo, probabilmente non si formeranno, nel corso dell'evoluzione, catene continue di 1 attraverso la matrice. Con una concentrazione pari o superiore a un mezzo, le catene appaiono ma il reticolo non si riempie ancora totalmente di 1 e rimangono isole di 0 nello stato stabile finale. Si incontra la nucleazione, con la matrice che si riempie completamente di 1, quando la regola è modificata in modo da richiedere solo due 1 su cinque celle. La concentrazione critica è pari a 0,0822.

Il modello di Ising è uno strumento concettuale della fisica che appare superficialmente molto simile a un automa cellulare. Il modello è un reticolo rettangolare in cui ogni sito ha due valori possibili e interagisce solo con i suoi quattro adiacenti più vicini. Si impiega spesso questo modello per descrivere materiali ferromagnetici; ogni sito rappresenta uno spin atomico che deve puntare o verso l'alto o verso il basso. Al di sotto di una temperatura critica (la temperatura di Curie) gli spin tendono a essere allineati e il materiale è allora magnetizzato, ma a temperature superiori sono distribuiti più o meno casualmente.

Nel numero di dicembre ho parlato di una versione del modello di Ising creata con un programma di tabellone elettronico; il suo reticolo di celle si presta in modo naturale a essere studiato in termini di automi cellulari, per quanto si tratti di un reticolo con regole probabilistiche per simulare la temperatura. Avevo osservato un fenomeno curioso: a bassa temperatura gli spin non assumevano un allineamento uniforme in una direzione e adottavano invece una configurazione a scacchiera di spin alternativamente rivolti verso l'alto e verso il basso. A ogni passo nel tempo tutti gli spin si ribaltavano. In un ferromagnete, la disposizione a scacchiera è la configurazione di massima energia e dovrebbe quindi essere instabi-

le; è la disposizione caratteristica di un antiferromagnete.

Vichniac aveva già individuato il problema e l'aveva spiegato. Nella realizzazione standard del modello di Ising, a ogni iterazione solo uno spin può cambiare. Ne consegue che quando un particolare sito passa in rassegna il suo intorno, alcuni degli spin esaminati sono «vecchi» e alcuni sono «nuovi». In queste condizioni, non può nascere l'oscillante antiferromagnete. Solo quando vengono ricalcolati simultaneamente tutti gli spin viene favorito l'antiferromagnete ad alta energia. Vi sono strategie per evitare questa «catastrofe da retroazione», ma la lezione più significativa da trarne è che è fuorviante attenersi alla più semplice corrispondenza intuitiva tra il modello di Ising e gli automi cellulari.

Vichniac e altri del gruppo del MIT rilevano che gli automi cellulari hanno uno status fondamentalmente differente da quello di altri modelli fisici. Lo stragemma più comune per costruire un modello matematico del mondo naturale è stato a lungo l'equazione differenziale, che può descrivere il cambiamento in una certa grandezza come funzione della posizione e del tempo. Per esempio, le equazioni di Maxwell forniscono la variazione del valore di un campo elettromagnetico da punto a punto e da istante a istante. Tutte le grandezze in queste equazioni variano con continuità. Un automa cellulare, invece, è un sistema totalmente discreto. Lo spazio non è un continuum ma una matrice di celle; anche il tempo è spezzettato in passi discreti e mentre l'intensità di un campo può variare su un dominio continuo, le celle di un automa cellulare possono avere solo un numero finito di stati.

Naturalmente, lo spazio reale, il tempo e molte variabili fisiche sono ritenuti continui anziché discreti (almeno nella scala comunemente presa in considerazione). Non ne consegue, però, che le equazioni differenziali portino di per sé a modelli della natura più validi. Spesso non è il preciso valore numerico di una variabile a essere significativo, ma solo la sua dimensione globale, come il fatto che un particolare punto di un fiocco di neve in crescita sia ghiaccio o vapore acqueo. Gli automi cellulari rendono esplicita questa «discre-

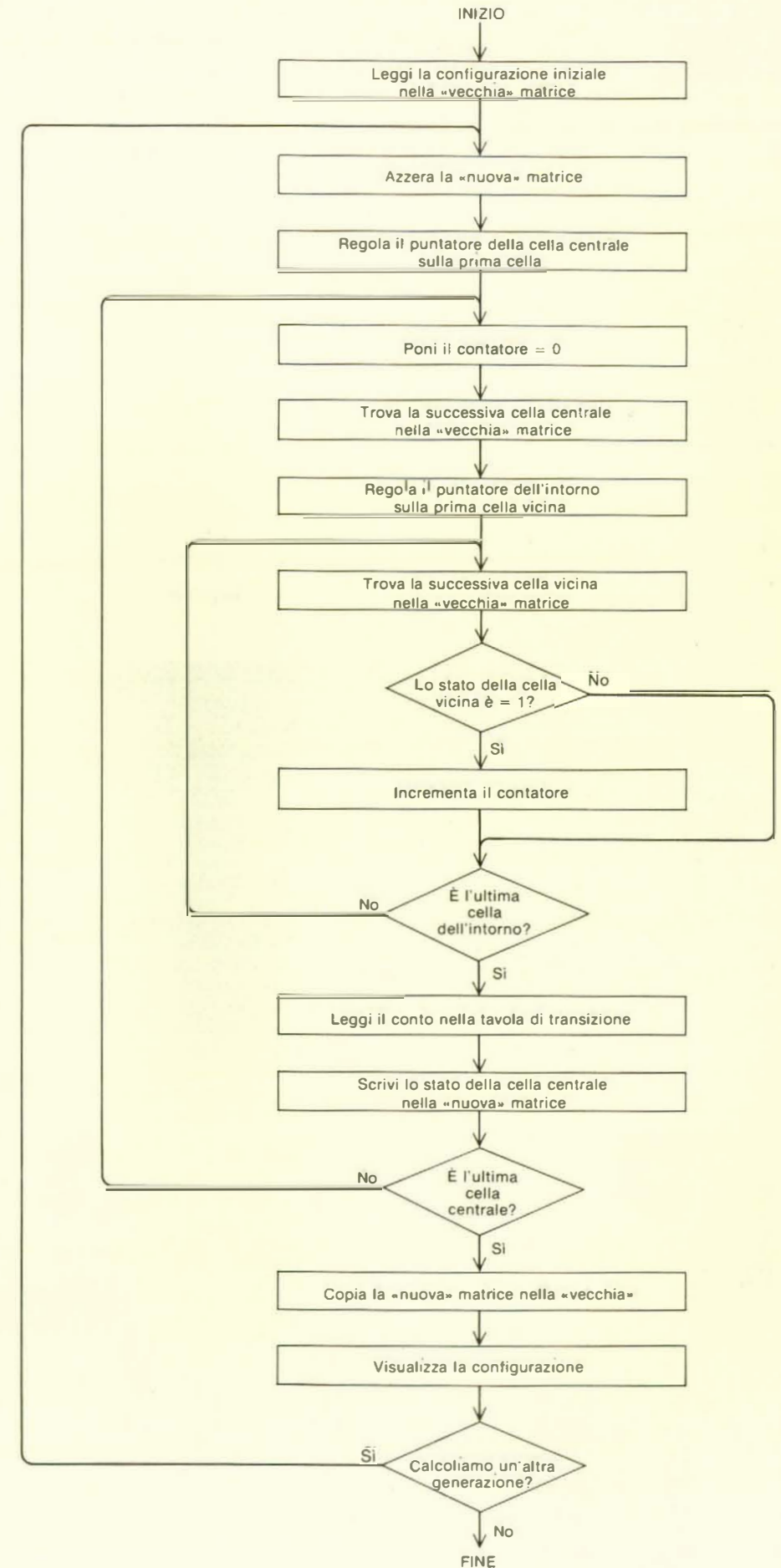
tezza» su un calcolatore digitale. Inoltre, si può calcolare esattamente la loro evoluzione nel tempo, senza bisogno di approssimazioni. Oltre a ciò, possono fare un uso molto più efficace delle risorse del calcolatore digitale.

Un programma per simulare un automa cellulare può essere scritto anche per i calcolatori più piccoli. Per Bak, del Brookhaven National Laboratory, ha di recente sostenuto in «Physics Today» che molte simulazioni si possono effettuare su un piccolo calcolatore personale più efficacemente e a minor costo che su più potenti macchine usate in partizione di tempo. L'esempio scelto a titolo illustrativo era una simulazione del modello tridimensionale di Ising, fatta con un Commodore VIC 20 e a un costo stimato in 4 dollari.

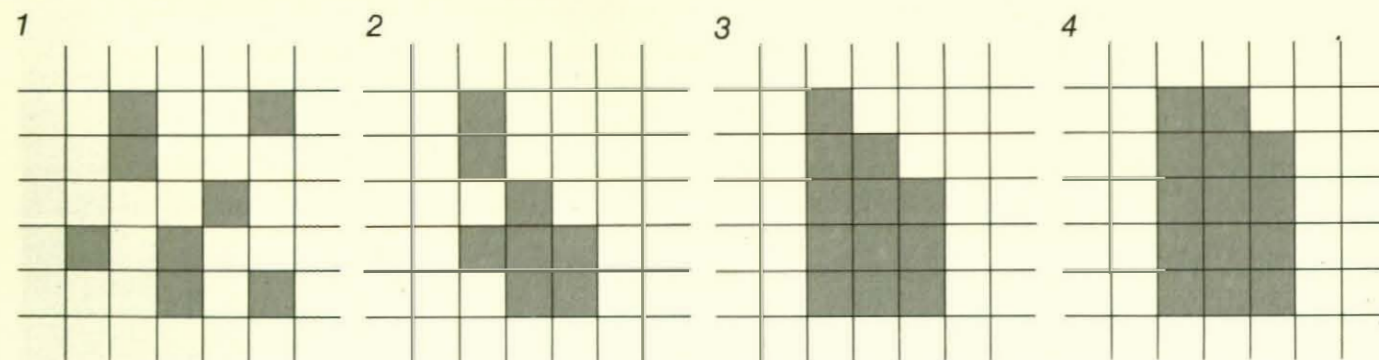
Il più semplice programma per automa cellulare include semplicemente il metodo che probabilmente si adotterebbe per effettuare la procedura a mano su carta millimetrata. Innanzitutto si stabilisce una matrice di celle, con ogni cella rappresentata da un elemento di memoria del calcolatore. A ogni passo nel tempo, il calcolatore deve occuparsi a turno di ogni cella, esaminare le sue adiacenti e calcolare il valore appropriato del successivo stato della cella. Il calcolo stesso è svolto comodamente cercando il valore in una tabella. Se si considerano solo regole di conto, la tabella richiede soltanto un'entrata per ogni possibile numero di celle «accese». Quando sono ammessi altri tipi di regole, la tabella può diventare piuttosto complicata.

Quando si scrive un programma di questo genere vanno tenuti a mente alcuni accorgimenti. Il più importante è che bisogna evitare di alterare il contenuto di una cella prima che il suo valore sia stato controllato da tutte le altre celle a cui è adiacente. Il modo più semplice per soddisfare questa condizione è conservare due copie della matrice; il programma esamina una copia per stabilire lo stato attuale dell'intorno e immette il risultato del suo calcolo nell'altra copia. Vanno definite anche condizioni di limite. Idealmente la matrice dovrebbe essere infinita, ma la cosa è ovviamente impraticabile. Una tecnica comune consiste nel congiungere i bordi di una matrice, in modo che celle su bordi opposti diventino adiacenti. In una dimensione, una matrice di questo genere è topologicamente un cerchio e in due dimensioni è un toro; pur essendo finita, non ha confini.

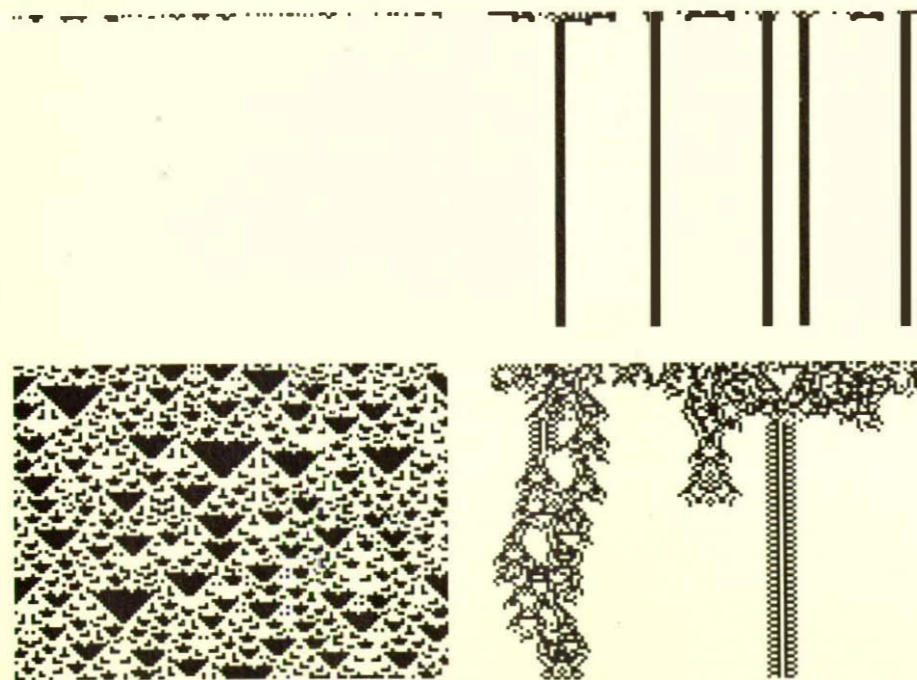
Un programma del tipo descritto sopra, che giri su un calcolatore digitale di uso generale, è una procedura sequenziale che simula le azioni di una schiera di molti calcolatori attivi simultaneamente. Molto meglio sarebbe avere davvero una rete di molteplici calcolatori con la struttura di una matrice cellulare. La costruzione di una macchina del genere non è affatto fuori questione: i singoli calcolatori sarebbero così semplici che ce ne starebbero molti su un unico chip di semiconduttore. Anche il fatto che solo calcolatori vicini hanno bisogno di comunicare l'uno con



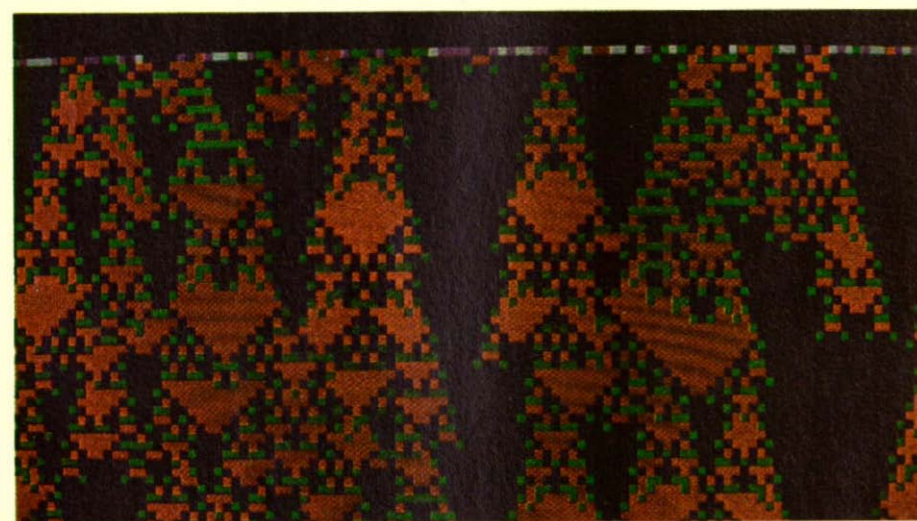
Algoritmo per automi cellulari basato su regole di transizione «di conto» o «totalistiche»



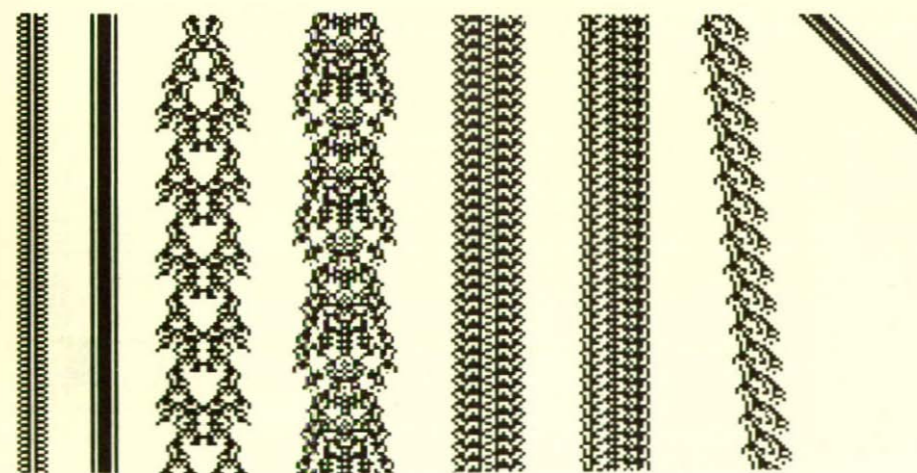
L'evoluzione di un automa cellulare secondo la regola dei due quinti



Le quattro classi di regole totalistiche in una dimensione



Stati successivi di un automa unidimensionale della classe 4



Alcuni componenti di un possibile calcolatore universale

l'altro ridurrebbe la complessità del congegno. Toffoli ha stimato che un simile elaboratore potrebbe operare un milione o forse anche un miliardo di volte più velocemente di un calcolatore di uso generale. Un lavoro preliminare su calcolatori di questo tipo è in corso al Massachusetts Institute of Technology e alla Thinking Machines Corporation di Waltham, Massachusetts.

Al posto di un chip specializzato, Toffoli ha costruito una macchina automa cellulare dedicata con normali componenti microelettronici. I calcoli sono effettuati in scritte, invece che per tutte le celle in una sola volta, ma, dato che è sintonizzato su un unico tipo di calcolo, il congegno è circa 1000 volte più veloce di un calcolatore di uso generale. La macchina in sé consiste di qualche scheda a circuiti stampati montata su un telaio; è collegata a una unità video a colori ed è controllata da un altro piccolo calcolatore, un Atari 800.

La macchina automa cellulare di Toffoli fornisce una matrice di 256 per 256 celle, ognuna delle quali può avere fino a 256 stati. Lo stato di ogni cella è ricalcolato 60 volte al secondo. Guardare un sistema che si evolve a questo ritmo è molto diverso dal guardare un congegno più lento. Invece di una successione di fotografie statiche si vede un film in movimento. Il gioco Vita non sembra più una maestosa progressione di configurazioni astratte; è più simile all'osservazione attraverso il microscopio di batteri e protozoi che nuotano, ruotano, procreano, mangiano e vengono mangiati.

Un automa cellulare a una dimensione richiede a un calcolatore molte meno risorse, sia spaziali sia temporali, di quante ne richieda un sistema a due dimensioni. Anche scrivere un programma per un sistema a una dimensione è più facile. La matrice lineare ha anche un altro vantaggio rispetto a quella planare: a causa della più semplice struttura geometrica, c'è maggiore possibilità di raggiungere una comprensione analitica dell'evoluzione dell'automata. E questo è proprio quanto ha cercato di fare, negli ultimi due anni, Stephen Wolfram dell'Institute for Advanced Study.

Una singola generazione di una matrice a una dimensione è semplicemente una fila di celle, ma si possono tracciare generazioni successive una di seguito all'altra. In questo modo si forma una matrice bidimensionale con un asse dello spazio e un asse del tempo e si può avere a colpo d'occhio l'intera evoluzione del sistema.

Wolfram ha scoperto che tutte le regole di transizione da lui studiate finora si possono inserire in quattro sole classi. La classe 1 è formata da quelle regole la cui evoluzione porta a uno stato stabile e omogeneo; per esempio, tutte le celle potrebbero assumere un valore pari a 1 o a 0. Le regole della classe 2 danno luogo a semplici strutture che possono essere stabili o periodiche, ma che in entrambi i casi rimangono isolate una dall'altra. Le regole della classe 3 creano configurazioni

Lecture da LE SCIENZE

edizione italiana di
SCIENTIFIC
AMERICAN

I GRANDI TEMI DELLA SCIENZA

I volumi della collana «Lecture» sono monografie dedicate ciascuna a un grande tema della scienza svolto attraverso i più importanti articoli scritti per «Le Scienze» dai maggiori protagonisti mondiali dei vari settori della ricerca e scelti da eminenti studiosi italiani.

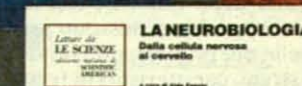
Elenco completo dei titoli disponibili:

- BIOLOGIA CELLULARE 1: Membrane e comunicazione (16 articoli, pp. 192)
- BIOLOGIA CELLULARE 2: Le basi della locomozione (14 articoli, pp. 150)
- BIOLOGIA CELLULARE 3: Il differenziamento (15 articoli, pp. 168)
- GENETICA (22 articoli, pp. 256)
- LA NEUROBIOLOGIA (22 articoli, pp. 272)
- STORIA NATURALE ED EVOLUZIONE (23 articoli, pp. 304)
- ZOOLOGIA 1: Comunicazione e comportamento sociale (19 articoli, pp. 190)
- ZOOLOGIA 2: Il mondo sensoriale (18 articoli, pp. 206)
- LE ORIGINI DELLA CIVILTÀ EUROPEA (20 articoli, pp. 232)
- L'ANTICO MEDITERRANEO (20 articoli, pp. 224)
- MEDICINA E SOCIETÀ (18 articoli, pp. 184)
- CONTRO LA FAME: L'alimentazione nel mondo (21 articoli, pp. 220)
- ARMI, STRATEGIE E DISARMO (26 articoli, pp. 318)
- DALLA SELCE ALL'ELETTRONICA: Il cammino della tecnologia (22 articoli, pp. 228)
- CHIMICA D'OGGI (22 articoli, pp. 264)
- LE PARTICELLE FONDAMENTALI: Quark, leptoni e l'unificazione delle forze (21 articoli, pp. 304)
- LA DINAMICA DELLA TERRA (24 articoli, pp. 302)
- I PIANETI DELLA STELLA SOLE (25 articoli, pp. 272)
- LE NUOVE FRONTIERE DELL'ASTROFISICA (19 articoli, pp. 240)
- RELATIVITÀ E COSMOLOGIA (11 articoli, pp. 134)

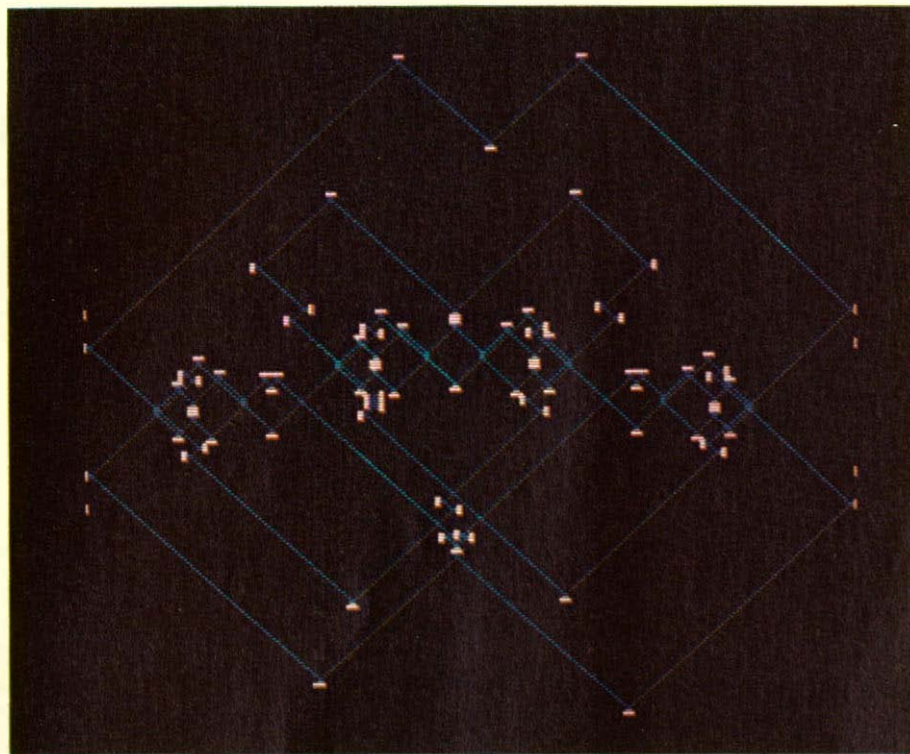
Per altre informazioni, si veda l'ultima pagina di questo fascicolo.

Questi volumi sono in vendita nelle principali librerie (distribuzione: La Nuova Italia Editrice) o si possono ordinare direttamente all'editore:

LE SCIENZE S.p.A.
Ufficio Vendite
Via del Lauro 14
20121 Milano



Formato: 21 x 29 cm.



Il calcolatore a palle di biliardo in azione

caotiche, anche se non casuali. Nella classe 4 ci sono le poche regole di transizione che generano strutture di sostanziale complessità spaziale e temporale.

Wolfram suppone che gli automi cellulari a una dimensione possono essere i più semplici sistemi ben definiti capaci di un comportamento complesso di autoorganizzazione. In natura molti sistemi dinamici continui hanno questa capacità: evolvono da uno stato casuale di partenza a una struttura altamente ordinata. (Un esempio è il fiocco di neve.) L'evoluzione può essere spiegata in termini di attrattori, che sembrano trascinare il sistema verso un sottoinsieme di tutte le possibili configurazioni.

È stato tracciato un parallelo tra le classi di automi cellulari e i tipi di attrattori che si osservano nei sistemi fisici. Un automa della classe 1 è analogo a un sistema continuo con il più semplice degli attrattori: un punto limite che invariabilmente porta il sistema allo stesso stato finale. L'evoluzione di un automa della classe 2 è invece simile a quella di un sistema con un ciclo limite, un insieme di configurazioni che si ripetono indefinitamente.

Gli automi della classe 3, con le loro disposizioni caotiche, si possono associare con quelle più interessanti entità dette attrattori strani, caratteristici di fenomeni fisici quali l'inizio di correnti di turbolenza. In un sistema governato da un attrattore strano, l'evoluzione procede verso un sottoinsieme di tutte le configurazioni possibili, ma il sottoinsieme può avere una struttura incredibilmente complicata. Quando si visualizza l'insieme come matrice di punti nello spazio, si tratta in molti casi di un frattale, una figura geometrica

con un numero frazionario di dimensioni.

Le distinzioni tra le classi di automi possono essere chiarite prendendo in considerazione un semplice esperimento. Supponiamo di avviare un automa cellulare in qualche configurazione iniziale scelta a caso e di farlo evolvere per molti passi nel tempo; si prenda nota dello stato finale. Si ritorni ora alla configurazione di partenza, ai cambi di valore di una singola cella e si faccia evolvere il sistema per lo stesso numero di passi. Che effetto avrà il piccolo cambiamento sullo stato finale? Per un automa della classe 1 non c'è alcuna conseguenza: un sistema della classe 1 raggiunge lo stesso stato finale indipendentemente dallo stato iniziale. Un automa della classe 2 può mostrare qualche effetto, ma limitato a una piccola area vicino al sito in cui è avvenuto il cambiamento. In un sistema della classe 3, invece, l'alterazione di una singola cella può provocare un disturbo che si propaga lungo tutta la matrice.

Le regole della classe 4 sono le più rare e le più interessanti. Alcune funzioni di transizione piuttosto semplici ricadono in questa classe; per esempio, nell'intorno definito in modo da includere la cella centrale e le due celle che le stanno a lato, la regola secondo cui la cella centrale è 1 se due o quattro celle dell'intorno sono 1 porta a configurazioni della classe 4. La sensibilità a piccole variazioni nelle condizioni iniziali è ancora maggiore nella classe 4 che nella classe 3. Si ritiene che per prevedere lo stato futuro di un automa della classe 4 non vi sia nessuna procedura generale più efficace di quella che consiste nel lasciare all'automato stesso il compito di calcolare lo stato.

Una congettura correlata alla precedente ha una portata ancora più estesa e suggerisce che gli automi della classe 4 si possano considerare dei calcolatori universali. La macchina di Turing è il più familiare tra i dispositivi di questo tipo; se una funzione può essere calcolata, è presumibile che una macchina di Turing possa farlo. Si può dimostrare che altri calcolatori sono universali mostrando la loro equivalenza a una macchina di Turing. Di numerosi automi cellulari a due dimensioni (incluso il gioco Vita) si è dimostrato che sono universali e una dimostrazione è stata fornita anche per un complicato sistema a una dimensione con 18 stati per cella. Gli automi della classe 4 sarebbero i più semplici calcolatori universali conosciuti. Si è individuata la maggior parte dei componenti essenziali. Un importante elemento mancante è un orologio: una struttura che emette una serie di impulsi a intervalli regolari, come il cannone ad alianti nel gioco Vita.

Considerare gli automi cellulari come calcolatori fa pensare che il loro comportamento di autoorganizzazione può essere caratterizzato nei termini delle loro capacità di calcolo. Così, per esempio, insiemi di configurazioni generati dall'evoluzione di un automa cellulare possono essere pensati come un linguaggio formale. Ogni configurazione è considerata una parola del linguaggio, formata da una sequenza di simboli che rappresentano i valori dei siti dell'automata cellulare secondo un insieme di regole grammaticali. Wolfram ha dimostrato che la configurazione generata da qualsiasi automa cellulare dopo un tempo finito può essere descritta da una semplice classe di linguaggi formali noti come linguaggi regolari. Per qualsiasi di questi linguaggi regolari è possibile trovare una semplicissima grammatica. Tale grammatica fornisce una descrizione minimale delle configurazioni dell'automata cellulare e si può assumere la sua dimensione come misura della complessità delle configurazioni. Per automi cellulari delle classi 1 e 2, la complessità tende a un limite finito per tempi lunghi, quindi le strutture generate da quei sistemi sono descritte da linguaggi regolari. Per automi cellulari delle classi 3 e 4, invece, la complessità di solito cresce rapidamente col tempo e appare evidente la necessità di linguaggi formali più complicati per descrivere il comportamento sul lungo periodo di tali sistemi.

Esiste una classe particolare di automi cellulari detti reversibili, o invertibili. Da qualsiasi configurazione di partenza, un automa reversibile che sia fatto evolvere per un qualsiasi numero di passi, poi fermato e fatto girare all'inverso, tornerà esattamente allo stato iniziale. Le configurazioni formate da un automa reversibile tipico hanno un aspetto qualitativamente differente rispetto alle configurazioni caratteristiche di un automa non reversibile. In particolare, se la configurazione iniziale è casuale, essa tende a rimanere casuale: non compare nessuna struttura di autoorganizzazione.

Una condizione necessaria per la reversibilità è che la regola di transizione sia deterministica tanto in avanti quanto all'indietro: ogni possibile stato di un intorno, cioè, deve avere sia un unico successore sia un unico predecessore. Il gioco Vita è non reversibile perché il predecessore di uno stato non può essere individuato senza ambiguità: se, per esempio, una cella è attualmente «morta», nella generazione precedente avrebbe potuto avere un numero qualsiasi di adiacenti vivi diverso da tre. Un modo sistematico per creare regole di transizione reversibili è stato inventato da Fredkin e in seguito approfondito da Margolus. L'essenza del metodo sta nel far dipendere lo stato successivo di una cella dai due stati precedenti dell'intorno. Lo stato al momento $t+1$ è dato da una funzione qualsiasi dell'intorno al momento t meno lo stato al momento $t-1$. L'inverso è allora chiaro: lo stato al momento $t-1$ deve essere dato dallo stato al momento t meno lo stato al momento $t+1$.

In ragione della condizione di determinismo bidirezionale, non ci può essere alcun attrattore nell'evoluzione di un automa reversibile. La presenza di un attrattore implica che molti stati iniziali si evolvono lungo percorsi che si fondono l'uno con l'altro; nell'evoluzione rovesciata, i punti di fusione diventerebbero punti di diramazione e il determinismo verrebbe meno. Analogamente, un automa cellulare reversibile non può mai entrare o uscire da un *loop*, ossia un ciclo di stati, perché anche in questo caso si produrrebbe un punto di diramazione in una direzione o nell'altra. Vista l'esclusione di attrattori e delle configurazioni di autoorganizzazione associate, potrebbe sembrare che le regole di transizione reversibili diano luogo ad automi cellulari piuttosto stupidi, ma vi sono in compenso altre

santi questi sistemi. Il fatto più notevole è che il contenuto informativo di una configurazione di celle in un automa reversibile è una grandezza che si conserva (non può aumentare o diminuire nel corso dell'evoluzione dell'automato). Questa proprietà rende i sistemi reversibili dei validi modelli di calcolo.

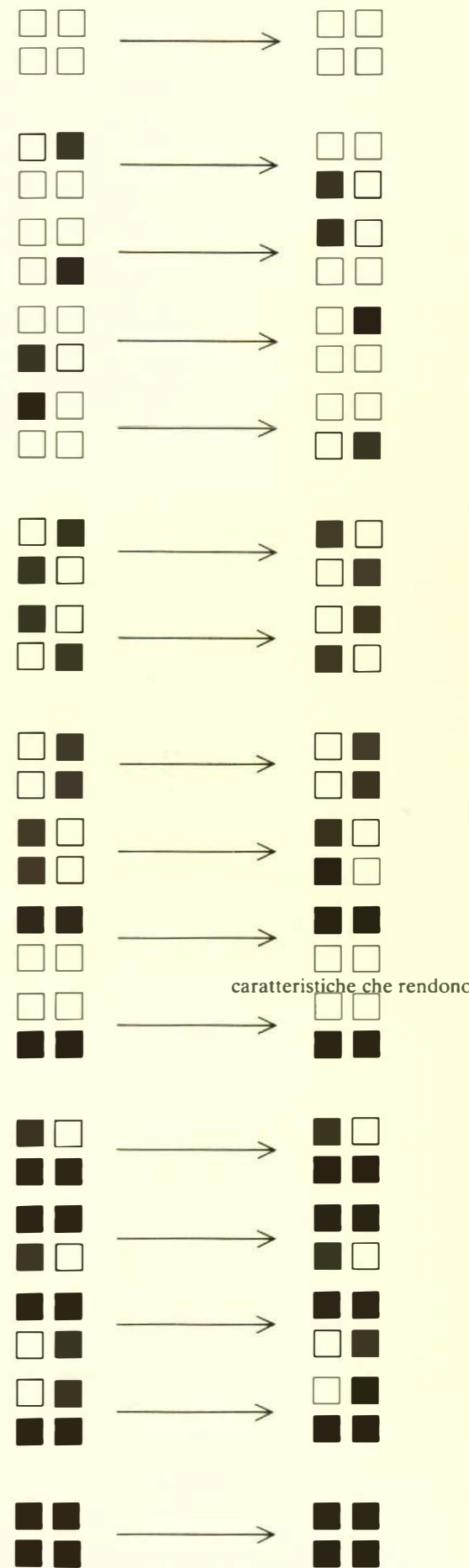
Margolus ha costruito un calcolatore automa cellulare basato su un immaginario sistema meccanico studiato per la prima volta da Fredkin: il modello del calcolo a palle da biliardo. Nel modello, i bit di informazione (1 e 0) sono trasportati da ideali palle da biliardo che si muovono senza attrito e rimbalzano con perfetta elasticità l'una contro l'altra e contro altri ostacoli. La presenza di una palla in una posizione designata rappresenta un 1 binario e l'assenza di una palla rappresenta uno 0 binario. Con dei respingenti disposti in modo opportuno è possibile creare porte logiche analoghe a quelle di un calcolatore elettronico. In una porta AND, per esempio, una palla da biliardo passa attraverso la regione di uscita (e quindi registra un 1 binario) solo se due palle si avvicinano simultaneamente alla porta lungo specifiche traiettorie.

La versione del modello a palle da biliardo proposta da Margolus, basata su un automa cellulare, è un esempio di regola reversibile di transizione semplice ma piuttosto inconsueta. Le celle non sono considerate individualmente ma in blocchi di quattro: ogni possibile configurazione all'interno di un blocco è trasformata in un'unica configurazione prodotta. La regola è fatta in modo che un unico 1 su uno sfondo di 0 si propaghi lungo una delle quattro direzioni diagonali del reticolo a una velocità di una cella per passo; l'1 isolato è l'incarnazione di una palla da biliardo. Un blocco continuo di quattro 1 rimane immutato e agisce come perfetto riflettore. Quando si fa girare il modello sulla macchina automa cellulare di Toffoli, le «palle da biliardo» si muovono velocemente lungo lo schermo in complesse configurazioni intrecciate. Guardando questo movimento ordinato (anche se frenetico), è difficile tenere a mente che il programma non possiede nessuna rappresentazione dei percorsi delle palle ma applica semplicemente un'unica regola a tutte le celle.

Il modello a palle da biliardo e la sua realizzazione con l'automata cellulare hanno un risvolto importante per la teoria del calcolo. È stata avanzata la congettura che qualsiasi calcolatore debba avere componenti che dissipano sia energia sia informazione; secondo questa ipotesi, c'è un limite termodinamico al rendimento di un calcolatore, proprio come c'è un limite al rendimento di una macchina termica. Le perdite di informazione e di energia che si suppongono inevitabili dipendono direttamente dall'irreversibilità del processo di calcolo. (Quando un calcolatore somma i numeri 5 e 3 per ottenere 8, la procedura non può essere rovesciata perché ci sono infiniti numeri che, sommati, avrebbero potuto dare lo stesso risultato.)

Fredkin, Toffoli e Margolus affermano che il modello a palle da biliardo offre un argomento contrario all'idea della dissipazione inevitabile. Nel calcolatore a palle da biliardo non si perde alcuna informazione. In effetti, le stesse palle da biliardo non si possono creare né distruggere e tutta l'informazione che definisce la loro disposizione iniziale viene conservata pur con tutto l'evolversi del sistema. Gli ingressi per un'operazione di addizione possono essere recuperati semplicemente invertendo le traiettorie. In linea di principio, il calcolatore a palle da biliardo potrebbe operare senza alcun consumo interno di energia.

Il collegamento tra fisica e calcolo automatico è stato tracciato con particolare chiarezza da Toffoli in una affermazione che si potrebbe leggere come una descrizione del più grande degli automi cellulari. «In un certo senso - scrive - la natura ha continuato a calcolare lo "stato successivo" dell'universo per miliardi di anni; tutto quello che dobbiamo fare - e, in effetti, tutto quello che possiamo fare - è "saltare in groppa" a questo enorme calcolo in atto e cercare di scoprire quali parti di esso si avvicinano a dove noi vogliamo andare.»



Regola di transizione per un calcolatore a palle di biliardo