COLLEGIO SUPERIORE
DELL'UNIVERSITÀ DI BOLOGNA

CHIARA CORTESE, TOMMASO FADDA, KEVIN M. FRICK
STEFANO PULIGHEDDU, SIMONE TENTORI

ASTRAZIONI STENOGRAFICHE

CONCETTI CHIAVE PER VIVERE CONSAPEVOLMENTE LA NOSTRA SOCIETÀ

COLLEGIO.UNIBO.IT

Bononia
University Press

ASTRAZIONI STENOGRAFICHE

I

SUPERVISIONE
BEATRICE FRABONI

CURATORE
MATTEO CERRI

CHIARA CORTESE, TOMMASO FADDA, KEVIN M. FRICK
STEFANO PULIGHEDDU, SIMONE TENTORI

ASTRAZIONI STENOGRAFICHE

CONCETTI CHIAVE PER VIVERE CONSAPEVOLMENTE LA NOSTRA SOCIETÀ

COLLEGIO.UNIBO.IT

Bononia
University Press

Bononia University Press
Via Foscolo 7
40123 Bologna
tel. (+39) 051 232882
fax (+39) 051 221019

ISBN 978-88-6923-676-1
ISBN online 978-88-6923-677-8

www.buponline.com
e-mail: info@buponline.com

Progetto grafico: Alessio Bonizzato

Layout e impaginazione: Design People (Bologna)

Prima edizione: novembre 2020

SOMMARIO

PREFAZIONE

Nel 2007 James R. Flynn fa emergere in un suo libro un nuovo concetto: l'astrazione stenografica. Flynn è uno psicologo dell'intelligenza, famoso per l'effetto che porta il suo nome e che descrive l'aumento progressivo che il quoziente intellettivo ha avuto nel corso dei decenni del secolo scorso; con l'idea delle astrazioni stenografiche, Flynn si rivolge al mondo dell'educazione. Le astrazioni stenografiche sono infatti dei concetti elementari, semplici ma essenziali, che permettono a un cittadino moderno di vivere appieno la società contemporanea ma che possono non aver fatto parte del cammino scolastico di ciascuno di noi. Originariamente Flynn ne identifica tredici, che rapidamente diventano venti, ma questo è, per sua stessa ammissione, solo un primo traguardo. Spetta a tutti i ricercatori declinare le astrazioni stenografiche per i diversi settori del sapere.

L'idea di Flynn è molto seducente, specialmente nei tempi che stiamo vivendo: se da un lato la nostra vita viene resa più semplice dalla tecnologia, dall'altro la quantità di nozioni e di concetti che dobbiamo possedere per orientarci in questo mondo diventa sempre più grande. Da questi spunti è quindi nato un Corso d'insegnamento presso un'istituzione d'eccellenza, il Collegio Superiore dell'Università di Bologna, che si proponeva di esplorare con gli studenti alcune astrazioni stenografiche nuove. Un'idea altamente sperimentale, che, fortunatamente, può ancora essere implementata in queste rare isole di libertà didattica che sono le scuole d'eccellenza. Il corso si è concentrato su tre materie: fisica, fisiologia ed economia e ha chiesto agli studenti che vi hanno partecipato di preparare un testo esplicativo sull'astrazione stenografica da cui fossero stati colpiti maggiormente.

Il risultato è andato ben oltre le aspettative iniziali. Per questo abbiamo deciso di spingere la nostra sperimentazione ancora più avanti pubblicando cinque dei testi prodotti dagli studenti: cinque racconti di scienza che ci spiegano un concetto elementare; qualcosa di nuovo, la cui novità non è nell'interpretazione guidata dall'esperienza, propria di un ricercatore anziano, ma quella vista con gli occhi di chi di questa società si appresta a diventare un membro attivo. Il domani, in fondo, si trova sempre vicino ai più giovani. Da queste iniziative emerge anche quanto sia elevata la preparazione culturale e la passione per la scienza degli studenti. Chi

denigra e sottovaluta i giovani non fa altro che avvelenare la sorgente da cui dovrà bere nel prossimo futuro.

È quindi con un non velato ottimismo che abbiamo deciso di non lasciare che la vita di questo libro si esaurisca in una lettura compiuta ma di trasformarlo nel primo volume di una collana. Astrazioni Stenografiche si propone di portare tutti gli anni nuovi racconti di idee e di concetti che possano essere utili alla nostra conoscenza, e dare una piccola scossa al nostro ottimismo: prepariamo insieme le mani che guideranno il nostro futuro.

Beatrice Fraboni
Direttrice del Collegio Superiore

Matteo Cerri
Curatore della collana

LA PERSISTENZA DEL FANTASMA

STEFANO PULIGHEDDU

La Sindrome dell'Arto Fantasma rappresenta una patologia peculiare che ha messo in difficoltà generazioni di medici, restii a credere alla possibilità che un arto non più presente potesse mantenersi nella coscienza dell'invalido e persino dar dolore – come fosse un vero e proprio fantasma. Con lo sviluppo della fisiologia del Novecento ed in particolare grazie ai progressi di fine secolo si è riusciti ad avere una visione articolata sul tema, proponendo diverse teorie tra cui la riorganizzazione delle fibre corticali. Per quanto riguarda le possibilità di terapia, ancora non vi è un trattamento particolarmente efficace; comunque, l'intervento più promettente pare essere la *mirror therapy*, che ha rivoluzionato il panorama terapeutico precedente.

Phantom Limb Syndrome is an interesting condition that has puzzled generations of doctors reluctant to believe the possibility that a limb no longer present could still cause pain—as if it were a real ghost. As the twentieth century brought significant developments to physiological knowledge, and thanks to the scientific progress made at the end of the century, it was possible to develop an extensive understanding of the subject, proposing various theories like Cortical Remapping Theory. Regarding the possible therapies to date, there is not yet an effective first-line treatment for phantom limb pain; however, the most promising intervention seems to be Mirror Therapy, which has completely changed the previous therapeutic panorama.

PREFAZIONE

Dove si trova il dolore? Sarebbe facile rispondere che il nostro corpo sia la sede di questa sensazione assolutamente spiacevole, così come controintuitivamente necessaria. In fondo, ci facciamo sempre male a qualcosa: «ho mal di schiena», dirà l'anziano; «ho mal di pancia», dirà il bambino. Ai nostri occhi, il dolore si impossessa di una parte del nostro corpo, spesso dimenticata, e ce la fa odiare. Perché a prima vista il dolore è un male; un nemico esiziale dell'umanità.

In realtà, il dolore è qualcosa di molto più complesso e, come spesso succede, potremmo scoprire di non poterne fare a meno. Ci sono infatti malattie rare per le quali le persone non sentono dolore. Fatto che accorcia la loro vita media. Come potrebbero infatti rendersi conto che il fuoco scotti? Immaginare di vivere senza poter provare dolore è difficile tanto quanto cercare di spiegare a queste persone cosa sia il dolore stesso. Eppure, nel nostro corpo, ci sono zone immuni al dolore. Il fegato, per esempio. O il rene. Organi che non possono essere conquistati dal dolore e che, proprio per questo, possono mettere a repentaglio la nostra vita. Quando un tumore li colpisce, noi siamo gli ultimi a saperlo.

Ma esiste anche un altro organo che, se da un lato non può provare dolore, dall'altro ne è la vera sede; la vera origine. Il cervello. È qui che le informazioni portate dai sensi si trasformano in dolore. Quest'ultimo è infatti più di senso: è un modo di essere. Utile, certamente. Ma anche parecchio difficile da vivere. Basti pensare che, al contrario di tutte le altre esperienze sensoriali, non possiamo rivivere il dolore. La nostra memoria non ce lo permette. E non possiamo neanche morire di dolore, anche se quest'ultima osservazione, riflettendoci bene, potrebbe non essere proprio una consolazione.

Perché non c'è niente che rifuggiamo di più del dolore. Lo rifugiamo così tanto e ci sono dolori così intensi, che a volte saremmo tentati di togliere al nostro corpo la parte che duole. Potremmo illuderci che, escindendo quella che, erroneamente, pensiamo essere la sede del dolore, il dolore ci abbandoni e ci lasci riposare. Purtroppo però, il prezzo che paghiamo per l'utilità del dolore è in termini di presenza: il dolore infatti è in grado di restare sempre con noi. Si chiama dolore fantasma. Uno dei fenomeni più curiosi della fisiologia, per cui una persona ha male ad una parte del corpo che non ha più. Un mistero che ha permesso alla ricerca di addentrarsi in profondità nel cammino della comprensione del dolore. Dal quale emerge che non possiamo allontanarci dal dolore: esso è ineluttabile perché esse si trova nello stesso posto dove siamo noi: nel nostro cervello.

In questo capitolo, Stefano Puligheddu ci racconta la storia dell'arto fantasma consentendoci di capire come questo apparentemente mistico fenomeno, sia invece incernierato nella realtà. E capire meglio il dolore, è forse il primo passo che possiamo fare per non averne più paura.

Matteo Cerri
Ricercatore di Fisiologia
Dipartimento di Scienze Biomediche e Neuromotorie
Università di Bologna

I. IL RACCONTO DEL VETERANO

Nel luglio del 1866 un breve articolo dal titolo *The Case of George Deadlow* fa la sua apparizione sull'Atlantic Monthly[1], famoso mensile statunitense fondato pochi anni prima (nel 1857) che si occupa ancor oggi di cultura, letteratura, salute ed altri svariati temi. Il pezzo, oggi consultabile dagli archivi online dell'Atlantic, inizia così:

> *The following notes of my own case have been declined on various pretexts by every medical journal to which I have offered them.*

L'incipit enigmatico preannuncia al lettore quella sfumatura di incredulità che accompagna il fenomeno che si accinge a narrare. Comincia dunque a delinearsi la curiosa storia di George Deadlow, veterano della guerra civile che in questa aveva perso tutti e quattro gli arti. Entrambe le gambe ed entrambe le braccia, a causa di ferite da pallottole, esplosioni o cancrene che non avevano lasciato spazio ad altre soluzioni per la medicina della seconda metà dell'Ottocento se non l'amputazione, il taglio netto. Ritrovatosi ad essere un "torso", racconta di come abbia trovato distrazione, quando non diletto ed interesse, nell'indagare le sue sensazioni e quelle dei pazienti nella sua stessa condizione mentre si trovava in lunga degenza in un ospedale per veterani a Philadelphia.

Egli si mette dunque ad osservare con attenzione, e nota che le persone che hanno subito un'amputazione conservano per molti mesi l'usuale sensazione di possedere ancora l'arto. Esso dà prurito o dolore, a volte tanto forte da credere di avere un crampo, ma non viene mai percepito freddo o caldo. Questo dolore sembrerebbe inoltre direttamente proporzionale alla persistenza della sensazione di aver ancora l'arto attaccato al corpo: qualora non vi sia sofferenza correlata, la percezione immaginaria della propaggine da tempo rimossa spesso diminuisce gradualmente, fino a sparire interamente.

George elabora quindi una sua spiegazione del fenomeno: i nervi, attraverso cui le impressioni esterne sono captate e trasmesse al midollo spinale ed al cervello, pur se troncati rimangono capaci di essere stimolati, ad esempio da irritazioni. Sono così indotti a trasmettere dei segnali al cervello, che abituato normalmente a riferirli alle parti ormai perse, genera delle sensazioni ingannevoli. Il nervo è come un *bell wire*, il filo di una campana che può essere tirato in qualsiasi punto lungo il suo decorso

[1] https://www.theatlantic.com/magazine/archive/1866/07/the-case-of-george-dedlow/308771/

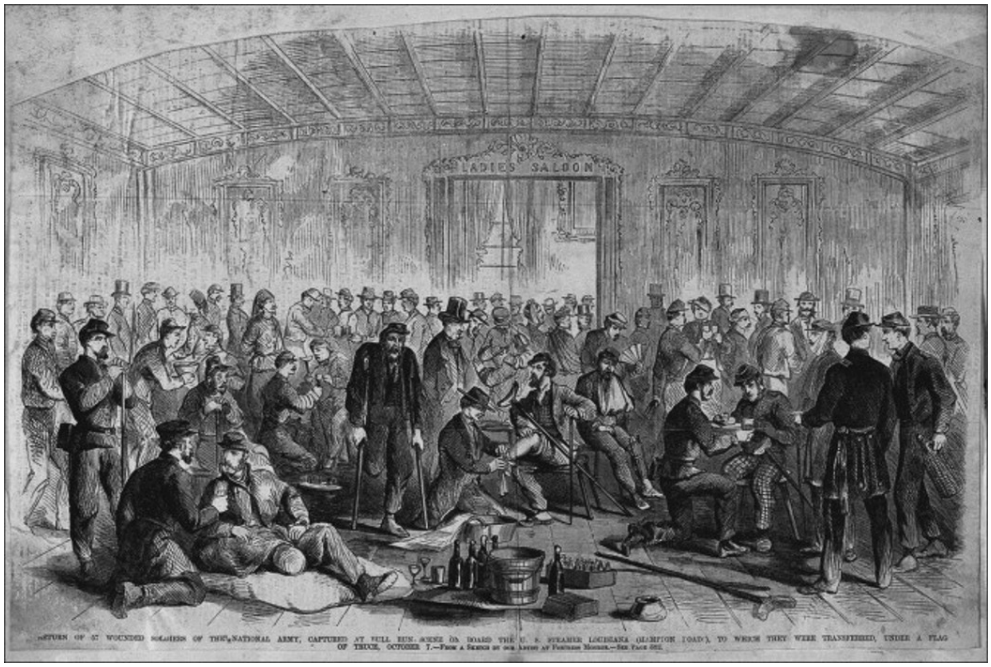


FIG. 1 *Return of wounded Confederate prisoners, under a flag of truce, during the American Civil War*, incisione su legno, Wellcome Collection. Attribution 4.0 International (CC BY 4.0).

senza alterarne il risultato, ovvero il rintocco. Le stimolazioni del nervo troncato sono spesso causate da cambiamenti nel moncone durante la guarigione: e dunque come questo cicatrizza, essi cessano e la stimolazione scompare. Però questo processo non è così omogeneo e lineare; a volte la sensibilità si può perdere, riacquistare, riperdere e via: il mutilato dirà in certi momenti «Posso sentire il mio pollice» e in altri «Ecco che sento il mio mignolo». Inoltre, aggiunge, quasi ogni persona che ha perso l'arto superiore oltre l'avambraccio lo sente come se fosse piegato, percepisce la mano assai vicina al gomito (una sorta di avambraccio corto) e spesso avverte le dita fortemente flesse. Per quanto riguarda l'arto inferiore, essi avvertono la presenza del piede ma gli pare che la gamba sia accorciata; se la ferita arriva alla coscia, gli sembra che il piede sia giusto ad una decina di centimetri dal ginocchio.

La storia prosegue narrando l'incontro di George con un medium singolare e l'invito da parte di questo a partecipare ad una seduta spiritica. Egli racconta dunque che durante il circolo esoterico vengono evocati dei fantasmi, i quali (con sorpresa generale) comunicano tramite il medium una strana presentazione di sé: «UNITED STATES ARMY MEDICAL MUSEUM, NOS. 3486, 3487». Ed ecco che Deadlow si alza sui monconi riconoscendo i codici di classificazione mu-

seali delle sue gambe, ormai sotto formaldeide da mesi; ma colto dall'eccitazione e dal dolore dei monconi incapaci di reggerlo, perde conoscenza.

Una vicenda decisamente intrigante, che si muove nel contesto degli Stati Uniti dilaniati dalla guerra di secessione. Il racconto è dettagliato e realistico, eppure puramente fittizio: George Deadlow non è mai esistito. La sua figura rappresenta la realtà di tantissimi individui, partiti verso il fronte del conflitto tra gli Stati Uniti e la Confederazione, che in seguito alle violenze delle battaglie e alle difficili condizioni igieniche si ritrovavano con ferite terribili, spaventose. Gli avanzamenti della tecnologia bellica non andarono di pari passo con lo sviluppo della medicina americana, di modo che i medici militari si trovarono impreparati di fronte alle terribili lacerazioni che gli si presentavano, in un tempo in cui la sterilizzazione era ritenuta inutile (se non persino dannosa) e l'anestesia un lusso raro. Abituati spesso solamente ad estrarre denti ed incidere ascessi, davanti a situazioni disastrose come arti anneriti dalla cancrena, molli e crepitanti, non potevano far altro che ricorrere all'estrema soluzione dell'amputazione, nella speranza di evitare ai soldati una morte incombente. Incalzati dalla fila di vittime che gemeva dal dolore, tiravano fuori la sega circolare e chiedevano aiuto per immobilizzare il paziente; spiegavano velocemente ciò che intendevano fare a soldati spesso giovani e dal viso terrorizzato, prima di accingersi a recidere con violenza il più velocemente possibile.

Quando il conflitto finalmente terminò, dopo quattro anni di massacri, la nazione americana dovette confrontarsi con quest'enorme massa di mutilati che non riuscivano a reinserirsi nella società: si stima che circa 60.000 uomini subirono l'amputazione di almeno un arto durante la guerra. Incapaci di lavorare nelle fattorie, marginalizzati e depressi, si ritrovavano sperduti e privati di un futuro. Si confrontavano con l'ostentato ringraziamento della nazione; che se da un lato li immortalava come eroi, dall'altro li lasciava abbandonati alle loro difficoltà, nel tentativo di lasciarsi alle spalle la memoria di un duro conflitto. Solo un qualche centinaio fece domanda per ottenere un arto artificiale, che spesso consisteva di una protesi scomoda e poco efficiente; gli altri dovettero imparare a far fronte alla vita quotidiana con l'uso degli arti restanti, e l'immagine tipica del veterano mutilato divenne quella del mendicante.

Ecco quindi che diverse persone scambiarono il racconto per reale: fioccarono le donazioni per l'Ospedale dei Mutilati di Philadelphia e alcune persone vi si recarono nel tentativo di conoscere di persona il protagonista, il quale invece non era mai esistito. Si trattava di un personaggio scaturito dagli innumerevoli feriti che furono in cura dal dottor Silas Weir Mitchell, vero autore dell'articolo e medico-chirurgo in una struttura creata dall'Army Medical Department per dedicarsi alle malattie del Sistema Nervoso, il Turner's Lane Hospital.

Mitchell crebbe a Philadelphia ereditando dal padre, medico anche lui, la passione per la materia: si iscrisse alla facoltà di medicina alla Thomas Jefferson University di Philadelphia, dove ottenne il titolo di Medical Doctor nel 1850. Cominciò la sua carriera studiando il veleno del serpente a sonagli, ma con lo scoppio della guerra civile nel 1861 la sua carriera cambiò direzione: prese incarico come chirurgo a contratto al Turner's Lane Hospital, dove si specializzò nelle malattie nervose. Qui curò e studiò un gran numero di pazienti con varie patologie e sindromi neurologiche, occupandosi in particolare delle sintomatologie da dolori neuropatici cronici. Egli scrisse la storia di George Deadlow per rappresentare quell'insolita condizione che affliggeva i mutilati – quella persistenza che ricorda un fantasma vero e proprio – che aveva incontrato ripetutamente tra i suoi pazienti. Dopo aver pubblicato questo racconto anonimamente, Mitchell lavorò sui suoi materiali clinici pubblicando diversi articoli per riviste di medicina e libri di neurologia, tra cui *Injuries of Nerves and Their Consequences* nel 1872. Costruì così la sua straordinaria carriera, per la quale sarebbe poi stato ricordato come «Padre della moderna neurologia Americana», cominciando col dolore fantasma che aleggiava sul disastro della Guerra Civile.

2. UNO SGUARDO PIÙ ACCURATO

La Sindrome dell'Arto Fantasma è oggi definita come la condizione nella quale un soggetto percepisce delle sensazioni da un arto che non esiste più. Alcuni pazienti riportano di sentire ancora presente l'arto nella sua interezza, possono descrivere la posizione in cui si trova e muoverlo attorno, finanche eseguire dei compiti specifici. Alcune volte i soggetti percepiscono una fede presente nel dito amputato o un orologio che usavano indossare; altri possono sentire calore o freddo (contrariamente a quanto riportava George Deadlow), sensazioni di bagnato o prurito, pressione. Spesso compaiono, anche in combinazione, dolore, formicolii o parestesie. A complicare il quadro, l'insorgenza di queste percezioni può essere indotta da un'ampia varietà di *triggers*, di eventi scatenanti: ad esempio, il fumo, cambiamenti della pressione barometrica, l'esposizione al freddo, il tocco ed i rapporti sessuali. Col passare del tempo, il numero dei soggetti affetti dalla sindrome decresce e la sensazione dell'arto fantasma a volte scompare: tuttavia, uno studio[2] del 1984 concluse che

[2] R. Sherman, C.J. Sherman, L. Parker, *Chronic phantom and stump pain among American veterans: Results of a survey*, in Pain; 1984, 18, pp. 83-95.

più del 70% dei pazienti continuava ad esserne affetto per più 25 anni. Tutto ciò ha una grande importanza quando si considera che queste sensazioni, ed in particolare il dolore cronico suscitato, possono inficiare notevolmente la qualità di vita delle persone che ne soffrono. Dopo l'amputazione, meno della metà riescono a ritornare al proprio lavoro, e il dolore dell'arto fantasma complica l'utilizzo di arti prostetici: da ciò si capisce che l'obiettivo primario del trattamento è la riduzione (se non abolizione) del dolore, per poter tornare a condurre una vita il più possibile simile a quella precedente all'operazione.

Sensazioni di parti del corpo che permangono in seguito alla loro rimozione non riguardano soltanto gli arti, ma possono in realtà ritrovarsi anche in seguito all'estrazione di un dente, ad una mastectomia, o successivamente alla perdita di un occhio. In particolare quest'ultimo caso è assai interessante e prende il nome di "Sindrome dell'occhio fantasma". Si può presentare anche in questo caso dolore, ma in più sono presenti delle allucinazioni visive che consistono principalmente in percezioni di base (forme, colori); al contrario, le allucinazioni visive causate da una perdita ingente ma non integrale delle funzioni visive sono meno comuni (insorgono solo in un paziente su dieci) e sono spesso rappresentate da immagini dettagliate. L'insorgenza della sindrome fantasma è ad ogni modo meno frequente in relazione all'occhio (30% dei pazienti) rispetto all'arto, per il quale la percentuale di pazienti che la sviluppano si aggira tra 50% e 78%.

2.1 LA SENSIBILITÀ

Naturalmente, la responsabilità maggiore nel determinare questa sindrome appartiene al nostro sistema di percezione sensoriale, un complicato intrico di recettori e neuroni che fatichiamo ancora a comprendere del tutto: meglio, si potrebbe forse dire che ne siamo ben lontani.

Viviamo immersi in un ambiente dinamico, complesso, che ci bombarda di segnali e stimoli in un frenetico susseguirsi di micro-eventi. In ogni momento, il nostro corpo sta straordinariamente raccogliendo decine di migliaia di stimoli attorno e dentro di noi, li sta filtrando per consentirci di non distrarci dalla nostra attività e li sta raggruppando, per permetterci di prenderne coscienza senza perderci in ogni microscopica informazione. La sensibilità, se ci si riflette, è spettacolare. Innata in ogni individuo, svolge un lavoro che le nostre migliori reti neurali artificiali faticano anche solo a minimamente tentare d'imitare; il nostro occhio è un modello a cui guardano centinaia di team di ricerca sparsi per il mondo per riuscire a comprendere e copiare la sua struttura, ad esempio nelle macchine

fotografiche (l'occhio umano ha una risoluzione di 576 Megapixel, una Nikon D5600 invece "solo" di 24,2). Eppure, per quanto sia immediata l'intuizione di cosa sia la sensibilità, il suo studio è decisamente più vasto e complicato: la classica divisione da libro di scuola nei famosi cinque sensi (tatto, udito, gusto, vista e olfatto) è conosciuta da tutti, ma la scienza è andata in realtà ben oltre.

Il nostro sistema nervoso ha bisogno per poter agire efficacemente di ricevere informazioni relative a ciò che accade all'interno ed al di fuori del corpo. A tal scopo necessita un sistema di "traduzione" di questi stimoli, assai variegati: gli serve un intermediario capace di trasformarli nel suo linguaggio fatto di impulsi elettrici. Questa sorta di "interprete" non è altro che l'insieme dei recettori sensoriali, cellule specializzate in tale ruolo. La sensibilità può essere considerata come divisa in due macrocategorie: una sensibilità meccanocettiva ed una sensibilità termo-dolorifica. La prima comprende la sensibilità tatto-pressoria, vibrazionale, distensiva-viscerale (immaginiamo ad esempio la sensazione di stomaco pieno dopo un lauto pranzo) e propriocettiva (relativa alla posizione del corpo nello spazio e allo stato di contrazione dei muscoli). La sensibilità termo-dolorifica è invece una sensibilità più antica della meccanocettiva, che raccoglie informazioni sulla temperatura e segnala danni tissutali attraverso le sensazioni dolorose. Il segnale viene poi trasmesso lungo i nervi fino al midollo spinale, per poi risalire facendo tappa in diverse stazioni per l'integrazione e la gestione degli stimoli: ad esempio una tappa obbligata per tutti i segnali è il talamo, un ammasso di neuroni e assoni al centro del nostro encefalo che si occupa di quel processo di "filtrazione" dei segnali rilevanti che abbiamo citato all'inizio. Infine, il segnale viene smistato verso un centro sensoriale della corteccia; ad esempio, per quanto riguarda le stimolazioni tattili, questo è la *corteccia sensoriale primaria*, o *area somestesica primaria*, nella porzione anteriore del lobo parietale. Invece, le fibre della sensibilità propriocettiva giungono oltre che alla corteccia sensoriale primaria anche al cervelletto. Nell'area sensoriale i neuroni sono organizzati in localizzazioni precise, venendo a determinare una sorta di rappresentazione del nostro corpo detta "homunculus sensitivo"; infatti, queste aree non sono disposte casualmente, ma prendendo in considerazione il corpo dal basso verso l'alto (dai piedi alla testa) le corrispettive sezioni corticali saranno mediali per le porzioni inferiori (piedi, gambe, etc.) e sempre più laterali andando superiormente (il viso sarà laterale). Inoltre, questa mappa detta somatotopica non sarà fedelmente in scala: le aree più vaste saranno quelle corrispondenti alle zone più sensibili, come volto o mani, mentre parti del corpo grandi ma poco sensibili come la schiena corrisponderanno ad aree di piccole dimensioni.

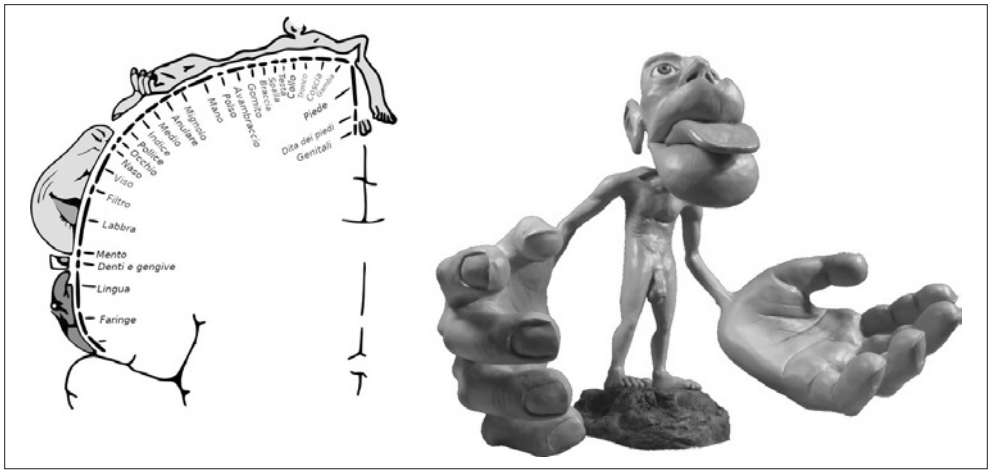


FIG. 2 L'*homunculus* sensoriale presenta un'anatomia distorta, dove le dimensioni delle parti del corpo dipendono dalla densità di terminazioni sensoriali. Elaborazione dell'immagine: "Front of Sensory Homunculus.gif" by Mpj29 (licensed under CC BY-SA 4.0).

Se da un lato nella Sindrome dell'Arto fantasma entra in gioco la sensibilità meccanocettiva (la sensazione di avere l'immaginario arto teso e contratto), dall'altro la sfida maggiore nel trattamento di questa condizione è rappresentata dalla percezione di dolore che essa suscita: una sfida che purtroppo non abbiamo ancora vinto. Il dolore, questa sensazione spiacevole che abbiamo conosciuto tutti almeno una volta nella vita, è generato attraverso dei recettori specifici, chiamati nocicettori. Per quanto raggruppati in un unico nome, la realtà è che sono decisamente variegati, dovendo essere capaci rispondere a stimoli così diversi; si trovano inoltre quasi ovunque, sulla pelle come sulla superficie degli organi. Tra le poche eccezioni figura il cervello, di modo che spesso i neurochirurghi utilizzano una blanda anestesia locale ed operano col paziente sveglio, per poter verificare in tempo reale di non star danneggiando inavvertitamente aree del linguaggio o motorie. Anche il rene ed il fegato sono "immuni" al dolore, determinando però anche risvolti negativi come la diagnosi tardiva di tumori importanti, che crescono quasi asintomatici.

La sensazione di dolore viene trasmessa da questi nocicettori attraverso delle fibre nervose particolari lungo il midollo spinale (eccetto il dolore dell'area del viso, che salta questa tappa), raggiungendo l'encefalo. Le fibre che ne trasportano il segnale raggiungono non solo delle aree del cervello come la corteccia sensoriale e la corteccia prefrontale, dedicata al ragionamento razionale, ma anche aree come ad esempio l'area limbica e l'amigdala, profondamente coinvolte nell'elaborazione delle emozioni, dell'ansia, della paura e degli istinti. Queste

fan parte di un sistema antichissimo, formato da vari nuclei nervosi, chiamato nel complesso sistema limbico: un network conservato fin dai rettili, racchiuso attorno al talamo nella profondità dell'encefalo, che svolge un ruolo nell'elaborazione della componente affettiva ed emozionale del dolore. L'unico altro senso che condivide con il dolore una connessione così importante e funzionale al sistema limbico è l'olfatto, anch'esso una sensibilità ritenuta tra le più antiche ad aver avuto origine.

Inoltre, vi è un'altra classificazione che risulta utile operare, per inquadrare con più precisione il tipo di dolore che ci si trova di fronte. Abbiamo definito il dolore dell'amputato come neuropatico, per distinguerlo da altri due tipi di dolore: il nocicettivo, generato dai nocicettori, che raccolgono lo stimolo doloroso esterno (chimico, infiammatorio, termico...) e l'idiopatico, aggettivo complicato che indica l'aver causa sconosciuta. Il dolore neuropatico è invece causato da una lesione della via nervosa, che risulta in un'alterazione della trasmissione o dell'elaborazione del segnale. Si tratta probabilmente del tipo di dolore più complicato da gestire: compare in ritardo rispetto alla lesione scatenante, ha una presentazione variabile (può essere descritto come un bruciore diffuso o caratterizzarsi per improvvisi episodi di dolori acuti e lancinanti) ed è percepibile anche in assenza di una lesione chiaramente identificabile e permanente: soprattutto, manca di trattamenti efficaci. Nemmeno gli antidolorifici più potenti di cui disponiamo, gli oppiacei, hanno su di esso una grande efficacia.

3. LA DECIFRAZIONE DEL MISTERO

3.1 LE TEORIE MECCANICISTICHE

Il nome e la prima descrizione clinica si devono proprio al dottor Silas Weir Mitchell, che pose le basi per la sua discussione nell'ambiente medico. Eppure, questa condizione era già stata riportata per la prima volta nel 1552 dal chirurgo francese Ambroise Paré, che la notò come Mitchell nei feriti di guerra[3]. Compare poi tra gli scritti di René Descartes, viene citata dal medico tedesco Aaron Lemos, dall'anatomista scozzese Sir Charles Bell e dal suo connazionale William Portfield, medico del diciottesimo secolo, che la sperimentò in prima persona in seguito alla perdita

[3] G. Keynes (ed.), *The apologie and treatise of Ambroise Paré*, Chicago: University of Chicago Press, 1952, p. 147.

della gamba: fu il primo a ipotizzarne la causa nell'alterata percezione sensoriale dei nervi[4]. Con il lavoro pionieristico di Mitchell si ebbe dunque la prima stesura di un'ipotesi fisiopatologica, che perdurò a lungo come unica spiegazione. Una ripresa dello studio sull'argomento si ebbe con dei lavori negli anni '40, che continuarono ad imputarne la causa alla lesione del nervo ed in particolare alla formazione di un neuroma[5] (un accrescimento od un ispessimento anormale del tessuto nervoso, tipico del nervo reciso in seguito ad un trauma) nonostante fossero riportati casi di dolore associato alla sindrome insorto poco dopo l'amputazione, prima dunque che il neuroma si fosse costituito.

3.2 IL NEUROMATRIX

Per molti anni, l'ipotesi dominante per spiegare il fenomeno rimase quindi l'irritazione del sistema nervoso periferico a livello del sito d'amputazione (neuroma); tuttavia, questa teoria presentava dei punti deboli che rendevano vacillante la struttura. Ad esempio, all'inizio degli anni '90 lo psicologo canadese Ronald Melzack, uno dei maggiori contributori alla nostra attuale conoscenza del dolore, dimostrò che varie persone senza arti fin dalla nascita, quindi che non avevano sviluppato connessioni neurali che potessero esser poi venute a mancare (e nemmeno neuromi), erano comunque soggette all'esperienza dell'arto fantasma[6]. Secondo Melzack la causa della sindrome era infatti da cercarsi a monte: l'esperienza del corpo, diceva, viene creata da un ampio network di strutture neurali interconnesse (che lui denominò "Neuromatrix") capaci di determinare l'insorgenza della malattia. Il dolore può essere generato non solo dalla pura sensazione nocicettiva, egli affermava, ma anche dal conflitto tra la percezione visiva della assenza dell'arto, la percezione sensoriale alterata e tra la rappresentazione propriocettiva, motoria; la mancanza di coordinazione, l'incoerenza tra le varie aree dà vita alle varie sensazioni, tra cui quelle di dolore, tipiche della sindrome. La teoria del ruolo nell'insorgenza del dolore di strutture nervose venne dimostrata col tempo, ad esempio da studi che provarono la possibilità di riprodurre dolori tipici di altre sindromi (come l'angina durante l'infarto) attraverso la stimolazio-

---

[4] M. Rugnetta, *Phantom limb syndrome*, in *Encyclopædia Britannica*, 13 Jul. 2018, *https://www.britannica.com/science/phantom-limb-syndrome.*

[5] K.L. Collins et al., *A review of current theories and treatments for phantom limb pain*, in *J. Clin. Invest.,* 2018 Jun 1, 128(6), pp. 2168-2176.

[6] R. Melzack, *Phantom limbs and the concept of a neuromatrix*, in *Trends in neurosciences*, 13.3, 1990, pp. 88-92.

ne elettrica di alcune aree cerebrali[7]. Per indagare con più cura l'influenza della percezione integrata del nostro corpo sull'insorgenza del dolore nella sindrome dell'arto fantasma, vennero elaborati una serie di esperimenti interessanti ma allo stesso tempo relativamente semplici: ad esempio[8], presi dei soggetti sani, gli si faceva muovere gli arti in maniera incongruente mentre guardavano il proprio riflesso su uno specchio, o con la vista bloccata da uno schermo. Il gruppo dello specchio riportava in maggior numero sintomi come formicolii, perdita della sensibilità, prurito e dolore (simili a quelli della sindrome dell'arto fantasma) proprio per via del forte conflitto tra visualizzazione, input somatosensoriale e rappresentazione corticale.

3.3 LA RIORGANIZZAZIONE CORTICALE

Ma a completare il quadro, partendo da queste teorie, fu un'altra intuizione: la teoria della riorganizzazione corticale (*Cortical Remapping Theory*, CRT). Essa afferma che i neuroni che ricevevano inputs dall'arto amputato vanno incontro, in seguito alla perdita del corrispondente territorio di innervazione, ad una "riorganizzazione" e rispondano ad inputs provenienti da altre aree, in particolare dal viso. Questo perché l'area della corteccia somestesica deputata all'arto perduto viene invasa da neuroni della vicina area somatosensoriale facciale: conseguentemente, un amputato può sperimentare dolore fantasma attraverso la stimolazione di certe aree del viso. Questa espansione ed invasione viene attribuita alla mancata ricezione di informazioni sensoriali relative all'arto perso da parte della corrispondente area corticale, portando dunque a questa sorta di riarrangiamento delle fibre indotta dalla grande plasticità di cui dispone il nostro sistema nervoso[9].

Nel 1983 infatti si dimostrò[10] l'esistenza di riorganizzazioni neurali all'interno di alcune parti dell'area somatosensoriale nei primati e nel 1991 un team guidato

---

[7] F.A. Lenz *et al.*, *The sensation of angina can be evoked by stimulation of the human thalamus*, in *Pain*, 59.1, 1994, pp. 119-125.

[8] C.S. McCabe *et al.*, *Simulating sensory-motor incongruence in healthy volunteers: implications for a cortical model of pain*, in *Rheumatology*, 44(4), 2005, pp. 509-516.

[9] J.T. Wall, J. Xu, X. Wang, *Human brain plasticity: an emerging view of the multiple substrates and mechanisms that cause cortical changes and related sensory dysfunctions after injuries of sensory inputs from the body*, in *Brain Res. Brain Res. Rev.*, 39(2-3), 2002, pp. 181-215.

[10] M.M. Merzenich *et al.*, *Topographic reorganization of somatosensory cortical areas 3b and 1 in adult monkeys following restricted deafferentation*, in *Neuroscience*, 8(1), 1983, pp. 33-55.

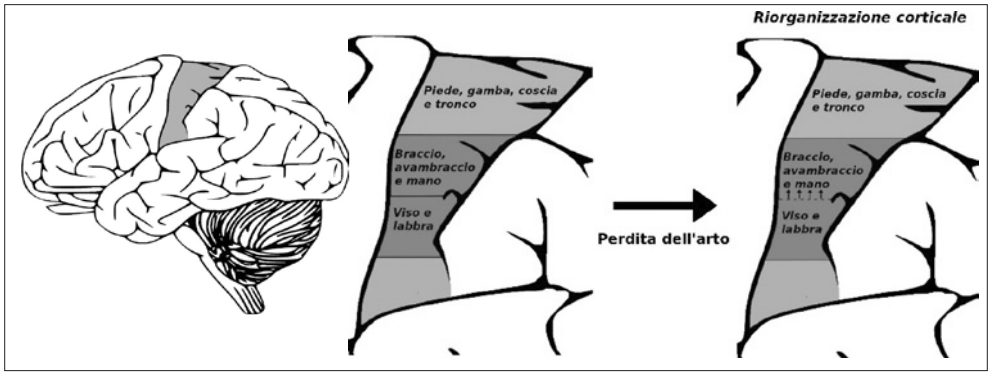


FIG. 3 Schematizzazione della riorganizzazione corticale.

da Tim Pons, ricercatore al laboratorio di Neuropsicologia del National Institute of Health, pubblicò un lavoro su Science descrivendo come la corteccia somatosensoriale primaria del macaco andasse incontro a riorganizzazioni sostanziali precisamente in seguito alla perdita di un input sensoriale, dimostrando precisamente la correlazione ipotizzata[11].

Venuto a conoscenza di questi risultati, un medico indiano che lavorava come ricercatore alla University of California, Vilayanur S. Ramachandran, considerò che si trattasse di un approccio molto promettente e decise di dirigere in quella direzione la sua ricerca.

3.4 IL LAVORO DI RAMACHANDRAN E SUCCESSIVI SVILUPPI

Vilayanur Subramanian Ramachandran nasce a metà del secolo scorso nel Tamil Nadu, in India. Il padre, ingegnere per l'Organizzazione per lo Sviluppo Industriale dell'ONU e diplomatico in Thailandia, lo conduce con sé a Bangkok, dove frequenta le scuole britanniche. Ritorna poi a Madras, nel cuore del Tamil Nadu, per frequentare lo Stanley Medical College e specializzarsi in chirurgia nel 1974. Ma già durante gli anni della specialistica comincia in realtà a lavorare su un filone di ricerca che lo accompagnerà per anni: la percezione visiva. Di gran talento, riesce ad ottenere la possibilità di svolgere il suo Ph.D. all'università di Cambridge, che conseguirà nel 1978. Studia i legami tra la percezione

[11] T.P. Pons *et al.*, *Massive reorganization of the primary somatosensory cortex after peripheral sensory deafferentation*, in *Science* 252.5014, 1991, pp. 1857-1860. doi:10.1126/science.1843843. PMID 1843843.

visiva e la fonetica, la visione stereoscopica, il movimento apparente: prima a Cambridge, poi alla Caltech dove svolge un post-doc e infine all'Università della California, San Diego, dove diventerà ordinario nel 1998. La sua ricerca scientifica conosce una nuova fase a metà di questo percorso, mentre è ancora Assistant Professor a San Diego: comincia, incuriosito, ad interessarsi alle neuroscienze cognitive ed in particolare ad alcune rare sindromi neurologiche, tra cui la Sindrome dell'Arto Fantasma.

In seguito agli studi di Pons, Ramachandran condusse dei semplici esperimenti che cambiarono la comprensione della patologia. Egli organizzò uno studio per dimostrare che le sensazioni dolorose siano dovute alla riorganizzazione della corteccia somatosensoriale del cervello umano, riuscendo a confermarlo nel 1994 assieme al suo team[12]. Continuando le sue ricerche, pubblicò inoltre un lavoro dove mostrava che accarezzare parti differenti del viso generava la sensazione di venire toccati in parti differenti dell'arto mancante, e successive analisi cerebrali degli amputati mostrarono lo stesso tipo di riorganizzazione corticale che Pons aveva rinvenuto anni prima nel macaco[13]. Inoltre, grazie a questa ipotesi è possibile spiegare perché il dolore fantasma insorga meno in seguito alla perdita dell'occhio rispetto all'arto: la causa potrebbe risiedere nella minore estensione dell'area che lo rappresenta all'interno della corteccia somatosensoriale primaria, che rende quindi meno probabile una riorganizzazione post-amputazione.

Il quadro tuttavia non riscuote l'approvazione della totalità della comunità scientifica: alcuni studi dibattono sulla possibilità che la mappatura originale possa venire conservata proprio attraverso il fenomeno del dolore fantasma[14], bloccando così il processo di riorganizzazione. Oppure, alcuni mettono in dubbio l'importanza del ruolo della corteccia somatosensoriale nell'insorgenza del dolore[15], per quanto i risultati a proposito non siano conclusivi. Comunque, la CRT rappresenta probabilmente la teoria maggiormente condivisa ed approvata dalla comunità scientifica.

---

[12] T.T. Yang *et al.*, *Noninvasive detection of cerebral plasticity in adult human somatosensory cortex*, in *NeuroReport*, 5(6), 1994, p. 7014. doi:10.1097/00001756-199402000-00010. PMID 8199341.

[13] V.S. Ramchandran, W. Hirstein, *The perception of phantom limbs*, in *Brain*, 121(9), 1998, pp. 1603-1630. doi:10.1093/brain/121.9.1603. PMID 9762952.

[14] T.R. Makin *et al.*, *Reassessing cortical reorganization in the primary sensorimotor cortex following arm amputation*, in *Brain*, 138(Pt 8), 2015, pp. 2140-2146.

[15] W. Penfield, M. Faulk, *The insula, further observation on its function*, in *Brain*, 78(4), 1955, pp. 445-470.

3.5 ALTRI ATTORI DA TENERE IN CONSIDERAZIONE

In realtà, per quanto si possa dare il ruolo di protagonista alla corteccia somatosensoriale, rimangono in scena delle strutture che si ritiene influiscano sul quadro, sebbene in maniera ancora poco chiara: ad esempio il talamo[16], che potrebbe partecipare attivamente alla riorganizzazione delle fibre localmente ed a livello corticale[17]. Inoltre, la memoria propriocettiva (la memoria della posizione del corpo nello spazio e dello stato di contrazione dei muscoli) detiene un'importanza non trascurabile, e ciò è intuibile alla luce del fatto che gli amputati riferiscono di percepire l'arto mancante come se si trovasse nella posizione precedente alla rimozione e tenendo presente l'alta frequenza dell'associazione tra la presenza del dolore e la sensazione che l'estremità persa abbia assunto una posizione scomoda[18]. D'altronde, è fondamentale ricordare che l'ipotesi meccanicistica che correla il fenomeno patologico con la costituzione del neuroma presenta una buona evidenza al suo seguito, ed è quindi da tenere ben presente il ruolo del sistema nervoso periferico nell'insorgenza della sintomatologia. In seguito al taglio completo del nervo, le terminazioni nervose diventano più attive e sensibili a stimoli chimici e meccanici[19]; sensibilità aumentata da stati di allerta in cui vi è rilascio di ormoni adrenalinici[20], come dimostrato da un recente studio sui ratti.

4. ESORCIZZARE IL FANTASMA

In alcune persone, il dolore fantasma può lentamente scomparire spontaneamente nel corso di alcuni mesi o finanche qualche anno: inoltre, ricordiamo, la sindrome non si manifesta in tutti i pazienti. Comunque, gran parte dei trattamenti utilizzati

---

16 E.R. Ergenzinger *et al.*, *Cortically induced thalamic plasticity in the primate somatosensory system*, in *Nat. Neurosci.*, 1(3), 1998, pp. 226-229.

17 C.X. Li *et al.*, *Forelimb amputation-induced reorganization in the ventral posterior lateral nucleus (VPL) provides a substrate for large-scale cortical reorganization in rat forepaw barrel subfield (FBS)*, in *Brain Res.*, 1583, 2014, pp. 89-108.

18 J. Katz, R. Melzack, *Pain 'memories' in phantom limbs: review and clinical observations*, in *Pain.*, 43(3), 1990, pp. 319-336.

19 K.C. Kajander, S. Wakisaka, G.J. Bennett, *Spontaneous discharge originates in the dorsal root ganglion at the onset of a painful peripheral neuropathy in the rat*, in *Neurosci. Lett.*, 138, 1992, pp. 225-228.

20 Y.-F. Lü *et al.*, *The Locus Coeruleus–Norepinephrine System Mediates Empathy for Pain through Selective Up-Regulation of P2X3 Receptor in Dorsal Root Ganglia in Rats*, in *Frontiers in Neural Circuits*, 11, 2017. https://doi.org/10.3389/fncir.2017.00066.

nella pratica clinica degli ultimi vent'anni non ha sfortunatamente mostrato consistenti miglioramenti dei sintomi. Le idee spaziano tra gli approcci più vari: dalla farmacoterapia all'impiego di tecniche non invasive o alternative come l'ipnosi o l'agopuntura, la stimolazione spinale, la terapia delle vibrazioni, fino a giungere persino ad approcci chirurgici; ma mancano prove significative di una definitiva maggiore efficacia di uno di loro, se non forse per la Mirror Therapy.[21]

4.1 FARMACOTERAPIA

L'approccio farmacologico possiede svariate possibilità d'azione, ma viene normalmente affiancato ad altre terapie non farmacologiche non invasive per migliorare la gestione del dolore. Sono utilizzati gli antidolorifici, da composti più blandi come i FANS fino a potenti sostanze come gli oppiodi (morfina, ketamina). Purtroppo, questi non hanno un'efficacia così risolutiva come si potrebbe immaginare, proprio per le caratteristiche peculiari del dolore neuropatico; comunque, si ipotizza possano inibire la riorganizzazione corticale successiva all'amputazione[22], oltre che agire sulla trasmissione del dolore a livello del cervello e del midollo spinale; presentano però importanti effetti collaterali, tra cui lo sviluppo di una forte dipendenza. La classe di farmaci più utilizzata è quella degli antidepressivi triciclici (TCA), in particolare l'amitriptilina che si è dimostrata utile nel ridurre il dolore neuropatico[23]. Risultati interessanti provengono anche dall'impiego di vari tipi di farmaci antiepilettici[24] come la gabapentina, spesso usati in combinazione con i TCA.

4.2 TERAPIE CHIRURGICHE

Le terapie chirurgiche più comuni per il trattamento della sintomatologia dolorosa sono abbastanza invasive e vengono utilizzate solo quando gli altri tentativi falliscono; comunque, nonostante le possibilità siano molteplici, nessuna ha mostrato finora una significativa superiorità. Si tratta di neurectomie (ablazione di un nervo

---

[21] K. L. Collins *et al.*, *A review of current theories and treatments for phantom limb pain*, cit.

[22] E. Huse *et al.*, *The effect of opioids on phantom limb pain and cortical reorganization*, in *Pain*, 90(1-2), 2001, pp. 47-55.

[23] E. Kalso, T. Tasmuth, P.J. Neuvonen, *Amitriptyline effectively relieves neuropathic pain following treatment of breast cancer*, in *Pain*, 64, 1996, pp. 293-302

[24] F. Levendoglu *et al.*, *Gabapentin is a first line drug for the treatment of neuropathic pain in spinal cord injury*, in *Spine*, 29(7), 2004, pp. 743-751.

o più comunemente di un suo tratto, talvolta con radioterapia o talvolta con neurotossine), interventi di revisione del moncone (ad esempio la reinnervazione mirata del muscolo, per prevenire la formazione del neuroma), nonché la stimolazione elettrica cerebrale e la stimolazione elettrica spinale[25].

4.3 TERAPIE NON INVASIVE

Tra di esse si contano interventi di grandissima varietà, dimostrando la moltitudine d'idee differenti con cui si è cercato di alleviare questa sindrome. Ricordiamo tra queste la stimolazione transcutanea del nervo (TENS), che ha riportato risultati positivi in alcuni lavori[26], ed un'altra terapia con una discreta diffusione: l'agopuntura, che può essere capace di alleviare il dolore tramite l'inserimento di sottili aghi d'acciaio sterilizzati in specifici punti della pelle.

Lo scenario subì in realtà una rivoluzione in seguito agli studi di Ramachandran, proprio per la sua visione olistica del quadro: per quanto possa sembrare poco ortodosso ed in un certo qual modo straordinario, riscosse grande successo l'invenzione e l'uso in clinica della cosiddetta "mirror box", concepita dallo scienziato indiano ragionando sull'eziologia della malattia. Si tratta di un rimedio dall'efficacia promettente, bassissima invasività e quasi a costo zero.

Essa è costituita da una scatola con due buchi, ove infilare in uno l'arto monco e nell'altro l'arto integro: essa presenta poi uno specchio, a mezzo, diretto verso l'arto intatto del paziente, il quale restituendo il riflesso dell'arto sano mentre svolge una serie di esercizi soddisfa l'illusione del soggetto di riuscire a "muovere" l'arto fantasma, svincolandolo dalle posizioni potenzialmente dolorose che egli percepiva aver assunto.

La prima prova in clinica rappresentò un successo promettente, riuscendo a diminuire il dolore nel 93% dei pazienti[27]: il primo amputato che la provò affermò di essere riuscito a "muovere" il suo arto mancante per la prima volta in decenni. Il principio terapeutico si basa quindi sull'impiego dell'illusione ottica per fornire dei feedback

---

[25] B. Subedi, G.T. Grossberg, *Phantom limb pain: mechanisms and treatment approaches*, in *Pain Res. Treat.*, 2011, 2011, 864605

[26] L.M. Black, R.K. Persons, B. Jamieson, *Clinical inquiries. What is the best way to manage phantom limb pain?*, in *J. Fam. Pract.*, 58, 2009, pp. 155-158; J. Katz, R. Melzack, *Auricular transcutaneous electrical nerve stimulation (TENS) reduces phantom limb pain*, in *J. Pain Symptom. Manag.*, 6(2), 1991, pp. 73-83.

[27] B.L. Chan *et al.*, *Mirror therapy for phantom limb pain*, in *N. Engl. J. Med.*, 357(21), 2007, pp. 2206-2207.

FIG. 4 Esempio di mirror box.

visivi capaci di "ingannare" il cervello a credere di essere ritornato in possesso dell'arto perso. Alcuni studi hanno correlato il funzionamento della terapia con l'attivazione dei neuroni specchio[28], neuroni coinvolti nell'esecuzione di un movimento finalizzato e nell'osservazione dello stesso movimento svolto da un altro individuo. Un lavoro del 2014 riporta come osservare delle persone muovere le proprie gambe abbia lenito il dolore associato alla sindrome in vari soggetti che avevano perso entrambe le gambe, corroborando l'ipotesi che il feedback visivo abbia un ruolo cruciale nella patologia; si tratta di uno spunto interessante, in particolare quando abbinato all'osservazione dell'esistenza di una generale mancanza di studi sulla sindrome dell'arto fantasma in soggetti privi della visione[29]. Una review del 2018, pur sottolineando la mancanza ancora di un adeguato corpus di studi al proposito, ha concluso basandosi su 15 studi ritenuti idonei che la terapia potesse essere efficace nel ridurre il dolore da arto fantasma, riducendo l'intensità e la quantità degli episodi dolorosi attraverso un trattamento semplice, economico e non invasivo.

4.4 LE TERAPIE CHE VERRANNO

La sfida della sindrome dell'arto fantasma è ancora lungi dall'essere conclusa, continuando ad essere un campo esplorato con interesse dalla ricerca scientifica: il pano-

---

[28] K. Guenther, *It's all done with mirrors: V.S. Ramachandran and the material culture of phantom limb research*, in *Med. Hist.*, 60(3), 2016, pp. 342-358; M.L. Tung *et al.*, *Observation of limb movements reduces phantom limb pain in bilateral amputees*, in *Ann. Clin. Transl. Neurol.*, 1(9), 2014, pp. 633-638.

[29] B.A. Petersen *et al.*, *Phantom limb pain: peripheral neuromodulatory and neuroprosthetic approaches to treatment*, in *Muscle & nerve*, 59.2, 2019, pp. 154-167.

rama si arricchisce infatti di diversi altri approcci, nonostante questi non abbiano ancora raggiunto un'ampia diffusione nella pratica clinica. Un esempio è l'utilizzo di trattamenti neuroprostetici[30], che combinano l'uso di una protesi con la stimolazione dei nervi periferici della parte residuale dell'arto. Si tratta di una terapia parzialmente invasiva, perché prevede il posizionamento di elettrodi a livello delle fibre nervose; questo, insieme al fatto che ci siano ancora poche prove della stabilità nel tempo della capacità di evocare impulsi adeguati, ne complicano il bilancio rischio/beneficio. Sono al momento allo studio innovativi metodi d'impianto per aumentare la durata temporale di questi sistemi, (elettrodi che stimolano la rigenerazione assonale, utilizzo di trapianti autologhi di muscolo[31]) e ci sono buone speranze che possano contribuire significativamente a prevenire la formazione del neuroma e ridurre l'intensità del dolore. Yanagisawa[32] ha ottenuto un risultato interessante grazie all'utilizzo di protesi robotiche controllate dal paziente attraverso un'interfaccia cervello-computer, basata sull'attività cerebrale. Il soggetto doveva cercare di muovere l'arto fantasma, e il programma si occupava di recepire l'attività corticale legata al movimento utilizzandola per comandare l'arto robotico: il risultato ha indicato un calo del dolore significativo rispetto al gruppo di controllo, in cui la mano robotica effettuava invece dei movimenti casuali. Un'altra strategia terapeutica, che rappresenta un'intuizione recente, è l'affiancamento della terapia di Ramachandran ad una tecnologia in forte crescita negli ultimi anni, la realtà virtuale (VR): il soggetto si ritrova immerso in un contesto modificabile dall'operatore, garantendo la massima immersività della simulazione, in cui l'arto fantasma appare ricostruito e si muove sotto il controllo del paziente. Il campo è ancora giovane, ma alcuni studi[33] dimostrano significative riduzioni del dolore, trasmettendoci speranze incoraggianti per l'avvenire. Ortiz-Catalan ed il suo team hanno posizionato dei

---

30 C.A. Kubiak *et al.*, *Regenerative peripheral nerve interface for management of postamputation neuroma*, in *JAMA surgery*, 153.7, 2018, pp. 681-682.

31 P.P. Vu *et al.*, *A regenerative peripheral nerve interface allows real-time control of an artificial hand in upper limb amputees*, in *Science Translational Medicine*, 12.533, 2020;

32 T. Yanagisawa *et al.*, *Using a BCI Prosthetic Hand to Control Phantom Limb Pain*, in C. Guger, N. Mrachacz-Kersting, B.Z. Allison (eds), *Brain-Computer Interface Research*. *A State-of-the-Art Summary*, 7, Springer, Cham, 2019, pp. 43-52.

33 C. Mercier, A. Sirigu, *Training with virtual visual feedback to alleviate phantom limb pain*, in *Neurorehabil. Neural Repair*, 23(6), 2009, pp. 587-594; M. Ortiz-Catalan *et al.*, *Phantom motor execution facilitated by machine learning and augmented reality as treatment for phantom limb pain: a single group, clinical trial in patients with chronic intractable phantom limb pain*, in *Lancet*, 388(10062), 2016, pp. 2885-2894.

sensori all'interno del muscolo per misurare il potenziale elettrico, in maniera tale da prevedere quale movimento il soggetto cerchi di fare: immediatamente dopo, l'arto virtuale nello schermo esegue questi movimenti. Dunque, quando il paziente guarda lo schermo riesce ad immergersi in uno scenario più realistico rispetto alla terapia con la mirror box: tra i 14 pazienti che hanno partecipato a tale ricerca, il livello di dolore cronico è diminuito mediamente del 50% in seguito alla terapia. Osumi ed il suo team[34] hanno recentemente designato un esperimento di Mirror therapy virtuale, in cui i soggetti cercavano di afferrare degli oggetti virtuali simultaneamente con la mano intatta e con la mano fantasma; altri studi simili hanno mostrato che delle sessioni giornaliere di VR training possono influire sull'intensità del dolore provato. L'entusiasmo è tanto, nuovi approcci continuano ad essere sviluppati nel tentativo di migliorare la qualità di vita delle persone affette dalla sindrome dell'arto fantasma; tuttavia è necessario notare che si tratta di trattamenti non ancora sufficientemente sviluppati. Gli studi sono pochi, riguardano gruppi ristretti di pazienti e presentano un certo rischio di bias: per quanto i risultati siano promettenti, è necessario proseguire nella ricerca con studi clinici più ampi e strutturati. Ma appare chiaro che i prossimi anni potrebbero far concretizzare molte speranze, portando delle nuove risposte a tutte le persone che convivono con questa difficile condizione. Secondo i dati sui ricoveri ospedalieri del Ministero della Salute, solo nel 2018 vi sono state ben 11.612 dimissioni di pazienti dagli ospedali in seguito ad un'amputazione[35].

---

[34] M. Osumi *et al.*, *Characteristics of phantom limb pain alleviated with virtual reality rehabilitation*, in *Pain Medicine*, 20.5, 2019, pp. 1038-1046.

[35] Ministero della Salute, *Rapporto annuale sull'attività di ricovero ospedaliero. Dati SDO 2018*, 2020.

I NON PIÙ PROMESSI SPOSI

BREVE STORIA DEL DIVORZIO TRA LOCALITÀ E REALISMO NELLA FISICA MODERNA

SIMONE TENTORI

Nel ventesimo secolo la meccanica quantistica ha fatto il suo ingresso sotto i riflettori del mondo, scuotendo profondamente le fondamenta della fisica. La nuova teoria era in aperto contrasto con il realismo locale, il principio su cui si basava la concezione dell'Universo fino allora creduta vera, tanto da dare il via ad uno ampio dibattito tra gli scienziati dell'epoca, vedendo contrapposti due giganti come Einstein e Bohr.
In questo breve saggio ripercorreremo la storia dello scontro fra la meccanica quantistica e il realismo locale, le sue implicazioni sulla natura della realtà in cui viviamo e una soluzione al problema, spesso lasciata da parte, riguardante la vera essenza del libero arbitrio.

In the twentieth century, Quantum Mechanics made its appearance in the world, shaking the foundations of Physics. The new theory was in open contrast with the local realism, the principle on which was built the conception of the world that time. This contradiction led to a wide debate between physicists, even between two giants like Einstein and Bohr.
In this review we will retrace the history of the clash between Quantum Mechanics and local realism, its implications on the reality in which we live and a often forgotten solution to the problem concerning the true nature of free will.

PREFAZIONE

«La Luna è ancora lì quando nessuno la osserva?» Questa frase provocatoria, attribuita a Einstein durante il lungo dibattito con Bohr sulla oggettività del mondo esterno, rappresenta uno dei momenti più profondi del pensiero scientifico.

Intorno alla metà degli anni Venti del secolo scorso, il formalismo della meccanica quantistica stava iniziando a prendere forma. Un grande quantità di dati potevano essere correttamente interpretati dalla nuova teoria, e molte previsioni sperimentali sarebbero state verificate da lì a poco con una precisione strabiliante. Ma accanto all'indubbio successo sul piano sperimentale, le interpretazioni che scaturivano dal formalismo sarebbero rimaste per decenni oggetto di controversie, e a tutt'oggi lo stato delle cose è lontano dall'essere risolto.

Da allora, l'atteggiamento della comunità scientifica si è biforcato: da un lato, una componente predominante prese la strada di quella che poi sarebbe divenuta la posizione ortodossa, *shut-up and calculate*, invitando a evitare di chiedersi cosa sia la realtà ma cercando di carpire – seguendo il sentiero gnoseologico di Bohr – cosa filtri di essa dal formalismo della teoria.

L'altra componente, minoritaria, intraprese con pervicacia la strada opposta. Come ebbe a dire Bryce De Witt «non è saggio ignorare ciò che il formalismo sta cercando di dirci»: ignorarne le profonde implicazioni sul piano ontologico e gnoseologico rappresenterebbe per molti un gettare la spugna da parte del pensiero scientifico.

È in questo contesto che si inserisce il lavoro di Bell. Le sue famose disuguaglianze ci hanno mostrato che ciò che in un contesto storico viene visto come un sterile dibattito epistemologico, successivamente può essere sottoposto al vaglio dell'esperienza. Il divorzio tra realismo e località, raccontato in questo brano, è un primo passo verso la comprensione del mondo esterno. Ammesso che un mondo esterno esista davvero, e che possiamo sondarlo in modo libero e indipendente, come viene discusso suggestivamente nel finale.

Maximiliano Sioli
Professore associato di Fisica Sperimentale
Dipartimento di Fisica e Astronomia "Augusto Righi"
Università di Bologna

INTRODUZIONE

Era il 22 gennaio 1990 quando John Stewart Bell al CERN pronunciò uno storico discorso che riassumeva come la concezione del mondo era dovuta cambiare in seguito alla formulazione della meccanica quantistica:

> La sola idea di un'inquietante azione a distanza è repulsiva per i fisici. Se avessi un'ora, vi inonderei di citazioni di Newton, Einstein, Bohr ed altri grandi uomini. Vi direi quanto sia impensabile essere capaci di modificare una situazione distante facendo qualcosa qui. Penso che i padri fondatori della meccanica quantistica non avessero davvero bisogno delle argomentazioni di Einstein sulla necessità di escludere un'azione a distanza, perché la loro attenzione era rivolta altrove. L'idea di un determinismo o di un'azione a distanza era per loro talmente repulsiva che preferirono voltarsi dall'altra parte. Bene, è una tradizione, e nella vita a volte dobbiamo imparare a imparare nuove tradizioni. E potrebbe accadere che dobbiamo non tanto accettare un'azione a distanza, quanto l'inadeguatezza di "nessuna azione a distanza".

Quell'*inquietante azione a distanza*, che tanto scosse i fisici nel secolo scorso è, al giorno d'oggi, raccontata e narrata in saggi, opuscoli e manuali di meccanica quantistica che riempiono ormai ogni libreria dotata di una sezione scientifica. Un vasto pubblico, con competenze e abilità critiche diverse, si è avvicinato al mondo dei quanti e, a volte, ne è uscito confuso. Nei vari saggi divulgativi, infatti, si tende spesso a portare agli occhi dei non esperti le tesi più fantasiose e più accattivanti per far innamorare il pubblico della materia. Questa operazione, tuttavia, non è immune da rischi, soprattutto nel caso della meccanica quantistica, in quanto teoria che gode del maggior numero di interpretazioni esistenti, che cercano di comprendere e spiegare il significato reale di fenomeni come il principio di indeterminazione e l'entanglement, tanto lontani dall'esperienza quotidiana.

Capita quindi che il lettore, passando da un saggio all'altro, venga sballottato da un'interpretazione all'altra della meccanica quantistica, senza capire veramente a fondo le differenze fra esse, o ancora peggio che vi è una profonda differenza fra una teoria fisica e la sua interpretazione.

Se poi il lettore non è esperto, quello che può emergere nella sua mente, è una personale interpretazione della meccanica quantistica e della realtà che lo circonda, magari lontanissima da quella che la fisica descrive.

In quest'ultimo decennio, poi, vari truffatori hanno iniziato ad accostare la parola "quantistica" ad ogni altro termine, creando veri e propri mostri: come le diete quantistiche e la psicologia quantistica. L'intento è quello di vendere il nulla a persone poco avvezze al significato della parola, oppure confuse dalla mole di articoli, talvolta troppo esemplificativi, talvolta addirittura errati, che compaiono su internet.

Da qui nasce questo breve saggio, per cercare di rispondere ad una domanda fondamentale: «Che cosa ci dice, veramente, la meccanica quantistica sulla realtà in cui viviamo?».

Per iniziare a parlare dell'argomento, però, sarà prima necessario descrivere brevemente come mai si rese necessaria l'introduzione di una teoria che superasse la meccanica newtoniana.

1900: IL PROBLEMA DEL CORPO NERO

Era il 1877 quando per la prima volta Ludwig Boltzmann avanzò l'idea che i livelli energetici in una molecola potessero essere discreti, e non continui come si era da sempre immaginato. Ci vollero poco più di vent'anni perché il suo suggerimento venisse raccolto da un fisico tedesco, Max Planck, per risolvere quello che era uno dei problemi più grandi per la fisica del tempo: la catastrofe ultravioletta. Chiamiamo corpo nero un oggetto ideale capace di assorbire tutta la radiazione elettromagnetica che lo colpisce. Il nome non deve trarre in inganno: il sole e le stelle sono ottime approssimazioni di corpo nero, pur non essendo affatto neri. Lo studio di questo oggetto era diventato molto importante, perché conoscere il suo potere emittente (la potenza emessa per unità di superficie, in funzione della temperatura T e della frequenza ν della radiazione emessa) avrebbe permesso di calcolare il potere assorbente (il rapporto fra potenza assorbita e incidente) di qualunque corpo, altrimenti difficile da misurare. Con la fisica classica, si arriva a calcolare che la densità di energia spettrale (cioè la densità di energia per unità di frequenza) del corpo nero vale:

$$\mu_\omega = \frac{\omega^2 k_B T}{\pi^2 c^3} \qquad (1)$$

dove k_B è la costante di Boltzmann e c la velocità della luce. Questa formula risulta abbastanza corretta a basse frequenze, ma fallisce, e anche di parecchio, man mano che queste aumentano. Ma c'è di peggio: da questa formula, integrando in ω è pos-

sibile trovare la densità di energia che risulta infinita. Quest'ultimo risultato prende il nome di catastrofe ultravioletta: quello che la fisica classica prevede è che un corpo a temperatura ambiente emetta raggi x e gamma, e che per aumentare anche di un solo grado la sua temperatura sia necessaria un'energia enorme. Sperimentalmente tutte queste conclusioni sono false (nessuno di noi ha mai fatto una radiografia stando davanti ad un fornello), ma nessuno era in grado di comprendere dove fosse l'errore. La questione era intricata, e per risolverla è stato necessario ripensare le cose a partire dalla base. Fu Planck che in un articolo, scritto nel 1900, risolse il problema supponendo che un oscillatore (nel nostro caso un atomo emittente radiazione elettromagnetica) con frequenza ω potesse assumere solo valori di energia multipli di $\hbar\omega$ dove $\hbar$ è una costante universale chiamata costante di Planck ridotta. Con questa assunzione si trova che:

$$\mu_\omega = \frac{\hbar\omega^3}{\pi^2 c^3 \exp\left(\frac{\hbar\omega}{k_B T}\right) - 1} \tag{2}$$

La legge di Planck riproduce con estrema precisione i risultati sperimentali ed evita la catastrofe ultravioletta. Per quanto strano fosse, era necessario allora accettare che l'energia non potesse assumere per gli oscillatori tutti i valori possibili, ma solo essere un multiplo di un quanto elementare.

LA MECCANICA QUANTISTICA IN BREVE

A partire da questa intuizione la teoria fu sviluppata e portata avanti dai più grandi scienziati del Novecento: Einstein (Effetto fotoelettrico, 1905), Bohr (quantizzazione del momento angolare nell'atomo di idrogeno, 1913), Compton (Compton scattering 1922), Pauli (Principio di esclusione, 1925), Born (Approssimazione di Born, 1926), Heisenberg (principio di indeterminazione, 1927), Dirac (Equazione di Dirac, 1928) e molti altri. Tassello dopo tassello si andò formando quella che oggi è sicuramente una delle teorie scientifiche corroborate dal maggior numero di prove sperimentali. Di seguito ricordiamo i principali punti della meccanica quantistica:

- lo stato di un sistema quantistico è descritto da una **funzione d'onda**, chiamata ψ;
- **principio di sovrapposizione**: se per un sistema sono possibili due stati ψ_1 e ψ_2 allora anche una loro combinazione lineare descrive uno stato del sistema: $\Psi = \alpha\psi_1 + \beta\psi_2$ con la condizione di normalizzazione: $\alpha^2 + \beta^2 = 1$

- la dinamica di un sistema quantistico isolato libero è descritta dall'**equazione di Schrödinger**:

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi \tag{3}$$

- a ogni **osservabile** accessibile mediante misura sperimentale è associato un particolare tipo di operatore hermitiano. Un operatore è un oggetto matematico che agisce sulla funzione d'onda;
- **riduzione del pacchetto d'onda**: ogni qualvolta si effettua una misurazione su un sistema quantistico, esso viene sempre trovato in un particolare autostato dell'operatore hermitiano associato alla quantità misurata. Ad ogni autostato corrisponde sempre un determinato autovalore, che rappresenta il risultato della misura. La funzione d'onda ϕ è un autostato di un operatore $\hat{A}$ se $\hat{A}\phi = \lambda_\phi\phi$, cioè se l'azione dell'operatore $\hat{A}$ su ϕ restituisce ϕ stessa moltiplicata per uno scalare λ_ϕ, detto autovalore associato all'autostato ϕ. Se l'operatore è hermitiano tutti gli autovalori associati ai suoi autostati sono numeri reali. Questo principio è importante quando lo si considera insieme al principio di sovrapposizione. Sappiamo che prima della misura il sistema può essere in uno qualunque degli stati per lui possibili, dopo di questa invece il sistema si trova nello stato rivelato dalla misurazione. Attorno a questo punto solitamente si sviluppano le varie interpretazioni della meccanica quantistica, tramite il postulato di Born[1] è possibile calcolare con quale probabilità ci aspetteremo il sistema in un certo stato.

Dai precedenti punti è possibile ricavare il **principio di indeterminazione**: alcune grandezze in meccanica quantistica vengono dette coniugate, in tal caso il prodotto dei loro errori di misurazione non può essere nullo, è il caso di posizione e quantità di moto:

$$\Delta q\Delta p \geq \frac{\hbar}{2} \tag{4}$$

Questa relazioni ci dice che non è possibile determinare con infinita precisione sia la posizione che la quantità di moto di una particella. Questo principio è molto importante, sancisce infatti l'impossibilità di esistenza di un demone di Laplace. Laplace, matematico francese del Settecento, aveva immaginato un demone in gra-

[1] In realtà non è un postulato, negli anni '70 del Novecento fu ricavata come legge a partire da altri assiomi della meccanica quantistica da Finkelstein (1963) , Hartle (1968) e Graham (1973)[2][3][4]

do di conoscere con infinita precisione la posizione e la quantità di moto di ogni particella dell'universo. Se ciò fosse stato possibile, quel demone avrebbe potuto predire, istante per istante, l'evoluzione dell'universo e quindi il futuro. Il principio di indeterminazione di Heisenberg fa crollare il determinismo, non è infatti possibile avere informazioni infinitamente precise contemporaneamente su impulso e quantità di moto di una particella.

Per rendere chiaro quanto detto possiamo pensare al principio di funzionamento di un polarizzatore, come quello presente in molti occhiali da sole. Questa lente permette il passaggio della luce (e in ultima analisi dei fotoni), solo se polarizzata lungo una certa direzione. Il fotone è un sistema quantistico e prima di arrivare alla lente è in una sovrapposizione di stati, pertanto la sua polarizzazione è indefinita. Quando il fotone arriva alla lente sta avvenendo una misura della sua polarizzazione, infatti in base a questa il polarizzatore lo lascerà passare o meno. Dopo l'incontro con l'apparato di misura il fotone ha uno stato di polarizzazione definito, orizzontale o verticale, la misura è stata effettuata e l'esito non è più indeterminato. Ma lo era davvero in precedenza? Vari scienziati, fra cui Albert Einstein in un primo momento, non erano assolutamente disposti a rinunciare al determinismo della fisica. Un'alternativa possibile era allora quella che esistessero delle **variabili nascoste**, che non potevano essere misurate, che determinavano in anticipo se il fotone avrebbe attraversato il polarizzatore o no, cioè in quale stato quantistico il sistema si sarebbe trovato dopo la misura.

1935: IL PARADOSSO EPR

Einstein, Podolsky e Rosen nel 1935 pubblicarono nella rivista *Physical Review* un importantissimo articolo dal titolo: *Can Quantum-Mechanical Description of Physical Reality be Considered Complete?* [5]. In quattro pagine descrivevano un possibile esperimento che avrebbe messo a dura prova le fondamenta della meccanica quantistica per i successivi trent'anni.

L'articolo inizia con la definizione di cosa possa considerarsi reale e cosa completo in fisica; per citare il lavoro originale: «Una teoria è completa se ogni elemento della realtà fisica ha una sua controparte nella teoria». Questo sarà, come vedremo, il punto contestato dai tre scienziati nei confronti della meccanica quantistica.

Il concetto di realtà è più sottile da definire, bisogna intanto dire che ciò che è reale non può essere stabilito a priori con un ragionamento filosofico ma deve

emergere da esperimenti e misure. Nell'articolo Einstein, Podolsky e Rosen danno la seguente definizione: «Se, non disturbando il sistema in alcun modo, possiamo predire con certezza, il valore di una quantità fisica, allora deve esistere un elemento della realtà fisica corrispondente a quella quantità».

Dopo questa premessa di carattere epistemologico viene descritto un esperimento mentale, di cui qui ne viene fornita una versione più intuitiva formulata da David Bohm e Aharonov [6], rispetto a quella dell'articolo originale, ma che porta allo stesso risultato.

Immaginiamo una molecola che abbia spin[2] totale pari a zero e che sia formata da due atomi che chiameremo 1 e 2. Affinché lo spin totale della molecola sia zero, i due atomi devono avere valore dello spin uguale e opposto in ogni direzione. Ricordiamo che gli spin lungo due assi diversi (per esempio lungo x e y) sono soggetti al principio di indeterminazione, come vale per la quantità di moto e la posizione, perciò non possono essere determinati contemporaneamente.

Il sistema molecola, per il principio di sovrapposizione si trova, rispetto a una particolare direzione prescelta, in una combinazione di due stati: quello in cui l'atomo 1 ha spin positivo e il 2 negativo e quello in cui l'atomo 1 ha spin negativo e il 2 positivo[3].

$$\Psi = \frac{1}{\sqrt{2}}[\psi_1(\uparrow)\psi_2(\downarrow) + \psi_1(\downarrow)\psi_2(\uparrow)] \tag{5}$$

dove tra parentesi abbiamo indicato la direzione dello spin.

Immaginiamo ora di riuscire a dividere la molecola nei due atomi senza alterare il loro spin e di mandare l'atomo 1 all'osservatrice Alice e l'atomo 2 all'osservatore Bob. Nonostante siano divisi, poiché il valore totale dello spin della molecola era inizialmente pari a 0, avranno ancora spin uguali e opposti lungo ogni asse.

Immaginiamo ora che Alice misuri lo spin dell'atomo che riceve lungo l'asse z e ottenga un valore positivo ($\uparrow$), allora Bob, prima ancora di effettuare la misura, se Alice glielo comunica, saprà di trovare lo spin della sua molecola lungo z orientato negativamente ($\downarrow$). Se invece Alice avrà misurato lungo l'asse z uno spin $\downarrow$, Bob sarà sicuro di trovare uno spin $\uparrow$.

---

[2] Lo spin è un grado di libertà interno delle particelle, è un vettore e quindi ha componenti lungo le tre direzioni x, y e z. È privo di un analogo nella meccanica classica e per il caso che stiamo considerando può assumere lungo ogni asse solo valori pari a $\frac{\hbar}{2}$ o $-\frac{\hbar}{2}$.

[3] Il termine $\frac{1}{\sqrt{2}}$ è inserito per rispettare la condizione di normalizzazione.

Secondo la definizione di realtà che abbiamo dato più sopra, lo spin lungo l'asse z dell'atomo di Bob è reale, poiché in base alla misura di Alice, senza interferire con il sistema, è possibile determinare con certezza il valore di quella grandezza. La misura lungo l'asse z renderebbe inoltre indeterminati gli spin lungo gli altri assi, a causa del principio di indeterminazione, perciò non prevedibili con certezza e quindi non reali. Nulla vieta, tuttavia, ad Alice, nell'esperimento ideato, di misurare lo spin lungo l'asse x o lungo l'asse y, anziché lungo z. In questo caso lo spin dell'atomo 2, avrebbe assunto valore di realtà lungo quegli assi.

Secondo Einstein, Podolsky e Rosen, la misura effettuata da Alice non può in alcun modo condizionare l'atomo di Bob, determinando lo spin lungo un asse e rendendolo indefinito lungo gli altri, perché ciò andrebbe contro la relatività ristretta, in quanto l'informazione viaggerebbe istantaneamente dall'atomo di Alice a quello di Bob, superandolo la velocità della luce. Allora, concludono i tre scienziati, lo spin lungo ciascuno dei tre assi doveva essere reale e determinato fin dall'inizio, prima ancora che Alice faccia la sua misura, perché la sua decisione di misurare lungo un asse piuttosto che un altro non può influenzare la realtà dello spin dell'altro atomo. Ma ciò è vietato dal principio di indeterminazione che stabilisce che solo lo spin lungo una direzione possa essere determinato.

L'unico modo per risolvere il problema, secondo l'articolo originale, è ammettere che la meccanica quantistica non è completa, cioè che la funzione d'onda non descrive completamente il sistema, ma esistono delle *variabili nascoste* che se conosciute potrebbero permettere di determinare lo spin anche lungo le altre direzioni con certezza.

Questo esperimento mentale prese il nome di paradosso EPR, ma come poteva essere risolto? A ben guardare il ragionamento di Einstein, Podolsky e Rosen sembra inattaccabile, ma è proprio così? C'è un'assunzione di base nel paradosso, che è nascosta nella frase: «ciò andrebbe contro la relatività ristretta, in quanto l'informazione viaggerebbe istantaneamente dall'atomo di Alice a quello di Bob». È davvero necessario che affinché ciò che accade in un punto dello spazio (la misurazione dello spin dell'atomo 1) influenzi ciò che accade in un altro ci debba essere un'informazione che viaggia tra i due luoghi?

La risposta a questa domanda non è banale: essa infatti non è categoricamente sì, ma è tale solo se viviamo in un universo locale, cioè in cui due oggetti distanti non possono avere influenza istantanea l'uno sull'altro. In un universo non locale, due sistemi per influenzarsi non devono necessariamente scambiarsi un'informazione, e quindi anche se l'influenza fosse istantanea non ci sarebbe superamento della

velocità della luce perché nessun segnale viaggerebbe tra i due. Einstein non mise nemmeno in considerazione questa possibilità, perché era un convinto sostenitore della località della realtà in cui viviamo.

Ma siamo veramente sicuri di vivere in un universo in cui vige il cosiddetto realismo locale[4] e in cui la meccanica quantistica risulterebbe allora incompleta? È questa la vera domanda da porsi dunque.

1964: LE DISUGUAGLIANZE DI BELL

Su questo punto si sviluppò un'accesissima discussione fra sostenitori delle variabili nascoste, inserite con lo scopo di completare la descrizione della meccanica quantistica e sostenitori della sua completezza. Il dibattito rimase però insoluto fino al 1964 quando John Stewart Bell, un fisico irlandese, pubblicò su *Physics* un breve articolo di sei pagine dal titolo: *On the Einstein Podolsky Rosen Paradox* [7] che presentava un interessante teorema riguardante la questione del paradosso EPR. Il lettore non interessato ai dettagli del teorema può saltare direttamente alla fine del paragrafo, tenendo a mente che ciò che Bell dimostra è che se esiste una teoria con delle variabili nascoste che riproduce gli stessi risultati della meccanica quantistica, è necessario che valga la disuguaglianza (11).

Nell'articolo originale Bell utilizza l'esperimento mentale proposto da Bohm e che è stato riportato più sopra. Immaginiamo che le particelle 1 e 2 abbiano spin rispettivamente $\vec{\sigma}_1$ e $\vec{\sigma}_2$, e che il loro spin venga misurato lungo un asse, diciamo $\hat{a}$, allora se la misura della componente dello spin della particella 1 lungo l'asse $\hat{a}$ (matematicamente si indica come $\vec{\sigma}_1 \cdot \hat{a}$) risulta pari a 1[5], allora $\vec{\sigma}_2 \cdot \hat{a}$ risulterà pari a 1 e viceversa. Poiché abbiamo visto che la funzione d'onda non prevede il risultato della singola misura, è possibile aggiungere delle specifiche per rendere completa la descrizione dello stato. Supponiamo che questa maggiore specificità dipenda dalla media di alcuni parametri λ. Non è importante cosa λ rappresenti, se un solo parametro, un set di parametri o di funzioni, il risultato che otterremo è indipendente da ciò. L'esito A della misura $\vec{\sigma}_1 \cdot \hat{a}$ non dipende allora solo da $\hat{a}$ ma dipenderà anche da λ. Allo stesso modo l'esito B della misura $\vec{\sigma}_2 \cdot \hat{b}$ lungo l'asse $\hat{b}$ dipenderà

[4] Il realismo locale unisce il principio di località, enunciato precedentemente, con quello di realismo, secondo il quale tutti gli oggetti debbano possedere dei valori preesistenti per ogni possibile misurazione prima che queste misurazioni vengano effettuate

[5] In unità di $\frac{\hbar}{2}$

da $\hat{b}$ e λ (per indicare che B dipende da quei due parametri scriviamo B($\hat{b}, \lambda$). Ricordando che il valore della proiezione dello spin lungo un asse può essere positivo (allora alla misura assegneremo esito +1) o negativo (−1) otteniamo che:

$$A(\hat{a},\lambda) = \pm 1 \quad B(\hat{b},\lambda) = \pm 1 \tag{6}$$

dove A indica sempre una misura effettuata sulla prima particella, B sulla seconda. Chiamando $\rho(\lambda)$ la distribuzione di probabilità[6] di λ avremo che il valore di aspettazione (il valore medio) del prodotto delle due componenti $\vec{\sigma}_1 \cdot \hat{a}$ e $\vec{\sigma}_2 \cdot \hat{b}$ varrà[7]:

$$P\left(\hat{a},\hat{b}\right) = \int d\lambda \rho(\lambda) A(\hat{a},\lambda) B\left(\hat{b},\lambda\right) \tag{7}$$

Questo dovrebbe essere in accordo con il valore di aspettazione fornito dalla meccanica quantistica che è pari a $-\hat{a} \cdot \hat{b}$. Poiché $\rho(\lambda)$ è una distribuzione di probabilità vale per definizione che:

$$\int \rho(\lambda) d\lambda = 1$$

Poiché, a causa dell'equazione (6), sappiamo che il prodotto fra $A(\hat{a}, \lambda)$ e $B(\hat{b}, \lambda)$ può valere −1 o 1, allora l'integrale nell'equazione (7) varrà al minimo −1 e al massimo +1. Se scegliamo $\hat{a} = \hat{b}$, allora stiamo misurando gli spin lungo lo stesso asse, e i loro valori, come ribadito più volte nell'esperimento che stiamo considerando, devono essere opposti, perciò $A(\hat{a}, \lambda) = -B(\hat{a}, \lambda)$ e analogamente $B(\hat{b}, \lambda) = -A(\hat{b}, \lambda)$. Inserendo questo risultato nell'equazione (7) otteniamo:

$$P(\hat{a}, \hat{b}) = -\int d\lambda \rho(\lambda) A(\hat{a}, \lambda) A(\hat{b}, \lambda) \tag{8}$$

consideriamo ora un altro asse $\hat{c}$ lungo il quale potrebbe avvenire la misura e calcoliamo $P(\hat{a}, \hat{b}) - P(\hat{a}, \hat{c})$:

$$P(\hat{a}, \hat{b}) - P(\hat{a}, \hat{c}) = -\int d\lambda \rho(\lambda) A(\hat{a}, \lambda) A(\hat{b}, \lambda) + \int d\lambda \rho(\lambda) A(\hat{a}, \lambda) A(\hat{c}, \lambda) =$$
$$= \int d\lambda \rho(\lambda) A(\hat{a}, \lambda) \left[A(\hat{c}, \lambda) - A(\hat{b}, \lambda)\right]$$

---

[6] L'integrale della distribuzione di probabilità $\rho(x)$ di una variabile aleatoria x fra due estremi a e b rappresenta la probabilità che la variabile x sia compresa fra a e b:

$$P(a \le x \le b) = \int_a^b \rho(x) dx$$

[7] Una nota tecnica: integrali in cui non indichiamo gli estremi sono da intendersi estesi a tutto il dominio di definizione della funzione.

Poiché il valore della misura può essere ±1 allora avremo che $A(\hat{b}, \lambda)^2$ dovrà essere uguale a 1, perciò possiamo scrivere l'equazione sopra come:

$$P(\hat{a}, \hat{b}) - P(\hat{a}, \hat{c}) = \int d\lambda \rho(\lambda) A(\hat{a}, \lambda) A(\hat{b}, \lambda) [A(\hat{c}, \lambda) A(\hat{b}, \lambda) - 1] \qquad (9)$$

Possiamo ora trovare un limite superiore al modulo di $P(\hat{a}, \hat{b}) - P(\hat{a}, \hat{c})$, ciò che otteniamo è che[8]:

$$|P(\hat{a}, \hat{b}) - P(\hat{a}, \hat{c})| \leq \int d\lambda \rho(\lambda) [1 - A(\hat{c}, \lambda) A(\hat{b}, \lambda)] \qquad (10)$$

Ma vediamo dall'equazione (8) che il secondo termine nella parentesi quadra è esattamente $P(\hat{b}, \hat{c})$ e perciò:

$$|P(\hat{a}, \hat{b}) - P(\hat{a}, \hat{c})| \leq 1 + P(\hat{b}, \hat{c}) \qquad \textbf{Disuguaglianza di Bell} \qquad (11)$$

Questa disequazione prende il nome di disuguaglianza di Bell e deve valere se esistono variabili nascoste di carattere locale. Tuttavia assumendo come valore di aspettazione quello predetto dalle meccanica quantistica: $P(\hat{a}, \hat{b}) = -\hat{a} \cdot \hat{b}$ troviamo che la disuguaglianza è falsa, infatti scegliendo per esempio $\hat{a}$ e $\hat{b}$ perpendicolari fra loro e $\hat{c}$ inclinato di 45° rispetto ad entrambi, si ottiene: $0.707 \leq 1 - 0.707$ che è palesemente falsa. Di conseguenza una teoria che incorpori variabili nascoste locali non è compatibile con le predizioni della meccanica quantistica. Resta dunque da stabilire sperimentalmente se la disuguaglianza di Bell è violata, oppure risulta sempre vera.

1975: L'ESPERIMENTO DI ASPECT

L'articolo scritto da Bell destò subito l'attenzione della comunità dei fisici, perché offriva una soluzione definitiva al paradosso EPR verificabile sperimentalmente.

Nel 1969 uscì un articolo di Clauser e altri [8] in cui si dimostrava che erano tecnicamente possibili esperimenti che potessero verificare le disuguaglianze, tuttavia si incontrarono varie difficoltà: nel 1972 le università di Berkeley e Harvard condussero indipendentemente l'esperimento e ottennero risultati contrapposti,

---

[8] Infatti $|P(\hat{a}, \hat{b}) - P(\hat{a}, \hat{c})| = | = \int d\lambda \rho(\lambda) |A(\hat{a}, \lambda) A(\hat{b}, \lambda) [A(\hat{c}, \lambda) A(\hat{b}, \lambda) - 1]| \leq \int | d\lambda \rho(\lambda) A(\hat{a}, \lambda) A(\hat{b}, \lambda) [A(\hat{c}, \lambda) A(\hat{b}, \lambda) - 1] | = \int d\lambda \rho(\lambda) |A(\hat{a}, \lambda) A(\hat{b}, \lambda)| |[A(\hat{c}, \lambda) A(\hat{b}, \lambda) - 1]|$ poiché $\rho(\lambda)$ è sempre maggiore di 0. Inoltre visto che A vale 1 o −1 e il termine tra parentesi quadre è sempre minore o uguale a 0 si ottiene il risultato in (10).

i primi una violazione delle disuguaglianze, mentre i secondi la loro validità e di conseguenza una contraddizione delle predizioni della meccanica quantistica. Entrambi i risultati tuttavia furono ritenuti inaffidabili, poiché le fonti di particelle *entangled*[9] non erano ottimali.

Un altro problema fondamentale di questi esperimenti era l'impossibilità di escludere che un segnale, a velocità inferiore o uguale a quella della luce, si propagasse fra le due particelle, poiché la velocità con cui venivano eseguite le misure e la loro distanza spaziale non erano sufficienti ad escludere questa possibilità. In sostanza non si poteva eliminare l'evenienza che nel momento della misurazione un segnale viaggiasse a velocità uguale o inferiore a quella della luce da una particella all'altra, comunicando lungo quale direzione e in che modo dovesse orientarsi lo spin.

Bisognò aspettare il 1975, quando l'allora 28enne Alain Aspect, propose un esperimento [9] sufficientemente meticoloso da determinare se le diseguaglianze di Bell fossero violate o meno. Gli esperimenti precedenti si basavano sull'assunzione che valesse il principio di località di Bell (che impedisce a dispositivi di misura separati o a particelle separate di scambiarsi informazioni), una richiesta molto forte, che può però essere resa più debole.

L'esperimento di Aspect, infatti, richiede un'ipotesi meno stringente: i dispositivi, o le particelle, possono trasmettersi l'informazione, purché lo facciano a velocità inferiore o uguale a quella della luce (altrimenti la teoria sarebbe in disaccordo con la relatività ristretta). Questa richiesta prende il nome di principio di separabilità di Einstein, e per verificare l'infrazione delle disuguaglianze è necessario fare in modo che sperimentalmente questa possibilità non si realizzi. Nell'esperimento, al posto dello spin, Aspect proponeva l'utilizzo della polarizzazione dei fotoni, che analogamente allo spin può essere assumere due valori.

Il motivo della scelta era dovuto al fatto che sperimentalmente è più semplice produrre coppie di fotoni correlati che atomi con le caratteristiche dell'esperimento di Bohm. La polarizzazione di un fotone è misurabile tramite un polarizzatore. Il problema negli esperimenti precedenti era legato al loro posizionamento: erano infatti o tenuti in direzione fissa, o, se le loro direzioni erano mutate nel tempo, ciò avveniva troppo lentamente per escludere che i polarizzatori potessero comunicare fra loro tramite un segnale a velocità pari o inferiore a quella della luce.

---

[9] Chiamiamo così le particelle correlate, prodotte, nel nostro caso, nel singoletto di spin di cui abbiamo parlato più volte.

Aspect per ovviare questo problema propose l'uso di commutatori ottici. L'apparato sperimentale proposto è quello rappresentato in Figura 1.

Una sorgente **s** emette fotoni *entangled*, in figura chiamati ν_a e ν_b, il fotone ν_a incontra poi il commutatore ottico C_a, in base alla cui orientazione (che varia continuamente nel tempo), viene mandato o verso il polarizzatore $I_1(a_1)$ o verso il polarizzatore $I_2(a_2)$ i quali lo lasciano passare o meno in base alla sua polarizzazione (è qui che avviene quindi la misura). L'eventuale passaggio del fotone attraverso il polarizzatore viene poi rivelato da un fotomoltiplicatore posto dietro di esso.

Il colpo di genio di Aspect fu nell'evitare la problematica dello scambio di informazioni fra i vari polarizzatori non tramite il cambiamento della loro orientazione, ma appunto grazie all'utilizzo dei commutatori ottici: questi cambiano orientazione in maniera stocastica e indipendente l'uno dall'altro, di modo però che due stati dello stesso commutatore separati da un tempo maggiore di $\frac{L}{c}$ (dove L è la distanza fra i due commutatori, ed è quindi il tempo che impiegherebbe la luce a percorrere la loro distanza) siano statisticamente indipendenti.

Questo era l'esperimento che Aspect aveva progettato, di cui ne realizzò varie versioni, la prima nel 1980 e, quella ritenuta più affidabile, nel 1982.

La misurazione della polarizzazione di un fotone lungo una direzione qualunque $\vec{a}$ può avere due esiti: $+1$ se il fotone è parallelo ad $\vec{a}$, -1 se è perpendicolare.

Nell'ultima versione dell'esperimento Aspect misurò il valore di S, una variabile definita come:

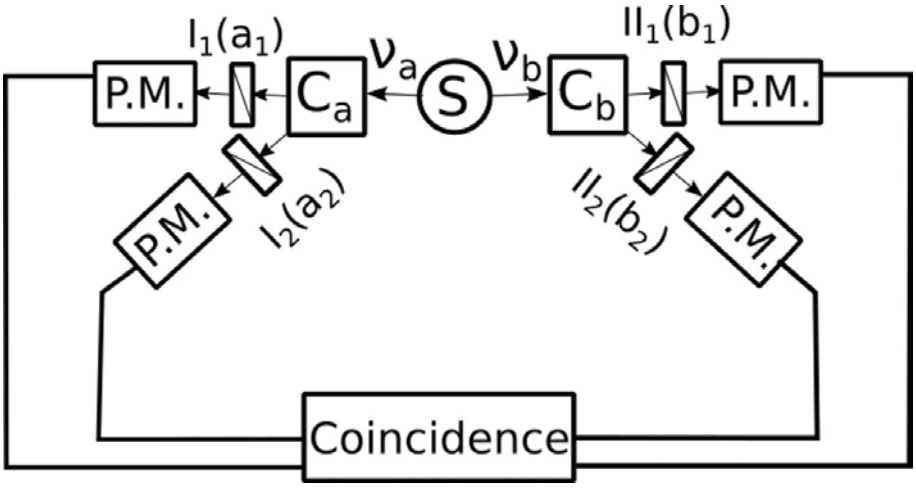


FIG. 1 L'apparato sperimentale proposto da Aspect nell'esperimento: s è la sorgente di fotoni, C_a e C_b i commutatori, i rettangoli barrati **I** e **II** i polarizzatori, **P.M.** indica la presenza di un fotomoltiplicatore che serve a rivelare il passaggio di un fotone.

$$S = E\left(\vec{a},\vec{b}\right) - E\left(\vec{a},\vec{b'}\right) + E\left(\vec{a'},\vec{b}\right) + E\left(\vec{a'},\vec{b'}\right) \tag{12}$$

ove

$$E\left(\vec{a},\vec{b}\right) = P_{++}\left(\vec{a},\vec{b}\right) + P_{--}\left(\vec{a},\vec{b}\right) - P_{+-}\left(\vec{a},\vec{b}\right) - P_{-+}\left(\vec{a},\vec{b}\right) \tag{13}$$

in cui $P_{xy}\left(\vec{a},\vec{b}\right)$ indica la probabilità che la misurazione della polarizzazione lungo $\vec{a}$ risulti x e lungo $\vec{b}$ risulti y.

I risultati furono pubblicati su *Physical Review Letters* in un articolo dal titolo: *Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities* [10]. Aspect misurò una violazione delle disuguaglianze di Bell con una precisione incredibile; il valore di aspettazione previsto dalla meccanica quantistica per il suo esperimento era $S = 2.70 \pm 0.05$, il valore sperimentale risultò $S = 2.697 \pm 0.015$, mentre assumendo l'esistenza di variabili nascoste locali si dovrebbe avere $S \leq 2$. Le disuguaglianza di Bell risultavano dunque violate e la meccanica quantistica ne usciva per l'ennesima volta confermata nelle sue predizioni. Ma questo implicava necessariamente rinunciare alle variabili nascoste e al realismo locale?

L'esperimento di Aspect lasciava aperte alcune scappatoie (il termine tecnico è *loopholes*):

1. nell'esperimento solo una piccola parte dei fotoni prodotti venne misurata; si può immaginare allora che il campione considerato non fosse rappresentativo, e che se la misura fosse avvenuta su tutti i fotoni prodotti le disuguaglianze non sarebbero state violate;
2. il cambio di direzione dei commutatori ottici non era completamente casuale e quindi la separabilità non era appieno garantita.

2016: THE BIG BELL TEST

Dalla spinta iniziale che diede Aspect gli esperimenti continuarono in maniera sempre più raffinata. Nel 1998, Zeilinger [11] e altri elaborarono un esperimento in cui la separabilità era raggiunta grazie ad una notevole distanza fisica tra le stazioni di misurazione, una scelta casuale e ultraveloce dei modulatori e con una registrazione dei dati effettuata in maniera indipendente. Nel 2001 Wineland [12] chiuse il primo *loophole* realizzando un esperimento in cui tutte le coppie di fotoni

venivano misurate e confermando la violazione delle disuguaglianze. I test di Bell proseguirono, cercando di raggiungere un grado di casualità sempre maggiore, fino al Big Bell Test [13], realizzato nel 2016 e i cui risultati furono pubblicati nel 2018. Il team dell'esperimento reclutò centomila persone per giocare ad un videogame in cui attraverso più livelli bisognava digitare, nel modo più casuale possibile, stringhe di 0 e 1. Il 30 Novembre 2016 per 12 ore il gioco fu attivo e i risultati prodotti dai giocatori vennero inviati in tempo reale a 13 laboratori in tutto il mondo in cinque continenti diversi, e utilizzati per determinare le condizioni di lavoro nei vari esperimenti (i bit prodotti, in base al contenuto, potevano muovere *phase shifter*, *polarization controller* e altri strumenti). I risultati del Big Bell Test dall'articolo originale sono riportati nella tabella più sotto.

Lead institution	Location	Entangled system	Rate (bps)	Inequality	Result	Stat sign
Griffith University	Brisbane	Photon Polarization	4	$S_{16} \leq 0.511$	$S_{16} = 0.965 \pm 0.008$	57σ
UQ & EQUS	Brisbane	Photon Polarization	3	$\|S\| \leq 2$	$S_{AB} = 2.75 \pm 0.05$ $S_{BC} = 2.79 \pm 0.05$	15σ 16σ
USTC	Shanghai	Photon Polarization	10^3	$PRLGB^{30}$	$I_0 = 0.10 \pm 0.05$	N\A
IQOQI	Vienna	Photon Polarization	$1.61 \cdot 10^3$	$\|S\| \leq 2$	$S_{HRN} = 2.639 \pm 0.008$ $S_{QRN} = 2.643 \pm 0.006$	81σ 116σ
Sapienza	Rome	Photon Polarization	0.62	$B \leq 1$	$B = 1.225 \pm 0.007$	32σ
LMU	Munich	Photon-atom	1.7	$\|S\| \leq 2$	$S_{HRN} = 2.427 \pm 0.0223$ $S_{QRN} = 2.413 \pm 0.0223$	19σ 18.5σ
ETHZ	Zurich	Transmon qubit	$3 \cdot 10^3$	$\|S\| \leq 2$	$S = 2.3066 \pm 0.0012$	$P < 10^{-99}$
INPHYNI	Nice	Photon time bin	$2 \cdot 10^3$	$\|S\| \leq 2$	$S = 2.431 \pm 0.003$	140σ
ICFO	Barcelona	Photon-atom ensemble	125	$\|S\| \leq 2$	$S = 2.29 \pm 0.010$	2.9σ
ICFO	Barcelona	Photon multi-frequency bin	20	$\|S\| \leq 2$	$S = 2.25 \pm 0.08$	3.1σ
CITEDEF	Buenos Aires	Photon polarization	1.02	$\|S\| \leq 2$	$S = 2.55 \pm 0.07$	7.8σ
UdeC	Concepción	Photon time bin	$5.2 \cdot 10^4$	$\|S\| \leq 2$	$S = 2.43 \pm 0.02$	20σ
NIST	Boulder	Photon polarization	10^5	$K \leq 0$	$K = (1.65 \pm 0.20) \cdot 10^{-4}$	8.7σ

Com'è evidente dai risultati le disuguaglianze sono violate senza ombra di dubbio in tutti i casi con una certezza impressionante.

È possibile individuare un ulteriore *loophole* dovuto all'assunto nel teorema di Bell che non vi sia correlazione tra la scelta del setting degli apparati sperimentali e qualunque altra cosa che possa influenzare gli esiti delle misure, la proprietà chiamata indipendenza statistica.

Nel 2017, per chiudere questo *loophole* è stato realizzato il *Cosmic Bell Test*, in cui gli apparati di misura sono stati impostati usando le lunghezze d'onda dei fotoni emessi da due stelle distanti migliaia di anni luce nella Via Lattea e causalmente disconnesse al momento dell'emissione dei fotoni. I risultati di questo esperimento confermano a loro volta, in maniera estremamente precisa, la violazione delle disuguaglianze di Bell.

Tutti i *loopholes* sono chiusi e sembra proprio necessario rinunciare al realismo locale. O no? Le variabili nascoste non sono del tutto escluse, possono esistere se sono non locali, la strada presa da Bohm e de Broglie nella loro interpretazione della meccanica quantistica. A differenza delle variabili locali tuttavia, la mag-

gior parte di questi modelli prevede fenomeni che contraddicono la relatività ristretta, in particolare le informazioni viaggiano più velocemente di *c*. Si potrebbe obiettare che anche la non località lo faccia, ma il problema è stato risolto tempo fa: il cosiddetto *no communication theorem* assicura che tramite l'entanglement non è possibile comunicare, cioè ciò che avviene in un punto dello spazio può influenzare istantaneamente un altro punto, ma con questo fenomeno non si possono trasmettere informazioni. In sostanza la non località non prevede e non permette il passaggio di un'informazione (o di un segnale) e non contraddice quindi la relatività ristretta.

A FAREWELL TO FREE WILL?

Assumiamo sempre implicitamente la libertà dello sperimentatore... Questa assunzione fondamentale è essenziale per fare scienza. Se questo non fosse vero, sono dell'idea che non avrebbe nessun senso porre interrogativi alla natura tramite esperimenti, poiché la natura potrebbe determinare quali siano le nostre domande, e guidarle in maniera che giungiamo ad una falsa immagine di essa

Anton Zeilinger, *Dance of the photons*

Challenging local realism with human choices, sfidare il realismo locale grazie alle scelte umane, questo il titolo dell'articolo originale del Big Bell Test, scelte umane che devono essere libere o, per lo meno, non predeterminate.

Questa ipotesi, a primo sguardo molto blanda, apre un ulteriore *loophole*, finora taciuto per la sua finezza, il cosiddetto *loophole* della libera scelta: affinché quanto abbiamo detto finora valga, le scelte degli sperimentatori non devono essere predeterminate.

Prima di affrontare questo *loophole* però, riavvolgiamo di un poco il nastro della nostra storia, e torniamo al 2006, quando John Conway e Simon Kochen pubblicarono su *Foundations of Physics* un articolo dal titolo: *The Free Will Theorem* [16].

Immaginiamo un esperimento con una particella di spin totale pari a 1, in cui misuriamo il quadrato di una sua componente lungo una direzione *w*, (il valore che potrà assumere sarà quindi 0 o 1). Chiamiamo *triplo esperimento*, la misurazione di tale valore lungo gli assi *x, y, z* e il valore misurato lungo ognuna di esse *j, k, l* rispettivamente. Per la validità del teorema devono essere verificate le seguenti tre condizioni.

- SPIN: un triplo esperimento lungo gli assi *x*, *y*, *z* dà come risultato 1, 0, 1 in un certo ordine.

- TWIN: è possibile produrre due particelle di spin 1 separate nello spazio, tali che, se il risultato di un triplo esperimento sulla prima fosse $s_x^2 = j, s_y^2 = k, s_z^2 = l$, allora il medesimo risultato si otterrebbe per un triplo esperimento sulla seconda.
- FIN: c'è un limite superiore alla velocità con cui le informazioni possono essere effettivamente trasmesse.

Mentre SPIN e TWIN sono verificate sperimentalmente, FIN è un postulato della relatività ristretta. Se le tre condizioni sono verificate, allora, se la scelta delle direzioni in cui effettuare gli esperimenti sulle particelle di spin 1, non è una funzione delle informazioni accessibili agli sperimentatori (cioè se le loro scelte non sono predeterminate), allora anche i risultati degli esperimenti non sono funzione delle informazioni a loro accessibili.

Nel 2009 fu elaborata una versione più forte del teorema [17], ma che ricevette delle critiche [18], questi dettagli matematici tuttavia esulano dallo scopo di questo scritto, e non saranno trattati, sebbene riportati per completezza.

Il teorema del libero arbitrio ci dice qualcosa in più. Infatti, se esso vale, allora è anche vero che se gli esiti degli esperimenti sono predeterminati, anche le scelte degli sperimentatori sono predeterminate. Questo ci porta direttamente ad affrontare il *loophole* della libera scelta. Il libero arbitrio come possibilità di scelta assoluta, senza limiti, è esclusa dalla fisica per come la conosciamo ora: non esiste nessuna legge fisica che permetta o contempli l'esistenza del libero arbitrio.

Quello che può esistere come libero arbitrio è l'impossibilità di prevedere in anticipo le azioni di una persona, non tanto per una conoscenza imperfetta dei dati necessari a prevederle, ma per l'intrinseco carattere probabilistico della meccanica quantistica.

Dal teorema del libero arbitrio sappiamo però che se gli esiti degli esperimenti sono predeterminati allora lo sono anche le azioni degli sperimentatori, negando di conseguenza anche l'esistenza del libero arbitrio come concepito nella seconda definizione data sopra. Questa strada prende il nome di superdeterminismo: il cosmo non ha possibilità di scelta, l'esito di ogni esperimento e azione umana è predeterminata dalle condizioni iniziali dell'universo.

Questa teoria, a prima vista estrema, ha dei pregi indiscutibili: salva il realismo locale e risolve in parte il problema della misura nella meccanica quantistica. Il prezzo da pagare è accettare che dalle condizioni iniziali dell'universo ogni cosa sia predeterminata e che non valga l'indipendenza statistica, una delle ipotesi del teorema di Bell, cioè che la distribuzione delle variabile nascoste non sia indipendente dal settaggio dell'apparato sperimentale. Tornando all'esperimento EPR, questo significa che Alice non avrebbe

potuto scegliere diversamente l'orientazione lungo cui misurare lo spin, perché essa era già predeterminata e quindi l'obiezione di Einstein, Podolsky e Rosen viene meno.

Varie critiche sono state avanzate al superdeterminismo, una soluzione a ciascuna di esse e un approfondimento di questa teoria si può ritrovare nell'articolo di Hossenfelder e Palmer: *Rethinking Superdeterminism*, pubblicato nel maggio 2020 [19].

Ritorniamo però al 2017, all'esperimento con la luce cosmica emessa dalle stelle della Via Lattea: non si era detto che escludeva ogni correlazione statistica e non dovrebbe, di conseguenza, rendere invalido il superdeterminismo? La risposta è no, l'articolo stesso conclude che per avere una correlazione statistica nell'esperimento, il meccanismo a variabili nascoste, sfruttando il *loophole* della libera scelta, dovrebbe aver agito prima dei tempi dell'invenzione della stampa da parte di Gutenberg, cioè più di sei secoli fa. Questo non è un problema per il superdeterminismo, per cui ogni evento è predeterminato dalle condizioni iniziali da cui si è evoluto l'Universo, ben prima quindi dell'invenzione della stampa di Gutenberg.

Resta la critica di Anton Zeilinger, riportata ad inizio paragrafo, ma quanto può avere veramente senso? La natura potrebbe anche averci portato ad avere una falsa immagine di essa, ma è davvero falso ciò che appare vero ogni volta che facciamo un esperimento? Se anche la natura ci avesse guidato a questa falsa immagine, in cosa sarebbe falsa, se essa si riproduce sempre con precisione?

La fisica non può e non deve avere fini ontologici, i *perché* della fisica sono in realtà dei *come*, i nostri modelli non incarnano la realtà, ma cercano di spiegarla e prevederla.

Siamo giunti alla fine di questo viaggio, la cui conclusione potrebbe averci molto più sconvolto di quanto sconvolse Einstein e gli altri fisici della prima metà del Novecento. Al momento non esistono esperimenti in grado di dire se la strada giusta da intraprendere sia quella del superdeterminismo o quella del divorzio fra realtà e località, ma fino a cinquant'anni fa nessuno avrebbe pensato che sarebbe stato possibile usare la luce di stelle lontane qualche migliaio di anni luce per determinare il settaggio di un apparato sperimentale, fino a trent'anni fa nessuno avrebbe pensato che fosse possibile realizzare un esperimento in cui centinaia di migliaia di persone inviavano dati a dei laboratori sparsi per tutto il globo giocando online.

Eppure lo abbiamo fatto.

Viviamo in un secolo in cui la fisica teorica procede molto più velocemente di quanto gli esperimenti possano fare, ma come abbiamo imparato da questa storia, gli esperimenti che oggi ci sembrano irrealizzabili, sono pronti ad essere le rivoluzioni scientifiche di domani.

Occorre aspettare, ma per conoscere la natura della realtà che viviamo, aspettare ne vale la pena.

RIFERIMENTI BIBLIOGRAFICI

[1] P.A.M. Dirac, *The Principles of Quantum Mechanics*, Oxford University Press, 1958.

[2] D. Finkelstein, *The logic of quantum physics*, Trans. N.Y. Acad. Sci., 25, 621, 1963.

[3] J. Hartle, *Quantum Mechanics of Individual Systems*, Amer. J. Phys. 36, 704, 1968.

[4] B.S. De Witt and N. Graham, Eds., *The Many-Worlds Interpretation of Quantum Mechanics,* Princeton University Press, 1973.

[5] A. Einstein, B. Podolsky, N. Rosen, *Can Quantum-Mechanical Description of Physical Reality be Considered Complete?*, Physical Review, 47, 1935, pp. 777-780.

[6] D. Bohm, Y. Aharonov, *Discussion of Experimental Proof for the Paradox of Einstein, Rosen and Podolsky*, Physical Review, 108(4), 1957, pp. 1070-1076.

[7] J.S. Bell, *On the Einstein Podolsky Rosen Paradox*, Physics, 1(3), 1964, pp. 195-200.

[8] J.F. Clauser, M.A. Horne, A. Shimony, R.A. Holt, *Proposed Experiment to Test Local Hidden-Variable Theories*. Physical Review Letters, 23(15), 1969, pp. 880-884.

[9] A. Aspect, *Proposed Experiment to test the non separability of quantum mechanics*, Physical D, 14(8), 1976, pp. 1944-1951.

[10] A. Aspect, P. Grangier, G. Roger, *Experimental Realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A New Violation of Bell's Inequalities*, Phys. Rev. Lett., 49, 91, 1982.

[11] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, A. Zeilinger, *Violation of Bell's Inequality under Strict Einstein Locality Conditions*, Phys. Rev. Lett., 81, 5039, 1998.

[12] M.A. Rowe, D. Kielpinski, V. Meyer, C.A. Sackett, W.M. Itano, C. Monroe, D. J. Wineland, *Experimental violation of a Bell's inequality with efficient detection*, Nature, 409, 2001, pp. 791-794.

[13] The Big Bell Test collaboration, *Challenging local realism with human choices*, Nature, 557, 2018, pp. 212-216.

[14] J. Handsteiner *et al.*, *Cosmic Bell Test: Measurement Settings from Milky Way Stars*, Phys. Rev. Lett., 118, 060401, 2017.

[15] A. Zeilinger, *Dance of the Photons*, New York, Farrar, Straus and Giroux, 2010, p. 266.

[16] J.H. Conway, S. Kochen, *The Free Will Theorem*, Foundations of Physics, 36(10), 2006.

[17] J.H. Conway, S. Kochen, *The Strong Free Will Theorem*, Notices of the AMS, 56(2), 2009.

[18] S. Goldstein, D.V. Tausk, R.Tumulka, N. Zanghì, *What Does the Free Will Theorem Actually Prove?*, Notices of the AMS, 57(11), 2010.

[19] S. Hossenfelder, T. Palmer, *Rethinking Superdeterminism*, Front. Phys., 8, 2020.

LA RIVOLUZIONE IN ORBITA

DALLA GUERRA POLITICA AD UNO SPAZIO PER TUTTI

TOMMASO FADDA

L'economia dello spazio è ad oggi un vasto insieme di attività che riguardano l'esplorazione, la ricerca, e l'utilizzo dello spazio cosmico, sia attorno alla terra che interplanetario. Storicamente sono sempre stati gli attori pubblici ad avere il controllo dello spazio, ma oggi sono sempre di più le imprese private che partecipano sia attivamente, sia lavorando per conto delle agenzie pubbliche allo sfruttamento dell'orbita terrestre, creando un nuovo mercato ricco di opportunità e in costante evoluzione. In questo capitolo si analizzerà l'evoluzione di dell'accesso allo spazio, dall'iniziale sfida tra URSS e USA alla supremazia di questi ultimi, fino all'attuale privatizzazione dello spazio, resa necessaria per ridurre i costi dell'esplorazione e allo stesso tempo fonte di molte nuove opportunità per tutti gli attori dell'industria spaziale.

The space economy is today a vast set of activities concerning the exploration, research, and use of cosmic space, both around the earth and interplanetary. Historically, public actors have always been in control of space, but today more and more private companies participate in the exploitation of the Earth's orbit both actively and working on behalf of public agencies, creating a new market full of opportunities and in constant evolution. This chapter will analyze the evolution of access to space, from the initial challenge between the USSR and the US to the supremacy of the latter, up to the current privatization of space, made necessary to reduce the costs of exploration and at the same time source of many new opportunities for all players in the space industry.

PREFAZIONE

L'astrazione stenografica intorno a cui ruota il saggio di Fadda è il concetto di concorrenza, che viene studiato nel contesto del settore aerospaziale. La concorrenza è un concetto chiave nella teoria economica, a partire dalla *mano invisibile* di Adam Smith fino ai contributi della moderna economia industriale. La teoria economica ci insegna che la concorrenza beneficia i consumatori e la società nel suo complesso, consentendo di ridurre i prezzi e incrementando la varietà e la qualità dei prodotti. La riduzione dei prezzi a sua volta porta ad un aumento della platea dei consumatori che possono acquistare i beni e ad un incremento della produzione, generando effetti positivi sull'economia nel suo complesso.

Il lavoro di Fadda si concentra sugli effetti della concorrenza nel mercato aerospaziale, analizzando l'evoluzione del settore dalla metà del secolo scorso ai giorni nostri. Le caratteristiche tecnologiche e istituzionali di questo settore hanno fatto sì che, per molti decenni, assumesse le caratteristiche di un monopolio governativo, poiché le agenzie spaziali di Stati Uniti e Unione Sovietica erano di fatto gli unici attori in grado di mantenere un programma di esplorazioni spaziali. Fadda studia gli sviluppi recenti legati principalmente all'avvento di società private che hanno modificato lo scenario competitivo. I recenti successi della compagnia Space X nei lanci di veicoli spaziali e in particolare la riuscita della missione "Demo 2", cioè il primo volo spaziale con equipaggio per l'astronave del programma spaziale privato ideato da Elon Musk, sanciscono infatti l'inizio di una nuova era delle esplorazioni spaziali. Secondo l'analisi di Fadda, l'ingresso dei privati nel settore aerospaziale ha coinciso con una riduzione dei costi di lancio, che a sua volta ha portato all'entrata di nuove imprese operanti anche in altri settori. Pertanto, il principale canale attraverso cui si sono realizzati i benefici della concorrenza nell'ambito del mercato aerospaziale è quello della riduzione dei costi, ovvero una riduzione dell'inefficienza produttiva legata al precedente regime di monopolio.

Il saggio ripercorre la storia delle esplorazioni spaziali a partire dagli anni Cinquanta e Sessanta del secolo scorso, caratterizzati dal predominio dei programmi spaziali governativi degli Stati Uniti e dell'Unione Sovietica. Dopo il ridimensionamento dei programmi spaziali russi, l'intero settore aerospaziale è stato di fatto dominato dalla NASA, che deteneva il monopolio anche a livello tecnologico, grazie all'elevato budget governativo dedicato al settore. Il monopolio della NASA è stato caratterizzato da un elevato livello dei costi di lancio dei veicoli spaziali, di cui Fadda analizza nel dettaglio le ragioni. Da un lato, il saggio descrive gli aspetti di natura tecnica che hanno contribuito alla determinazione di tali costi. Da questo

punto di vista, il settore potrebbe configurarsi come un esempio di quello che nella letteratura economica viene descritto come *monopolio naturale*, ovvero un mercato in cui la posizione di monopolio emerge in conseguenza delle caratteristiche tecnologiche, caratterizzate dalla presenza di elevati costi fissi. In tali contesti, il mercato tende "naturalmente" al monopolio in quanto la duplicazione dei costi di produzione renderebbe inefficiente l'ingresso di nuove imprese. D'altra parte, Fadda individua nel regime di monopolio la ragione stessa dell'inefficienza produttiva, legata all'assenza di incentivi da parte del monopolista a ridurre i costi di produzione per mantenere la competitività rispetto ad eventuali rivali.

Fadda documenta la progressiva riduzione dei costi di lancio che si è verificata nell'ultimo decennio, attribuendone la ragione principalmente all'ingresso del mercato di aziende private che hanno rotto il monopolio della NASA e dei suoi storici partner commerciali (Boeing, Lockheed Martin e McDonnel Douglas). Negli ultimi anni, Space X è divenuto un partner di primo piano per la NASA, dimostrando di saper conciliare un'elevata affidabilità e la capacità di contenere i costi. Fadda analizza in dettaglio i vari canali di riduzione dei costi che hanno consentito a Space X di acquisire un vantaggio competitivo rispetto alle imprese concorrenti. In particolare, la struttura verticalmente integrata di Space X, che sviluppa internamente le componenti e i sottosistemi, ha consentito di ridurre i costi di gestione e le infrastrutture necessarie. Inoltre, Space X fa un uso maggiore rispetto ai concorrenti di procedure automatizzate, che velocizzano il processo di produzione. Più in generale, la chiave del suo successo sembra essere la semplicità e la snellezza delle procedure e delle tecnologie, che le hanno consentito di raggiungere livelli qualitativi elevati con costi contenuti.

Il lavoro di Fadda analizza infine gli effetti provocati dall'aumento della concorrenza nel settore aerospaziale. In particolare, l'incremento del numero di operatori ha portato ad un'ulteriore riduzione dei costi, che ha a sua volta portato ad un aumento dei voli spaziali e all'esplosione del mercato dei satelliti per scopi commerciali.

Gli sviluppi di questa "privatizzazione dello Spazio" sono ad oggi ancora difficili da prevedere. L'espansione dei privati nell'industria aerospaziale non sembra destinata ad arrestarsi. È notizia recente che la collaborazione tra la NASA e le aziende private, ed in particolare Space X, potrebbe svilupparsi ulteriormente in futuro, portando ad un inedito partenariato per una missione lunare.

Il contributo del saggio di Fadda è duplice. Da un lato, utilizza gli strumenti della teoria economica per analizzare un settore che è stato poco studiato nella letteratura. D'altro canto, Fadda riesce a coniugare la conoscenza della teoria economica con le competenze tecniche necessarie per analizzare le caratteri-

stiche delle tecnologie di produzione e la struttura dei costi in un settore che presenta una complessità tecnologica molto elevata. Da questo punto di vista, il saggio rappresenta la sintesi perfetta del carattere interdisciplinare del Collegio Superiore dell'Università di Bologna, riuscendo ad integrare conoscenze dell'ambito economico-sociale con quelle ingegneristico-tecniche per analizzare l'evoluzione di un settore come quello aerospaziale i cui sviluppi sono al centro del dibattito attuale.

Elena Argentesi
Professoressa associata di Economia Applicata
Dipartimento di Scienze Economiche
Università di Bologna

L'ECONOMIA DELLO SPAZIO

Secondo la definizione fornita dalla rivista Business Insider l'economia dello spazio è «quell'insieme di attività economiche e di sfruttamento di risorse strettamente collegate all'esplorazione, la ricerca, la gestione e l'utilizzo dello spazio cosmico e coinvolge attori sia pubblici che privati». Con il termine economia dello spazio si intendono quindi una serie di attività che prevedono da un lato ricerca e sviluppo, per la progettazione e realizzazione di infrastrutture spaziali, come stazioni spaziali, stazioni di terra, satelliti, navette, lanciatori e rover, e dall'altro lo sfruttamento in particolare dei dati satellitari per applicazioni di difesa e commerciali come la geolocalizzazione tramite GNSS – global navigation satellite system –, per servizi metereologici, o ancora per la telefonia satellitare. Ad oggi sono stati inviati in orbita più di 9.000 satelliti, per una massa stimata attorno alle 10-20 mila tonnellate, dei quali la maggior parte è ad ora inattiva e costituisce, insieme a moltissimi piccoli detriti, la cosiddetta "spazzatura spaziale".

L'esplorazione spaziale è un'attività piuttosto recente: l'orbita terrestre è stata raggiunta da artefatti umani solamente nel 1957, e pertanto l'industria spaziale ha appena sessant'anni di vita, e segue ancora oggi metodi e approcci non ottimizzati e sicuramente migliorabili, sia nella progettazione che nella gestione dei budget. L'ambiente spaziale è inoltre stato a lungo di competenza esclusiva degli enti governativi delle principali potenze mondiali (USA e URSS). Tuttavia, negli ultimi decenni e in particolare a partire dagli anni 2000, il mercato spaziale ha subito un cambiamento radicale, che si è manifestato primariamente con l'ingresso nel mercato di alcune compagnie private.

Il passaggio conseguente da un regime di monopolio, nel quale era del tutto assente una corrente imprenditoriale interessata a ottenere un profitto dalle attività spaziali, ad un regime di oligopolio, ha creato i presupposti per l'esplosione dell'industria spaziale. Infatti, le nuove compagnie private, introducendo all'interno del mercato spaziale per la prima volta la necessità di rendersi appetibili al mercato privato e di produrre dei guadagni, hanno portato ad una maggiore spinta verso la riduzione dei costi, che si è tradotta nella possibilità di accesso allo spazio per molte industrie impegnate anche in altri settori industriali e dei servizi. È grazie a quest'amplificazione del bacino di aziende possibilitate ad accedere allo spazio che il mercato spaziale ha avuto una crescita esponenziale in questi ultimi anni, la quale tuttora non accenna a rallentare, sembrando anzi destinata a crescere ancora più velocemente nei prossimi anni secondo alcuni studi (vedi bibliografia).

I cambiamenti nel mercato hanno riguardato sia i lanciatori (ovvero i razzi, come lo Space Shuttle o l'europeo Ariane V), che operano come 'navette' per portare in orbita i satelliti, sia quest'ultimi, che costituiscono il vero oggetto delle missioni spaziali e che hanno subito una rivoluzione nei costi e nelle dimensioni anche maggiore rispetto ai lanciatori. Tuttavia, questa rivoluzione risulta meno trasversale, avendo coinvolto solo i satelliti commerciali piuttosto che i grandi satelliti per la ricerca scientifica sviluppati dalle principali agenzie spaziali mondiali.

UNA QUESTIONE DI BUDGET

La riduzione dei costi dei lanciatori ha portato ad un risparmio anche per gli enti governativi, che in realtà hanno cercato tale risparmio (in particolare la NASA) subappaltando sempre di più la realizzazione dei sistemi a contraenti esterni. La NASA infatti a partire dagli anni '90 ha visto sempre di più ridursi il proprio budget, proprio a causa dell'aumento delle spese per la ricerca spaziale e della riduzione delle motivazioni politiche volte alla supremazia tecnologica, che avevano costituito il motore primario della corsa allo spazio in particolare tra USA e URSS a partire dalla fine degli anni '50 fino ai primi anni '90 del Novecento, ovvero durante il periodo della Guerra Fredda.

In realtà la NASA ha potuto godere a lungo di un budget molto più elevato rispetto a quello di Roscosmos (l'agenzia spaziale russa), che già ai tempi della corsa alla Luna – e in maniera ancora maggiore una volta che questa fu raggiunta dagli Stati Uniti – ha sofferto di una cronica mancanza di budget, che ne ha limitato le prerogative politiche ed economiche.

Sotto la presidenza Kennedy, gli sforzi economici degli Stati Uniti aumentarono, portando il budget della NASA a circa 5 miliardi di dollari all'anno, 35 miliardi di dollari al cambio attuale, ovvero dieci volte il budget russo. Pur assorbendo solo una parte del budget della NASA, che doveva finanziare anche i programmi aeronautici, per il programma Apollo la spesa lievitò dal preventivo iniziale di 7 miliardi di dollari a 25,4 miliardi considerando tutte le spese sostenute, una cifra che destò scalpore all'epoca e che corrisponde a più di 100 miliardi di dollari al cambio attuale, impiegati nel solo arco temporale tra il 1967 e il 1972, ovvero più di 20 miliardi di dollari all'anno, una spesa oggi ingiustificabile anche per le maggiori economie mondiali, e che tuttavia risulta molto inferiore a quella del programma Space Shuttle, che ha visto investiti circa 199 miliardi di dollari distribuiti però tra il 1981 e il 2011, quindi 'solo' 5 miliardi di dollari all'anno.

Questo aumento sostanziale del budget iniziale, che non si ha avuto solo per Apollo ma anche per lo stesso Shuttle e in realtà per la maggior parte dei programmi più importanti della NASA, come quelli riguardanti la ISS o le missioni verso Marte, è ancor più sorprendente se si tiene conto che tali budget comprendevano già la possibilità di fallire durante la fase di sviluppo. Una possibile causa di questi errori di valutazione è probabilmente da associare alla volontà da parte dei dirigenti NASA di far approvare i programmi alle amministrazioni federali, che avrebbero poi dovuti rispondere dei costi ai contribuenti.

Va poi considerato che all'epoca della Guerra Fredda, grazie soprattutto all'egemonia geopolitica delle due nazioni, USA e URSS godevano di risorse naturali praticamente illimitate, senza peraltro avere alcuna coscienza ecologica, e pertanto l'approvvigionamento dei materiali necessari, anche i più esotici, e l'allestimento di siti di prova avevano costi irrisori, e gli enormi budget erano spesi in gran parte in capitale umano e manodopera per la realizzazione dei componenti.

Oggi questa egemonia non è più vera in particolare per gli Stati Uniti, e insieme ad un cambio di atteggiamento da parte dell'opinione pubblica ha portato ad un forte ridimensionamento del budget disponibile. Un ulteriore sviluppo dell'industria spaziale non può quindi avvenire se non grazie ad una riduzione dei costi di sviluppo e di lancio. Fortunatamente, diversi fattori hanno portato e stanno ancora portando ad una riduzione degli stessi, risulta pertanto interessante confrontare l'epoca delle grandi missioni governative con l'attuale "privatizzazione" dello spazio, con riferimento in particolare alla riduzione dei prezzi di lancio operata prevalentemente grazie all'azienda americana Space X.

AGLI ALBORI DELL'ERA SPAZIALE: LA POLITICA SOPRA I VANTAGGI ECONOMICI

Il lancio dello Sputnik nel 1957 da parte dell'allora URSS, rappresentò l'inizio dell'era spaziale per l'uomo. Dodici anni dopo, il 20 luglio 1969, lo sbarco dell'uomo sulla Luna rappresentò il culmine della prima fase dell'esplorazione spaziale. Questi due eventi ben rappresentano le motivazioni che hanno spinto lo sviluppo dell'esplorazione spaziale ai suoi albori, nonché le modalità con cui questo processo si è sviluppato. Se infatti l'esplorazione spaziale è stata sempre considerata dall'uomo come un'attività estremamente affascinante, indirizzata al progresso, alla ricerca dell'ignoto, alla conoscenza, il fondamentale impulso che ha consentito all'uomo di raggiungere lo spazio è venuto dalla politica. I primi programmi spaziali ebbero infatti successo – se si può parlare di successo, dati gli enormi fallimenti che si sono avuti lungo il percorso

di crescita – grazie a budget che oggi sarebbero improponibili per qualsiasi economia mondiale, considerando che le uniche due che potrebbero affrontare tale spesa, Cina e India, sono molto attente al budget per l'esplorazione spaziale.

Sebbene già dagli anni '20 del Novecento fossero stati numerosi gli esperimenti per cercare di mandare dei razzi in orbita, la prima era spaziale – gergalmente nota come "corsa allo spazio" – ebbe inizio dopo la Seconda Guerra mondiale ed in particolare con il crescere delle tensioni tra Stati Uniti e URSS, a partire dalla fine degli anni '50, anche se ufficialmente si parla di era spaziale a partire dal lancio dello Sputnik nel 1957. Queste tensioni portarono le due nazioni antagoniste a identificare la conquista dello spazio come la principale manifestazione di una supremazia economica e tecnologica che entrambi i Paesi volevano imporre sull'altro. Pertanto, il successo e la supremazia in tale ambito non avevano come obiettivo una diretta ricaduta economica, seppure questa poteva avvenire indirettamente, come ricaduta tecnologica nella vita di tutti i giorni – peraltro fino a pochi anni fa tale motivazione è stata la più utilizzata negli USA e in molti altri paesi per giustificare le enormi spese legate all'esplorazione spaziale, rappresentando inoltre, unitamente all'identificazione del progresso tecnologico come metro dell'evoluzione di una civiltà, la motivazione per cui Elon Musk ha investito molto in Space X – ma semmai potevano essere ritenuti ragione di un vantaggio politico-economico a livello globale, in quanto tasselli fondamentali per una supremazia e quindi influenza anche politica dei due paesi sull'economia e la politica dell'intero pianeta (Crawford, 1995).

Dopo la prima spinta all'esplorazione spaziale, culminata con lo sbarco dell'uomo sulla Luna, e quindi con una vittoria statunitense, i programmi spaziali russi sono stati ridimensionati, anche per colpa di una situazione economica difficoltosa all'interno dell'URSS. Sono quindi stati gli USA ad affermarsi non solo nel settore dell'esplorazione spaziale, ma in tutta l'industria aerospaziale, con la crescita esponenziale di aziende come Boeing, Lockheed Martin e McDonnell Douglas.

Seppure pochi anni dopo lo sbarco sulla Luna, con il termine del programma Apollo nel 1975, la prima era spaziale poteva dirsi conclusa, fino agli inizi del ventunesimo secolo è rimasta la NASA l'unica in grado di compiere significative missioni spaziali, per via di un budget consistente (nell'ordine dei miliardi di dollari ogni anno), seppur non più illimitato, e di un quasi monopolio sulle tecnologie a livello mondiale. Sebbene alcune delle principali aziende per la produzione delle infrastrutture spaziali fossero private, esse hanno lavorato per molto tempo utilizzando quasi unicamente i fondi pubblici forniti dal governo americano, avendo come unico committente la NASA. L'industria spaziale era quindi totalmente legata, sia a livello nazionale che a livello globale, al monopolio della NASA.

LE CAUSE DEGLI ELEVATI COSTI DI LANCIO

Durante l'epoca di questo "monopolio" da parte della NASA, i costi dei lanci sono stati elevatissimi, e del tutto insostenibili per qualsiasi azienda privata. Il motivo principale per cui non si ha avuto all'inizio una riduzione degli stessi è che la NASA, lavorando in un sostanziale regime di monopolio, non si è mai dovuta confrontare con una concorrenza interessata a guadagnarsi una percentuale del mercato, e quindi non ha mai dovuto preoccuparsi di essere economicamente competitiva nella progettazione, realizzazione e gestione dei lanciatori essendo essa stessa il cliente finale, oltre che il costruttore. A questo bisogna unire il budget elevato riservato ai programmi spaziali: da un lato, come già affermato, per l'obiettivo di supremazia politica e dall'altro perché alcuni programmi, in particolare quelli legati alla spedizione di satelliti in orbita, sono risultati di grande utilità al settore della difesa americana, e quindi sono stati finanziati anche con il budget riservato a tale utilizzo (si ricorda che gli Stati Uniti sono tuttora il paese a livello mondiale che spende di più per la difesa, più dei successivi dieci paesi per budget messi assieme, ovvero 732 miliardi di usd – dollari americani correnti – nel 2019).

A queste motivazioni, bisogna aggiungerne altre di natura tecnica, che hanno reso il costo dei lanci elevato per tutti i veicoli spaziali. Si tratta nello specifico dei requisiti tecnici da soddisfare durante la progettazione, il difficoltoso adempimento dei quali va tutt'oggi di pari passo con gli elevati costi di progettazione e sviluppo tipici dell'intero settore aerospaziale.

Un primo obiettivo è quello di massimizzare la quantità di carico utile (ovvero il carico non legato a necessità costruttive, ma quello che è utile al conseguimento degli obiettivi della missione), ovvero di strumenti che compongono la missione, che dipende dalla potenza del razzo; allo stesso tempo per avere la massima performance occorre ridurre al minimo il peso da trasportare, per esempio miniaturizzando gli strumenti o eliminando qualsiasi ingombro inutile. Questi obiettivi valgono sia per gli spacecraft che utilizzano un lanciatore per trasportare il payload (carico utile) in orbita o al di fuori di essa, come satelliti e sonde interplanetarie, sia per quelli, come lo Space Shuttle, che costituiscono contemporaneamente il payload e il lanciatore. Dati i costi molto elevati per la progettazione e programmazione di ogni singola missione, la ricerca per l'aumento della potenza dei razzi e per la riduzione del peso comporta degli sforzi ingegneristici enormi, che si traducono a loro volta in spese ancora più ingenti. Inoltre, dato che spesso nella storia del volo spaziale il carico utile massimo dei razzi – ovvero dell'infrastruttura comune tra le differenti missioni – è stato un fattore limitante, sul quale non poteva essere operata alcuna

significativa miglioria, le necessità delle missioni hanno portato ad intensificare gli sforzi per la riduzione dei pesi del payload, portando i progettisti a concentrarsi sull'ottimizzazione di ogni singola missione piuttosto che sugli aspetti comuni a tutte, con un conseguente aumento dei costi a dismisura.

Rispetto a queste problematiche si aggiunge quindi anche il maggiore costo delle tecnologie a singolo uso rispetto a quelle riutilizzabili. Nella storia dell'esplorazione spaziale, molto raramente si è potuto parlare dello sviluppo di tecnologie direttamente riutilizzabili in altre missioni. Questo concetto di tecnologie a singolo uso è stata un'importante causa del costo elevato delle missioni spaziali, e ha avuto origine, come nell'intero settore aerospaziale, dalla continua e rapida evoluzione tecnologica, che ha reso poco appetibile lo sviluppo di tecnologie riutilizzabili, che al momento del lancio della missione (spesso anche 10 anni dopo l'inizio della stessa), risultavano già obsolete.

Un altro aspetto anche tragico delle missioni con equipaggio è l'elevato dispendio di capitale umano. Nonostante la storia del volo spaziale sia piuttosto recente, e il numero degli astronauti che hanno raggiunto lo spazio sia tuttora molto limitato (552 persone risultano aver superato i cento chilometri di quota al 2016), non sono mancati gli incidenti, purtroppo anche mortali, soprattutto nelle prime fasi dell'era spaziale. Quasi tutte le vittime sono statunitensi o cittadini dell'ex Unione Sovietica, 14 decedute durante il collaudo dei sistemi spaziali, mentre 18 (tra questi presente anche un israeliano), sono decedute durante missioni ufficiali verso lo spazio. Occorre considerare che molti dei decessi durante i collaudi sono avvenuti durante le prime fasi dell'esplorazione spaziale, nel periodo tra il 1961 e il 1968, e comprendono tra i vari Jurij Gagarin, primo uomo nello spazio nel 1961, deceduto tragicamente nel 1968 cercando di evitare un pallone atmosferico durante un'esercitazione, nonché molti astronauti dei programmi Gemini e Apollo, che sono tuttavia spesso ricordati come un successo nonostante le ingenti perdite umane. Per quanto riguarda invece i decessi durante le missioni, non possono non pesare nel bilancio negativo i due incidenti rispettivamente dello Space Shuttle Challenger (28 gennaio 1968) e Columbia (1° febbraio 2003), che hanno provocato la morte di ben quattordici astronauti.

In seguito ai prematuri incidenti avvenuti durante le prime missioni *manned* (con equipaggio a bordo), sono state spese molte risorse per assicurare la piena sicurezza dell'equipaggio a bordo. Peraltro, per evidenti ragioni ambientali, non è accettabile un qualsiasi guasto all'interno del veicolo, tale da mettere in pericolo la vita dell'equipaggio. Per tale motivo occorre eliminare i rischi con un'accurata progettazione e una lunga fase di prove prima del lancio, nonché creando strutture ridondanti che comportano pesi e volumi maggiori, inutili allo svolgimento della missione in caso di funzionamento nominale di tutti gli apparati.

Quando si parla di volo umano, le problematiche non riguardano solo la sicurezza degli astronauti, ma anche il lunghissimo tempo necessario all'addestramento degli stessi (si parla in generale di una decina di anni almeno per ottenere la qualifica, senza avere la certezza di essere impiegati in una missione in orbita), sia in termini di personale che di strutture per l'addestramento, con ricadute significative non solo sui budget generali delle agenzie spaziali, ma anche su quelli della singola missione, dal momento che ogni nuova attività prevede un lungo e specifico addestramento per ridurre al minimo la possibilità di errori umani. Gli astronauti inoltre rendono intrinsecamente più costose le missioni anche a livello del vettore, necessitando di un volume ampio per il comfort e di molti chilogrammi di massa ausiliarie per le riserve di ossigeno e di cibo. Inoltre, i vettori adibiti al volo umano presentano prestazioni inferiori, potendo attuare accelerazioni e sollecitazioni ridotte sugli astronauti, rispetto ad un carico inanimato.

La ridondanza necessaria a garantire la sicurezza nelle missioni con equipaggio è una caratteristica comune anche alle missioni *unmanned* (senza equipaggio). È infatti mandatorio avere una bassissima probabilità di rotture nel manufatto spaziale, con conseguenti sforzi estremi e costosi nel design, nonché nei successivi test per verificare il corretto funzionamento dello spacecraft. Questo ha spesso comportato nella storia dell'esplorazione spaziale la realizzazione di due sistemi identici, uno finalizzato al volo e uno unicamente ai test. Questa procedura implica ovviamente un notevole aumento dei costi, ma risulta fondamentale per poter conoscere e verificare il funzionamento di ogni dettaglio dello spacecraft e la sua completa corrispondenza alle specifiche di progetto.

Un ultimo elemento, infine, che porta ad un maggiore impegno economico, è rappresentato dalla complessità dell'ambiente spaziale, in particolare per l'assenza di aria e degli effetti gravitazionali, e per la presenza di elevate radiazioni cosmiche, nonché gli sforzi che la struttura degli spacecraft deve sopportare durante il lancio nello spazio. Tutti questi elementi insieme comportano un approfondito studio di tecnologie innovative e molto specifiche per resistere a condizioni totalmente diverse da quelle atmosferiche.

TRA I COSTI PASSATI E I PREZZI MODERNI: LA RIVOLUZIONE DEL MERCATO DEI LANCIATORI

Sebbene le missioni spaziali siano risultato negli anni molto costose, ultimamente le spese hanno subito un fortissimo calo per molti fattori. In questo paragrafo si analizzerà l'entità della riduzione di tali costi, per quanto riguarda in particolare i

costi per il lancio degli spacecraft, per poi analizzare nei paragrafi successivi le motivazioni, già accennate, che hanno portato a tale riduzione.

Dato che la massa che i lanciatori riescono a spedire in orbita dipende dall'orbita di riferimento (ovvero a quale distanza dalla terra si vuole immettere in orbita il satellite o la navetta spaziale), l'analisi dei costi viene generalmente effettuata rispetto ad un'orbita di riferimento. In particolare, si usano le orbite LEO (*low earth orbit*) come riferimento, ovvero quelle più vicine alla Terra. I costi vengono calcolati come spesa (in usd) per mandare in orbita un chilogrammo di carico utile. Per la stazione spaziale internazionale (ISS), essi risultano più elevati, in quanto quest'orbita risulta molto inclinata rispetto al piano dell'equatore, per facilitare i lanci dagli spazi porti russi, che si trovano a latitudini elevate. Normalmente i siti di lancio risultano vicini all'equatore per sfruttare la velocità di rotazione terrestre, e dover cambiare il piano di rotazione del satellite implica usare molto combustibile per dare al satellite una componente di velocità in un'altra direzione. Pertanto, dovendo usare maggiore combustibile per dare al satellite la stessa velocità finale, ed essendo limitata la quantità massima di combustibile a disposizione del lanciatore, risulta ridotta la quantità di carico utile che può essere trasportata. Infatti, nei lanci spaziali risulta fondamentale la velocità finale del satellite, che serve a quest'ultimo per restare nella sua orbita e dipende unicamente dalle caratteristiche di tale orbita. Poiché il calcolo dei costi è dato dalle spese per il lancio, che restano le stesse, divise per il carico utile, i lanci verso la ISS risultano più costosi. Inoltre, per tutti i lanci, carichi più piccoli del massimo carico utile trasportabile, sistemi di fissaggio e alloggio del carico e un volume limitato, che non permette di raggiungere il peso massimo consentito, hanno spesso aumentato il costo di lancio per chilogrammo.

Il costo dei lanci spaziali ha avuto un iniziale calo nella prima fase dell'esplorazione spaziale, dopo essere stato naturalmente elevatissimo per quanto riguarda i primissimi tentativi, ma poi è rimasto alto per decenni ed è stato particolarmente elevato per quanto riguarda le missioni dello Space Shuttle. Nell'ultimo decennio, lo sviluppo dei razzi commerciali ha ridotto il costo tipico dei lanci nello spazio fino ad un fattore venti, mentre quello dei lanci della NASA verso la ISS (la stazione spaziale internazionale) è diminuito di un fattore quattro.

I primi tre sistemi sviluppati avevano costi di lancio per LEO superiori a 100.000 usd/kg, anche avvicinandosi a 1 milione di usd/kg nel caso di Vanguard (per il quale tuttavia sono considerate anche tutte le spese di sviluppo), che nel 1957 è stato il primo sistema di lancio mai utilizzato dalla NASA, nonché il più costoso. I costi sono poi scesi rapidamente con il Saturn V utilizzato per Apollo – circa 170 volte più economico del Vanguard – che, pur essendo stato lanciato la prima volta nel 1967, mantiene ancora oggi uno dei costi per kg più bassi mai ottenuti, superato

solamente da tre sistemi sovietici e dai due principali lanciatori di SpaceX. Occorre però sottolineare come i costi di lancio del Saturn V non comprendono quelli della navicella Apollo e del modulo di comando, che una volta sganciati dal vettore principale dovevano raggiungere la Luna e tornare indietro, avendo propri costi di lancio molto più elevati di quelli del Saturn.

Dal 1970 ai primi anni 2000 si ha quindi avuto un assestamento dei costi di lancio, soprattutto perché molti sistemi che hanno iniziato a volare molto prima del 2000 continuano ad essere utilizzati ancora oggi. In questo periodo il costo medio per lancio è stato di 18.500 usd/kg, con una gamma tipica tra i 10.000 e i 32.000 usd/kg. Dei 22 sistemi inizialmente lanciati dal 1970 al 2000, solo 7 hanno avuto costi inferiori a 10.000 usd/kg, mentre solo due superiori a 32.000, lo Shuttle con 61.700 usd/kg e il piccolo e costoso Pegasus, volato per la prima volta nel 1990, con circa 67.700 usd/kg. Il vantaggio principale del Pegasus resta la piccola massa del vettore e del payload. Infatti, anche se il costo per chilogrammo è elevato, il costo totale di lancio resta basso, mentre utilizzando lanciatori più grandi occorre trasportare il massimo carico consentito per ottenere le cifre riportate.

Dopo questa fase di stallo, i costi sono tornati a calare a partire dal 2010 con il Falcon 9 (2.700 usd/kg, anche se alcune stime in realtà considerano costi di lancio almeno doppi). Il Falcon Heavy porta ad un ulteriore guadagno attestandosi intorno a 1.400 usd/kg. Per lo Shuttle si è speso in media 20 volte più del Falcon 9 e circa 40 volte più del Falcon Heavy per ogni singolo lancio, mentre il prezzo di lancio (essendo il Falcon un vettore commerciale si può parlare a tutti gli effetti di prezzo per il cliente finale, che sia esso privato oppure la NASA) del Falcon 9 è un settimo del costo medio di lancio nel periodo 1970-2000.

TABELLA 1 Costi di lancio per sistema verso orbita LEO

Sistema	Costo di lancio (usd/kg) in LEO
Vanguard (1957-1959)	~ 1000000
Saturn V (1967-1973)	~ 6000
Atlas V (2002-presente)	~ 10000
Space Shuttle (1981-2011)	~ 61700
Pegasus (1990-presente)	~ 68000
Delta IV Heavy (2004-presente)	~ 15000
Falcon 9 Full Thrust (2013-presente)	~ 2700
Falcon Heavy (2018-presente)	~ 1400
Ariane V (1996-presente)	~ 11000
Soyuz-2 (2006-presente)	~ 5900

Per quanto riguarda i costi di lancio per la Stazione Spaziale Internazionale (ISS), la tabella 2 mette a confronto lo Space Shuttle e il Falcon 9 più capsula Dragon, evidenziando come il Falcon 9 riduca le spese di circa un fattore 4, con un impatto meno estremo rispetto a quello per l'orbita LEO, ma comunque sorprendente.

TABELLA 2 Costi totali di lancio dello shuttle e del Falcon 9 verso la ISS

Sistema	Shuttle	Falcon 9 + Dragon
Costo per lancio, M$ (2018)	1697	150
Kg trasportati sulla ISS	16050	6000
Costo specifico (migliaia usd/kg)	105,8	25

PRINCIPALI SISTEMI DI LANCI AMERICANI, RUSSI ED EUROPEI

DA APPROCCIO DIFFERENTE UNA RIDUZIONE DELLE SPESE DI SVILUPPO

Secondo alcuni studiosi, gli elevati costi di lancio sono stati «il più grande fattore limitante per espandere lo sfruttamento dello spazio e l'esplorazione» (Wertz e Larson, 1996, pp. 115-7). Se il settore pubblico non è mai stato in grado – o non ha mai avuto interesse in tal senso – di rivoluzionare il proprio approccio, sono stati allora le compagnie a portare i criteri della sostenibilità economica all'interno del mercato spaziale.

Le compagnie private hanno tentato di risparmiare sui costi iniziali di sviluppo dei vettori partendo da progetti sviluppati per missili balistici, che però sono progettati per prestazioni elevate, non per costi minimi, seguendo un'idea tra l'altro già utilizzata dalle agenzie spaziali nei primi decenni del Novecento, quando per raggiungere lo spazio si cercava sostanzialmente di costruire dei "proiettili" molto più grandi.

Confrontando i razzi con gli aerei, sembra invece la riusabilità la strada ovvia per risparmiare, ma l'esempio del programma Space Shuttle, che ha completato 135 missioni utilizzando 6 modelli in tutto – pertanto ogni modello è stato a lungo utilizzato, senza ottenere tuttavia i risparmi inizialmente sperati – smentisce tale ipotesi. I lanciatori riutilizzabili necessitano di investimenti più cospicui per lo sviluppo, e consentono un carico utile minore a causa della massa di carburante necessaria all'atterraggio. Il Falcon 9 è riutilizzabile ed è stato riutilizzato, ma per ora i risparmi sui costi di progettazione e costruzione sono considerati solo in previsione futura. Rispetto a quanto affermato prece-

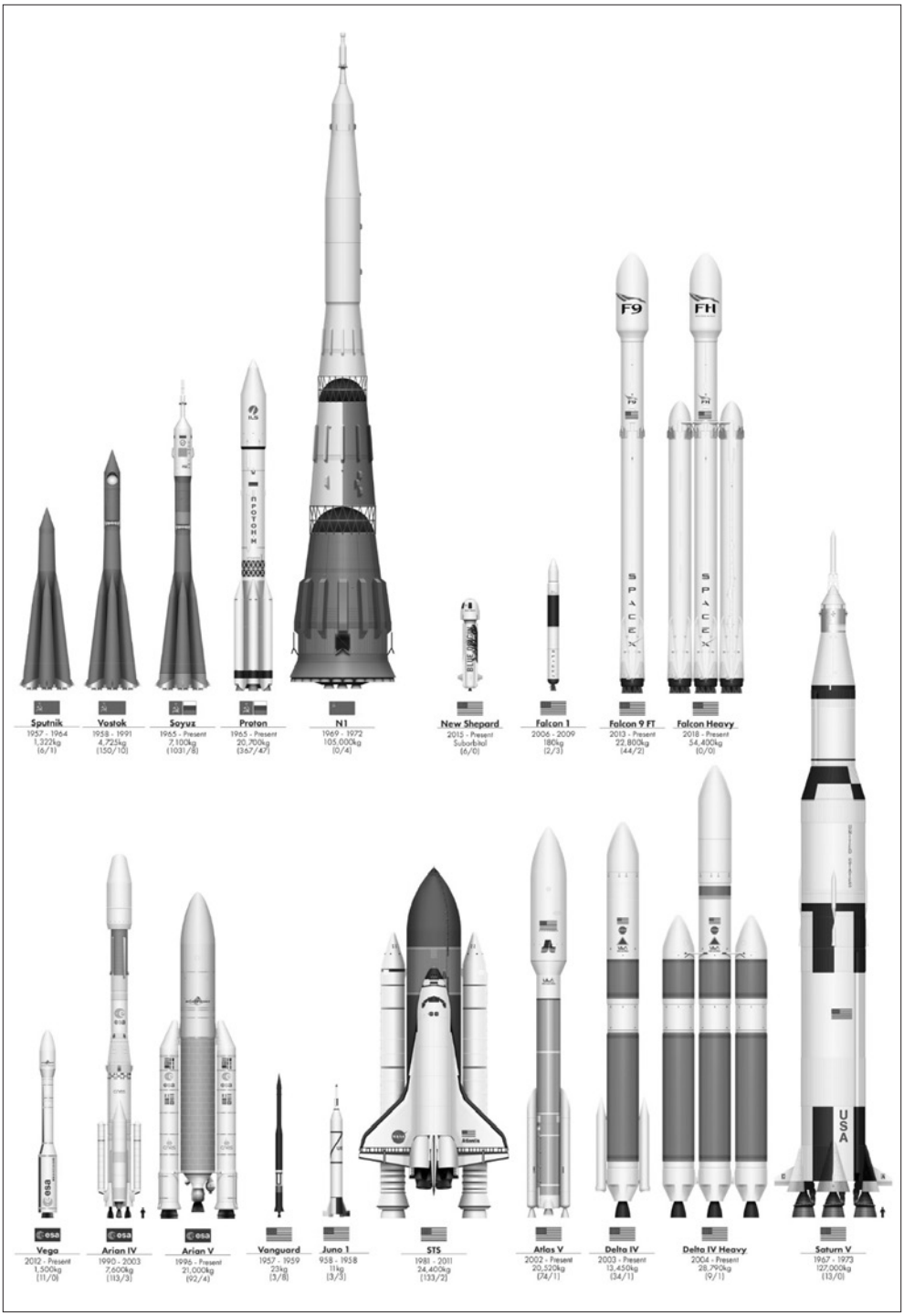

Sputnik
1957 - 1964
1,322kg
(6/1)
Vostok
1958 - 1991
4,725kg
(150/10)
Soyuz
1965 - Present
7,100kg
(1031/8)
Proton
1965 - Present
20,700kg
(367/47)
N1
1969 - 1972
105,000kg
(0/4)
New Shepard
2015 - Present
Suborbital
(6/0)
Falcon 1
2006 - 2009
180kg
(2/3)
Falcon 9 FT
2013 - Present
22,800kg
(44/2)
Falcon Heavy
2018 - Present
54,400kg
(0/0)
Vega
2012 - Present
1,500kg
(11/0)
Arian IV
1990 - 2003
7,600kg
(113/3)
Arian V
1996 - Present
21,000kg
(92/4)
Vanguard
1957 - 1959
23kg
(3/8)
Juno 1
1958 - 1958
11kg
(3/3)
STS
1981 - 2011
24,400kg
(133/2)
Atlas V
2002 - Present
20,520kg
(74/1)
Delta IV
2003 - Present
13,450kg
(34/1)
Delta IV Heavy
2004 - Present
28,790kg
(9/1)
Saturn V
1967 - 1973
127,000kg
(13/0)

FIG. 1

dentemente, occorre infatti precisare che i Falcon 9 e Heavy sono effettivamente molto più economici di altri lanciatori, ma il prezzo offerto ai clienti tiene già conto della riutilizzabilità dei vettori, il che non vuol dire che SpaceX sia già rientrato dei costi di sviluppo e costruzione dei propri vettori. È altresì vero che anche per quanto riguarda lo Space Shuttle le cifre indicano solo le spese per la preparazione del singolo lancio e non i costi di sviluppo, e queste sono molto maggiori rispetto a quelle sostenute da SpaceX.

Se anche la riutilizzabilità non sembra essere la chiave per l'economicizzazione del mercato dei lanciatori, vi sono altre caratteristiche che sono state considerate in letteratura come promotrici di tale trasformazione. Un ruolo importante è assegnato alla semplificazione della configurazione del velivolo, partendo da progetti preliminari meno avveniristici e che sfruttino in maniera intelligente il know-how già presente all'interno dell'azienda. Altrettanto importante è sfruttare (copiare) le soluzioni trovate da altri partner o dalla concorrenza, laddove vi sia una mancanza di know-how all'interno dell'azienda riguardo un certo elemento del progetto, prima di spendere tempo e denaro per ottenere soluzioni già sperimentate da altri. Questa semplificazione è accompagnata da un aumento della frequenza di utilizzo di ogni singola tecnologia, e quindi del numero di esemplari e di lanci per ogni sistema progettato: un risultato ottenibile grazie ad una maggiore duttilità del progetto, ricercata durante il design preliminare. Risulta pertanto fondamentale prevedere differenti tipologie di missioni affrontabili con un singolo sistema, criterio già utilizzato dalla stessa NASA nel settore aeronautico con lo sviluppo dei moderni caccia multiruolo, come il moderno F-35, in grado di agire da bombardieri, da velivoli di ricognizione e allo stesso tempo di affrontare combattimenti aerei.

Un altro aiuto alla riduzione delle spese viene dall'utilizzo di metodi di progettazione e produzione industriali, diversamente da quanto fatto all'interno della NASA per quanto riguarda la ricerca spaziale. In particolare, è necessario cambiare le priorità negli obiettivi da conseguire durante la fase di sviluppo. Per i vettori privati viene data la priorità all'ottimizzazione dei costi già durante la fase di design concettuale, cercando soluzioni già esistenti e preoccupandosi dei futuri costi di gestione di una certa tecnologia. Questa ottimizzazione procede anche nelle fasi successive, con la scelta dei singoli componenti – ad esempio l'hardware elettronico o le tubazioni idrauliche – adattando le specifiche del progetto alle componenti disponibili piuttosto che il contrario, tenendo in conto del costo, dell'adattabilità al progetto (è comunque meglio quando possibile avere un componente che non obblighi a modificare le specifiche del lanciatore), della reperibilità sul mercato e

delle eventuali spese e tempistiche di manutenzione delle componenti. Questa metodologia va ad assecondare la volontà di semplificazione del progetto e va unita ad un aumento dei margini di progettazione, ovvero dell'ampiezza dei valori entro i quali le prestazioni del lanciatore sono ritenute soddisfacenti, rendendo più facile il raggiungimento delle stesse.

Se la NASA ha sempre rifiutato di prendere in considerazione queste soluzioni, bisogna tuttavia sottolineare come, almeno agli inizi e fino agli anni '80, il forte investimento nella ricerca di nuove tecnologie e materiali ha avuto forti ricadute sull'industria statunitense e anche russa. Queste considerazioni restano valide ancora oggi, ma a partire dagli anni '90 la capacità di innovazione della NASA si è ridotta a causa di un parziale esaurimento delle più branche di ricerca più semplici, mentre al contempo con l'elevata crescita della capacità di innovazione delle industrie private, le agenzie pubbliche si sono trovate a operare con spese irragionevoli per il mercato ottenendo gli stessi risultati, se non peggiori. La forte interconnessione con il settore militare si è poi spesso manifestata nella necessità di mantenere segrete le tecnologie, e di svilupparne molte che non fossero disponibili sul mercato, azzerando i vantaggi dell'innovazione tecnologica per la società.

Nonostante ciò, la causa primaria dell'ingenza dei costi sembra comunque essere la mancanza di concorrenza. Il regime di monopolio ha visto a lungo l'ULA (United Launch Alliance) come fornitore e il suo principale cliente è stato il governo degli Stati Uniti (NASA e Dipartimento della Difesa) che ha sempre avuto come priorità l'alta affidabilità e le massime prestazioni, con un basso incentivo alla riduzione delle spese o al rispetto dei budget. L'ULA è una joint venture composta dai principali complessi aerospaziali statunitensi tra cui Boeing e Lockheed, che però hanno sempre avuto un rapporto diretto con la NASA, senza dover passare tramite bandi di assegnazione, e pertanto senza operare in regime di concorrenza. Inoltre, essa ha sempre operato avendo forti incentivi provenienti direttamente dal governo centrale, oltre che le commissioni della NASA. L'ULA ha tuttavia perso negli ultimi anni la maggior parte del mercato commerciale a causa della Russia e di Arianespace (il principale vettore sviluppato dall'Unione Europea), che sono a loro volta fortemente sovvenzionate rispettivamente dal Governo Russo e dai 22 stati membri dell'ESA (Zimmerman, 2012)

Prendendo come esempio lo Space Shuttle (in attività fino al 2011), una fonte predominante di inefficienza è stata legata ai costi operativi di lancio. Il fatto che siano stati necessari 10.000 appaltatori e oltre 1.000 dipendenti pubblici è indicativo della mancanza di semplicità operativa. Questo esercito di persone, unito alle

operazioni di missione e al personale delle operazioni dell'equipaggio ha costituito un terzo dei costi complessivi delle operazioni dello Shuttle. In aggiunta, la bassa frequenza di utilizzo dello Shuttle ha reso l'utilizzo di tutto questo personale estremamente inefficiente (Rutledge, 93-4063).

L'INGRESSO DI SPACEX NEL MERCATO DEI LANCIATORI SPAZIALI

SpaceX (Space Exploration Technologies Corporation), fondata nel 2002 da Elon Musk e con sede ad Hawthorne in California, è stata la prima compagnia privata a sviluppare e mandare in orbita un razzo utilizzando unicamente fondi privati. L'azienda è stata una delle prime insieme a Blue Origin a riuscire a portare ad altezza orbitale dei razzi utilizzando fondi privati, e in entrambi i casi ciò è stato reso possibile grazie al diretto coinvolgimento di due miliardari (Elon Musk per SpaceX e Jeff Bezos per Blue Origin), che hanno messo a disposizione ingenti capitali.

SpaceX in particolare ha avuto molto successo a livello industriale, diventando un partner affidabile per le missioni della NASA, e divenendo inoltre uno dei principali operatori per la messa in orbita dei satelliti di molte compagnie private: il solo Falcon 9, nei vari modelli ha completato più di 90 missioni, con un rateo di successo del 98% (il maggiore tra tutti i vettori spaziali), e ha mandato in orbita più di 1.000 satelliti, di cui 735 per il progetto Starlink. Blue Origin, pur non avendo subito sostanziali fallimenti negli ultimi anni – tralasciando quelli preventivati e quasi fisiologici nello sviluppo di un vettore orbitale, si trova più indietro rispetto a SpaceX, avendo ad ora sviluppato solamente un vettore suborbitale, in grado ovvero di superare i 100 km e quindi di tornare sulla terra con una traiettoria parabolica senza immettersi in un'orbita, tuttora in fase di test, mentre per un vettore orbitale l'azienda si trova ancora in piena fase di progetto. SpaceX al contrario ha effettuato moltissimi lanci orbitali ad oggi, sia per privati sia per la NASA, grazie in particolare al Falcon 9, e collabora attraverso la capsula Dragon al rifornimento della ISS. Inoltre, il 30 maggio 2020, con la missione Demo 2, SpaceX ha mandato sulla ISS per la primissima volta la capsula Crew Dragon con a bordo 3 astronauti. Tale missione ha rappresentato il primo volo con equipaggio di un velivolo statunitense, peraltro partito da suolo americano, dalla dismissione dello Space Shuttle nel 2011, e apre una nuova fase nella storia dell'azienda di Hawthorne.

Il motivo per cui SpaceX ha avuto successo è legato all'alta affidabilità che ha dimostrato, pur mantenendo costi molto ridotti rispetto alla concorrenza, soprattutto grazie alla struttura stessa dell'azienda. Questa è organizzata verti-

calmente, e sviluppa internamente la maggior parte dei componenti dei propri razzi, occupandosi personalmente di tutte le fasi del ciclo di vita del prodotto, tra cui progettazione, produzione, software, integrazione, test, lancio e operazioni di volo e di terra; la maggior parte di queste attività sono svolte in un'unica grande struttura. L'ULA, al contrario, è un'alleanza composta da molte aziende, con centinaia di subappaltatori che hanno decine di strutture sparse in tutto il paese, il che è una necessità politica per un programma pubblico finanziato dal governo.

SpaceX progetta cercando di mantenere una struttura semplice, per esempio il Falcon 9 utilizza 9 motori Merlin identici. Il Falcon Heavy utilizza invece la struttura base di 3 Falcon 9 assemblati assieme, con alcune modifiche per ospitare un maggiore volume di carico utile. Un altro fattore chiave nei bassi costi di SpaceX è la sua forza lavoro giovane e altamente motivata, composta da laureati di alto livello disposti a lavorare per lungo tempo sottopagati rispetto ai tecnici della NASA. Ancora, SpaceX utilizza apparecchiature di produzione automatizzate e all'avanguardia, una tecnologia precedentemente sconosciuta nell'industria spaziale, dove l'assemblaggio manuale dei componenti è ancora la norma (Greg, 2015).

Nel 2010, la NASA ha confrontato i costi di sviluppo del Falcon 9 con i modelli della NASA previsti utilizzando il metodo tradizionale cost-plus-fee. Utilizzando il modello NAFCOM (NASA-AF Cost Model), la NASA ha stimato che lo sviluppo del razzo le sarebbe costato 1.383 milioni di dollari con i metodi tradizionali. Il costo stimato di SpaceX è stato di 443 milioni di dollari, una riduzione del 68% rispetto all'approccio tradizionale.

La stessa SpaceX ha attribuito la propria efficienza nella gestione dei costi ad alcuni fattori chiave, come l'utilizzo di una forza lavoro più piccola, lo sviluppo interno all'azienda dei componenti e dei sottosistemi, un'organizzazione con meno livelli di gestione e meno infrastrutture e una cultura dello sviluppo commerciale. Il costo di Progettazione, Sviluppo, Test e Valutazione (DDT&E) dipende principalmente dalle dimensioni della forza lavoro necessaria. SpaceX stima che il subappalto di un dollaro di lavoro interno costerebbe da tre a cinque dollari a causa dei costi generali e dei profitti dei terzisti, se subappaltato. Questo avviene in quanto in ambito spaziale tutte le tecnologie sono molto specifiche e difficilmente utilizzabili su larga scala, pertanto anche i subappaltatori devono sviluppare tecnologie adattate allo specifico uso, senza pertanto applicare ottimizzazioni dovute ad un'industria di scala, e in più necessitano di un margine di profitto e di garanzie anche economiche dato l'alto rischio di insuccesso nello sviluppo.

Un altro fattore determinante del minor costo di SpaceX è stato la gestione moderna, che ha consentito uno sforzo ingegneristico più efficace. L'approccio di SpaceX alla progettazione dei razzi deriva da un principio fondamentale: la semplicità consente sia affidabilità che basso costo. Tutti i motori del Falcon 9 sono identici, mentre altri razzi utilizzano due o tre differenti propulsori per ottenere le prestazioni necessarie, a un costo più elevato. Lo stile organizzativo di SpaceX ricalca quello altamente efficiente della silicon Valley, piuttosto che quello della NASA, focalizzandosi sulla cultura imprenditoriale, su una produzione snella, una gestione orizzontale e una forte interconnessione tra i reparti in fase di progettazione e poi di produzione. Secondo Chaikin «questa è davvero la più grande innovazione di SpaceX: sta portando le pratiche standard di ogni altra industria nello spazio» (Chaikin, 2012).

GLI EFFETTI DELLA CONCORRENZA SULLA RIDUZIONE DEI COSTI

Oltre a SpaceX, sono molti i vettori che sono subentrati negli ultimi anni nel mercato dei lanciatori, sia compagnie private, sia altre agenzie nazionali e sovranazionali, che sono ora in grado di mandare satelliti e altre infrastrutture nello spazio con estrema regolarità. Tra queste, ad esempio, si possono nominare l'ESA (l'agenzia spaziale europea), la JAXA (Giappone), l'agenzia spaziale cinese e quella indiana, che stanno tutte crescendo moltissimo in questi ultimi anni. Le agenzie spaziali cinese e indiana sono in realtà tuttora molto focalizzate sul mercato interno, con ridotte collaborazioni internazionali, e operano in un sostanziale regime di monopolio interno, pur avendo un approccio molto più orientato allo sviluppo dell'industria privata nazionale, consentendo e incentivando la nascita di vettori privati. Sotto certi aspetti questo rimane valido per tutte le agenzie, ricordando che le autorizzazioni per il lancio anche delle compagnie private vengono comunque rilasciate dalle agenzie nazionali.

L'elevato numero di operatori ha in ogni caso aumentato il livello di opzioni e di concorrenza nel mercato; gli effetti generali di tale concorrenza hanno portato a un'ulteriore riduzione dei costi, con conseguente aumento del numero di voli spaziali, che ha portato e porterà ancora ad una riduzione dei costi a causa della curva di esperienza e alla crescita dell'affidabilità, ripagando più rapidamente gli investimenti per lo sviluppo iniziale e quindi giustificando maggiori investimenti nella progettazione dei lanciatori. In precedenza, il mercato dei lanci apparteneva a un numero limitato di entità sostenute dai governi, più interessate alla capacità

militare, all'affidabilità del lancio, al prestigio nazionale e alla creazione di posti di lavoro e stimoli economici che al risparmio economico o all'applicazione delle nuove tecnologie sviluppate. La nascita dei razzi commerciali ha invece fornito un diverso modello di business industriale che ha notevolmente ridotto le spese.

LE OPPORTUNITÀ DELLO SPAZIO ACCESSIBILE A TUTTI

Oggi, con la grande riduzione dei costi, si ha avuto un'esplosione del mercato dei satelliti: molte compagnie sono entrate nel mercato, potendosi permettere di lanciare i propri satelliti per offrire nuovi servizi (telecomunicazioni, internet, earth imaging, gps, meteorologia).

Fino al 2000, i satelliti di comunicazione erano il principale mercato di lancio commerciale. Il numero totale di lanci commerciali è stato di circa 25-35 all'anno fino a quando il Falcon 9 ha ampliato il mercato divenendo il più grande fornitore e superando l'Ariane 5 (ESA) come economicità di lancio. Poiché i satelliti per le comunicazioni utilizzano un numero limitato di slot spaziali (la posizione in orbita all'interno della quale è consentito posizionare il satellite) assegnati a livello internazionale, spesso il loro lancio è stato consentito grazie a forti pressioni politiche. Pertanto, il mercato commerciale del lancio di satelliti è stato tutt'altro che aperto e competitivo, dal momento che la maggior parte dei fornitori e dei clienti sono stati regolamentati e sovvenzionati dal governo.

Con simili limitazioni di mercato, costi inferiori potrebbero non avere un impatto significativo. Tuttavia, l'abbattimento dell'impatto degli stessi sul budget di missione (il budget complessivo per sviluppo, lancio e gestione del satellite) ha reso meno importante la riduzione dei pesi, consentendo per la prima volta di concentrarsi sulla riduzione dei costi dei satelliti stessi tramite la standardizzazione delle componenti. Inoltre, si è comunque avuto una fortissima spinta alla riduzione dei pesi e delle dimensioni, grazie anche allo sviluppo della microelettronica e di sottosistemi miniaturizzati. Questi due fattori hanno portato ad un enorme riduzione del costo di un satellite, aprendo il mercato anche a piccole compagnie che non avevano i satelliti come asset principale, e portando in questo modo il mercato alla produzione di satelliti piccoli ed economici, da poter mandare in orbita facilmente e in tempi rapidi. Ai vecchi satelliti per telecomunicazioni, in orbita alta, si stanno sostituendo costellazioni numerose di piccoli satelliti economici, spediti in orbita LEO. Essendo ridotte le spese di lancio, si preferisce oggi fare questi satelliti meno affidabili, riducendo al minimo le spese di sviluppo e sostituendoli quando si rom-

pono. Questo aspetto si contrappone alla storica tendenza, motivata appunto dagli elevati costi di lancio, a progettare satelliti grandi e molto affidabili, che potessero sopravvivere per il maggior tempo possibile. Grazie al nuovo modello di mercato, l'industria dei satelliti è esplosa, portando oggi ad una privatizzazione dello spazio, che risulta sempre più occupato da flotte di piccoli satelliti privati utilizzati per scopi commerciali. (da Wikipedia – Space Launch Market Competition).

Anche in ambito militare è molto l'interesse ad avere minori costi di lancio; le capacità militari degli Stati Uniti dipendono notevolmente dalle risorse spaziali, tra cui comunicazioni, posizionamento globale, meteo e satelliti di sorveglianza. I satelliti sono vulnerabili agli attacchi e difficili da nascondere o difendere, quindi la capacità di sostituirli rapidamente ed economicamente è molto attraente.

CONCLUSIONE

Il mercato dei lanci spaziali ha subito grandi cambiamenti, dai primi anni dell'era spaziale, legati a grandi missioni, mosse da motivazioni politiche e quindi mirate a "conquistare" lo spazio, fino ai giorni nostri, nei quali si sta sempre più affermando un mercato privato, orientato ad uno sfruttamento dello spazio per fini commerciali. In seguito a questa privatizzazione il valore economico del mercato spaziale è esploso ed è destinato a crescere ancora a lungo. Le motivazioni che hanno reso possibile questo cambiamento e questa "accessibilità" dello spazio per tutti sono da rinvenire nella sostanziale caduta del regime di monopolio applicato dalla NASA nella seconda metà dello scorso secolo, avvenuta a causa dell'ingresso di alcuni lanciatori privati, primo di tutti SpaceX. Il conseguente calo dei costi per il lancio dei satelliti ha avuto una serie di ricadute a catena, consentendo una rapida espansione del mercato privato.

BIBLIOGRAFIA

Chaikin, A., *Is SpaceX Changing the Rocket Equation?*, in *Air & Space Magazine*, January 2012.

Crawford, I.A., *Space development: social and political implications*, in *Space Policy*, 1995.

Greg, *SpaceX – Low cost access to space*, Harvard Business School, Technology and Operations Management, 2015.

Jones, H.W., *The Recent Large Reduction in Space Launch Cost*, 48th International Conference on Environmental Systems ICES-2018-81, July 2018.

Rutledge, W., *Launch Vehicle Cost Trends and Potential for Cost Containment*, AIAA Space Programs and Technologies Conference and Exhibit, 1993.
Wertz, J.R., Larson, W.J., eds., *Reducing Space Mission Cost*, Space Technology Series, Kluwer, 1996.
Zimmerman, R., *The Actual Cost to Launch*, in *Behind the Black*, April 6, 2012, http://behindtheblack.com/behind-theblack/essays-and-commentaries/the-actual-cost-to-launch/

SITOGRAFIA

https://en.wikipedia.org/wiki/space_launch_market_competition
https://it.businessinsider.com/space-economy-le-ragioni-economiche-dietro-la-nuova-corsa-allo-spazio-che-ci-riguarda-da-vicino/
https://www.industry.gov.au/sites/default/files/2019-03/global_space_industry_dynamics_-_research_paper.pdf

IL PARADOSSO DELL'INFORMAZIONE NEI BUCHI NERI

CHIARA CORTESE

San Francisco, 1983. Un imprenditore americano aveva organizzato una piccola conferenza di Fisica nel suo attico. Gli invitati erano alcuni tra i fisici più in vista a quel tempo ed erano stati chiamati semplicemente per discutere, a beneficio dell'imprenditore, le teorie più recenti e pioneristiche nel panorama della ricerca in Fisica. Di tutti i discorsi tenuti nel corso della conferenza, memorabile fu quello del noto scienziato Stephen Hawking. Nel corso del suo intervento, diede inizio ad uno dei dibattiti più interessanti degli ultimi anni ponendo la seguente domanda: qual è il destino dell'informazione caduta in un buco nero?
Hawking aveva già dimostrato l'evaporazione dei buchi neri e riteneva che l'informazione non si potesse conservare durante il processo. Questa teoria fu fin dall'inizio fortemente contrastata dagli scienziati Leonard Susskind e Gerard 't Hootf. Dal loro punto di vista, la perdita di informazione non era accettabile perché avrebbe minato uno dei pilastri della meccanica quantistica. Il dibattito che ne seguì è noto come "Guerra dei buchi neri".
In questo capitolo verranno presentati i tratti principali del paradosso dell'informazione dei buchi neri, partendo dalle proprietà fisiche dei buchi neri, comprese la temperatura e l'entropia, per arrivare a trattare il paradosso e le soluzioni proposte da Hawking, Susskind e 't Hooft.

San Francisco, 1983. A small physics lecture had been organized by an American entrepreneur in his loft. Some of the leading physicists of that time were there to simply discuss the most recent and avant-garde theories in the panorama of the physics research. Of all the speeches, the most remarkable was the one given by the famous scientist Stephen Hawking. He started one of the most fascinating physics debates in recent years with the following question: what is the fate of information fallen into a black hole?
Hawking had already demonstrated black hole evaporation and he believed that information couldn't be conserved during the process. This theory was from the beginning strongly opposed by the scientists Leonard Susskind and Gerard 't Hooft. From their point of view, information loss was not acceptable, it would undermine the basis of quantum mechanics. The ensuing debate is known as the "Black Hole War".
In this paper the main features of the black hole information paradox will be presented, starting with the physical properties of a black hole, including temperature and entropy, and then moving on to the paradox itself and the solutions proposed by Hawking, Susskind and 't Hooft.

PREFAZIONE

I buchi neri, originariamente congetturati studiando le soluzioni delle equazioni di Einstein della relatività generale, sono oggi divenuti oggetto di precise osservazioni sperimentali.

La rilevazione diretta delle onde gravitazionali, emesse dallo scontro catastrofico di buchi neri, ha permesso di osservarne la loro presenza. Similmente, una rete di radiotelescopi è recentemente riuscita a raccogliere la prima immagine di un buco nero, visto come un anello di radiazione che avvolge l'orizzonte degli eventi, il punto di non ritorno della materia assorbita dal buco nero.

La teoria di Einstein della relatività generale è la teoria che descrive la forza gravitazionale. In tale teoria la gravità è legata alla struttura ed alla geometria dello spaziotempo. In particolare, essa mostra come la forza gravitazionale, attraverso la sua natura attrattiva, permetta l'esistenza e la formazione dei buchi neri, oggetti con un campo gravitazionale così intenso da non lasciare sfuggire nulla di quanto attratto al suo interno: tutta la materia e la luce che attraversa il cosiddetto orizzonte degli eventi è intrappolata dal buco nero.

Questo è vero finché non si tiene conto della meccanica quantistica, la meccanica che governa la materia a scale di lunghezza molto piccole. Come mostrato originariamente dal fisico teorico Steven Hawking, gli effetti quantistici permettono che della radiazione possa uscire dal buco nero. Quindi, a causa della meccanica quantistica, i buchi neri non sono poi così neri come predetto dalla teoria della relatività generale.

Questo fenomeno inaspettato costituisce il punto cruciale su cui testare e capire la struttura quantistica della gravità, la cui formulazione è attualmente dibattuta e studiata intensamente, con incognite e paradossi che sembrano sfuggire ad una trattazione consistente e completa.

Il saggio di Chiara Cortese descrive in modo semplice e chiaro il paradosso dell'informazione, che emerge nella interpretazione della radiazione di Hawking emessa dai buchi neri. Come descritto nel saggio, i buchi neri sono oggetti molto complessi ed interessanti, che continuano ad essere al centro delle ricerche che riguardano la struttura della gravità classica e quantistica.

Fiorenzo Bastianelli
Professore associato di Fisica Teorica, Modelli e Metodi Matematici
Dipartimento di Fisica e Astronomia "Augusto Righi"
Università di Bologna

INTRODUZIONE

Attorno al 1980 l'attenzione del mondo della Fisica era rivolta principalmente all'infinitamente piccolo. Le teorie quantistiche, infatti, avevano aperto la strada alla fisica delle particelle elementari e si stavano facendo scoperte molto importanti in quel campo. La costruzione del Modello Standard aveva infatti consentito di descrivere con un'unica teoria tre delle quattro interazioni fondamentali: l'interazione debole, l'interazione forte e quella elettromagnetica. La quarta interazione fondamentale, quella gravitazionale, rimaneva al di fuori del Modello Standard per una sua peculiarità: non si riusciva a rinormalizzarla come teoria quantistica.

La Relatività generale di Einstein, che descrive l'interazione gravitazionale, ha un'impostazione molto diversa rispetto a quella quantistica: è una teoria "classica", priva del carattere probabilistico della meccanica quantistica. La ricerca in campo gravitazionale si concentrò, quindi, nell'elaborazione di una nuova teoria, più completa, che potesse unire Relatività generale e meccanica quantistica: la "gravità quantistica".

Oggi la gravità quantistica è per i fisici una vera e propria chimera: se venisse formulata e confermata sperimentalmente, sarebbe in grado di descrivere in modo coerente il mondo che ci circonda e di uniformare le leggi fisiche già esistenti. Le candidate al ruolo di gravità quantistica non mancano, prima fra tutte la teoria delle stringhe, che negli ultimi anni è stata al centro di accesissimi dibattiti e campo di ricerca di molti fisici teorici. In ogni caso, non vi sono ad oggi le dovute conferme sperimentali, perciò la gravità quantistica resta, per il momento, un'incognita.

Perché è così complesso lavorare ad una simile teoria? Come già accennato, la Relatività generale e la meccanica quantistica sono profondamente diverse. Le due hanno effetti importanti a scale completamente differenti: gli effetti quantistici intervengono a livello atomico e subatomico mentre la gravità è un'interazione a *lungo range* con effetti consistenti solo se sono coinvolti corpi massivi. Risulta chiaro, dunque, quanto sia difficile trovare un fenomeno che dipenda fortemente sia da effetti quantistici sia dall'interazione gravitazionale. Difficile però non vuol dire impossibile. È dall'astrofisica che arriva un soggetto caratterizzato da fenomeni di questo tipo: il buco nero.

Torniamo ora agli anni ottanta, precisamente al 1983. In un attico di San Francisco si teneva una piccola conferenza organizzata da un imprenditore americano appassionato di fisica. Gli ospiti erano tutti fisici di spicco ed erano stati chiamati semplicemente per discutere le teorie che in quel momento erano l'avanguardia della ricerca. Di tutti gli interventi, il più memorabile fu tenuto dal celebre Stephen

Hawking. Hawking nasce come relativista e pertanto i suoi studi si sono sempre concentrati su quanto lasciato in sospeso dalla Relatività generale, prima fra tutte l'esistenza dei buchi neri. Aveva già dimostrato una caratteristica molto interessante dei buchi neri, ovvero la loro evaporazione, ma su questo si tornerà più avanti. A San Francisco Hawking fece un'affermazione che aprì uno degli scontri di idee più appassionanti, e produttivi, della fisica moderna. È passato alla storia come "La Guerra dei buchi neri".

L'intervento di Hawking si concentrava su una particolare questione: qual è il destino dell'informazione caduta in un buco nero? La sua risposta fu che nell'evaporazione l'informazione va irrimediabilmente perduta. Alla conferenza la frase sconvolse soltanto due tra i fisici presenti: Leonard Susskind e Gerard 't Hooft. Per due esperti di quantistica come loro la perdita di informazione era inaccettabile, perché andava a minare le fondamenta della meccanica quantistica e quindi, dal loro punto di vista, della fisica. Furono loro i principali avversari di Hawking nella Guerra dei buchi neri.

Verranno ora presentate alcune delle proprietà dei buchi neri, per poi passare al paradosso dell'informazione e le soluzioni che vennero proposte da 't Hooft, Susskind e Hawking.

BUCO NERO: FORMAZIONE E PRINCIPALI CARATTERISTICHE

Il buco nero è uno dei soggetti più misteriosi di cui disponiamo a livello astronomico. Per comprenderne la fisica è fondamentale sapere da cosa è generato: dal collasso di una stella molto massiva. Il punto di partenza della trattazione sarà, quindi, l'evoluzione stellare.

Le stelle si formano quando una grande quantità di gas (soprattutto idrogeno) comincia ad essere gravitazionalmente instabile, ovvero quando la pressione interna del gas non è più sufficiente a compensare l'attrazione gravitazionale tra le molecole che lo compongono. La nube di gas entra in una fase di collasso su se stessa durante la quale le molecole del gas, che hanno un moto con direzione casuale, collidono tra loro sempre più frequentemente. Le collisioni ripetute comportano un riscaldamento del gas fino a che gli atomi di idrogeno non si fondono tra loro, formando elio. Lo splendore della stella è dovuto proprio al calore liberato da questa reazione, simile all'esplosione controllata di una bomba a idrogeno. Questo calore aggiuntivo aumenta ancora di più la velocità delle molecole e così le loro collisioni. La pressione interna del gas è un effetto macroscopico delle collisioni molecolari ed è quindi

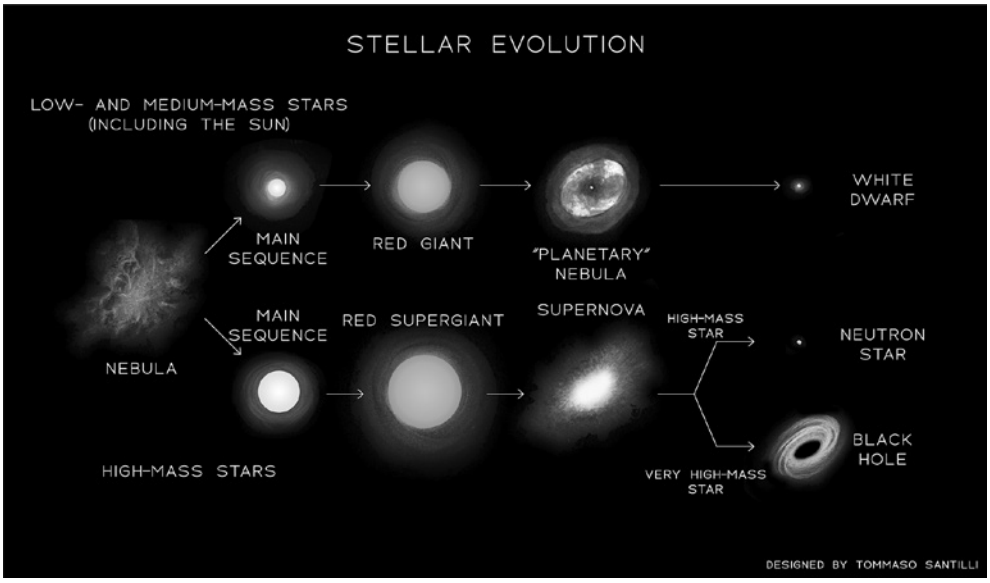


FIG. 1 Schema che riporta gli stadi principali nell'evoluzione stellare.

proporzionale alla loro intensità. Di conseguenza, la formazione di elio si protrae fino a che la pressione interna non torna a bilanciare l'attrazione gravitazionale. Il gas smette quindi di contrarsi e la stella rimane stabile, fintanto che il calore generato dalle reazioni nucleari è sufficiente ad impedire il collasso della stella su se stessa. Ciò avviene fino a che non si esauriscono le riserve di idrogeno e di altri combustibili nucleari. Maggiori sono le riserve iniziali, più rapido sarà il loro esaurimento: il calore necessario a mantenere stabile una stella è tanto più elevato quanto lo è la sua massa. Le stelle molto massive, quindi, tenderanno a terminare rapidamente le fonti di calore interno.

Una prima ipotesi di quanto accade dopo il raggiungimento di questo stadio arrivò nel 1928 da parte di Subrahmanyan Chandrasekhar, fisico e astrofisico indiano. Il modello da lui proposto prevedeva una contrazione della stella fino a che le sue particelle non si fossero trovate molto vicine tra loro. A questo punto sarebbero subentrate delle forze repulsive maggiori in modulo a quelle gravitazionali. Queste forze ulteriori dipendono dalle differenze tra le velocità delle particelle. In casi come questo la stella non solo evita di collassare ulteriormente, si espande fino al raggiungimento di un nuovo equilibrio, quello tra l'attrazione gravitazionale e la repulsione tra elettroni. Una stella che raggiunge questo stato finale è detta "nana bianca". Ne sono state osservate svariate: una delle prime fu avvistata in orbita attorno a Sirio.

Tuttavia, c'è un limite alla repulsione: la massima differenza di velocità tra le particelle è limitata dalla velocità della luce. Dalla Relatività speciale di Einstein sappiamo infatti che la massima velocità che un corpo può assumere è quella della luce, che ha come valore fisso $c = 299792458$ m/s. Ciò implica che, se la stella diventa particolarmente densa, l'attrazione gravitazionale prevale sulla repulsione e la stella continua a collassare su se stessa. Il limite massimo per la massa di una stella affinché essa possa contrastare il proprio collasso è noto come "limite di Chandrasekhar".

Per una stella con massa superiore al limite di Chandrasekhar esiste un'ulteriore configurazione stabile, ancora più contratta e densa rispetto alla nana bianca, con le interazioni gravitazionali bilanciate dalla repulsione neutrone-protone: la stella di neutroni.

La Relatività generale fornisce una visualizzazione un po' più tecnica di questi fenomeni. Se si va a considerare la traiettoria dei raggi di luce nello spazio-tempo, si nota come il campo gravitazionale della stella ne modifica la geometria. I coni di luce rappresentati in Fig. 2, che delimitano la regione di spazio raggiungibile da un osservatore posto nel loro centro, in prossimità della superficie della stella vengono leggermente deflessi verso l'interno e la deflessione aumenterà con la contrazione della stella. In corrispondenza di un raggio critico, i coni verranno completamenti deflessi all'interno e la luce non potrà evadere dal campo gravitazionale della stella. Dato che secondo la teoria della relatività niente può viaggiare ad una velocità supe-

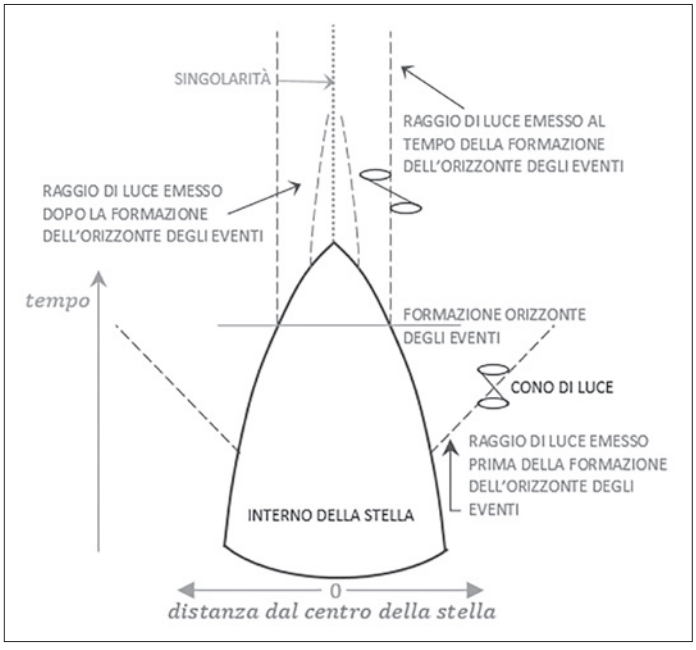


FIG. 2

riore a quella della luce, nulla potrà sottrarsi all'attrazione di un corpo simile. Si ha una regione dello spazio-tempo da cui non è possibile sfuggire per raggiungere un osservatore lontano. Tale regione è ciò che chiamiamo buco nero, nella definizione data da Hawking, e al centro di essa si trova una singolarità (un punto dello spazio-tempo con curvatura tendente ad infinito).

Il raggio critico della stella per cui si ottiene la deflessione completa della luce prende il nome di "raggio di Schwarzschild". Per ricavarne l'espressione matematica, si possono considerare un pianeta di massa M ed un satellite di massa $m \ll M$ in orbita attorno ad esso ad una distanza r (approssimabile al raggio del pianeta). Detta v la velocità del satellite, l'energia cinetica del satellite sarà $E_c = \frac{1}{2}mv^2$, mentre quella potenziale gravitazionale $U_G = G\frac{mM}{r}$, con $G = 6.67 \cdot 10^{-11}\text{Nm}^2/\text{kg}^2$ costante di gravitazione universale.

Si definisce "velocità di fuga" quel valore di v per cui $E_C = U_G$, ovvero $v_F = \sqrt{2GM/r}$. La velocità di fuga rappresenta quella minima che il satellite deve avere per poter sfuggire all'attrazione gravitazionale del pianeta. Il raggio di Schwarzschild è il raggio di un corpo celeste tale per cui la velocità di fuga è pari a quella della luce:

$$R_S = \frac{2GM}{c^2}$$

Esso individua il confine del buco nero, denominato orizzonte degli eventi (vedi Fig. 3). La proporzionalità con la massa è fondamentale: un aumento di essa, come nel caso della caduta di un corpo oltre l'orizzonte, comporterebbe un aumento del raggio di Schwarzschild.

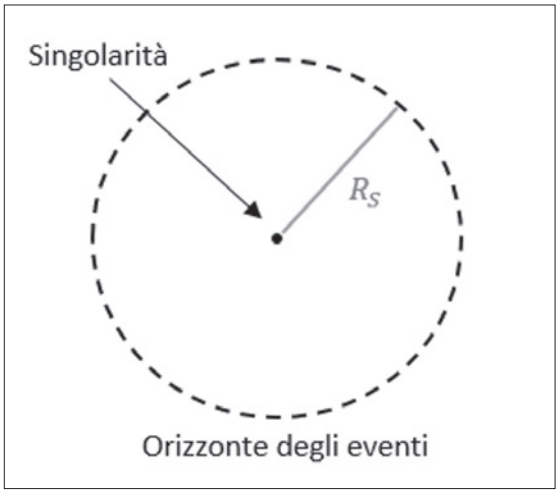


FIG. 3

Passiamo ora ad analizzare il destino della materia che cade in un buco nero.

Se il buco nero fosse di massa paragonabile a quella del Sole (circa $1.989 \cdot 10^{30}$ kg) la forza di gravità nei pressi dell'orizzonte degli eventi sarebbe estremamente intensa e assolutamente non uniforme, dato che agirebbe in direzione radiale. Qualsiasi corpo verrebbe disintegrato appena superato l'orizzonte.

In caso di buchi neri con raggio molto più elevato, l'attraversamento dell'orizzonte degli eventi sarebbe praticamente indolore: dato l'elevato raggio di curvatura, e quindi la lontananza dalla singolarità, la gravità sarebbe approssimativamente uniforme. Tuttavia, l'attrazione verso il centro del buco nero, con conseguente disintegrazione, sarebbe inevitabile.

L'aspetto dei buchi neri che suscita più dubbi non è tanto la singolarità, quanto l'orizzonte degli eventi, detto anche *punto di non ritorno*: questa superficie sferica puramente geometrica può essere attraversata da materia o radiazione in caduta nel buco nero, ma nessun segnale può uscire da essa. Da qui il termine "buco nero": un osservatore esterno non potrà osservare nulla di ciò che accade oltre il punto di non ritorno.

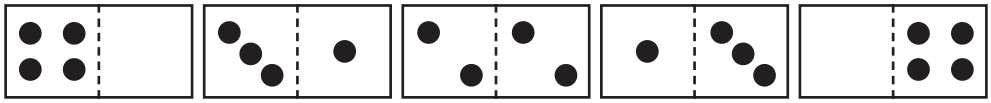
Una delle caratteristica chiave di questi "colossi" celesti è che osservatori diversi avranno percezioni paradossalmente diverse degli stessi eventi. Prendiamo come esempio teorico due astronauti, Andrew ed Agatha, il primo in orbita a distanza di sicurezza e la seconda in caduta verso il buco nero ma con un dispositivo per inviare segnali luminosi. Andrew vedrà i segnali luminosi di Agatha diminuire di frequenza mano a mano che si avvicina al punto di non ritorno, fino a vedere Agatha "ferma" in prossimità dell'orizzonte degli eventi. Dal suo punto di vista, Agatha non cadrà nel buco nero, perché l'attrazione gravitazionale annulla quasi completamente la propagazione dell'onda luminosa. Agatha, nel proprio sistema di riferimento, supererà senza accorgersene il punto di non ritorno e avrà ancora un po' di tempo prima di venire distrutta dalla singolarità. Questo tempo "bonus" dipenderà dalle dimensioni, e quindi dalla massa, del buco nero. Ad esempio, se la massa fosse pari a quella del Sole, Agatha avrebbe circa 10 μs di vita. Le due versioni, quella di Andrew e quella di Agatha, costituiscono un evidente paradosso. Si tornerà su questo aspetto più avanti.

TEMPERATURA ED ENTROPIA DEI BUCHI NERI: IL PARADOSSO DELL'INFORMAZIONE

Ad ogni sistema fisico possono essere associate varie grandezze che ne descrivono lo stato. Andremo a considerare soprattutto la temperatura e l'entropia.

L'entropia S è una misura del numero di configurazioni del sistema conformi ad uno specifico criterio riconoscibile ed è associata al grado di disordine di un

sistema. Prendiamo come esempio una scatola contenente, per semplicità, 4 molecole perfettamente identiche di un gas. Se la scatola contenesse al suo interno un divisorio che lasciasse passare indifferentemente le molecole da un settore all'altro, la possibile disposizione delle molecole sarebbe una tra le seguenti 5:



Immaginiamo ora di mettere tutte le molecole nel settore di sinistra della scatola e poi di lasciar espandere liberamente il gas. Andando poi ad aprire la scatola dopo un certo intervallo di tempo, le molecole saranno molto probabilmente disposte nella configurazione "c". Il fenomeno è descritto dal secondo principio della termodinamica: l'entropia di un sistema isolato aumenta sempre e, se si uniscono due sistemi, l'entropia del sistema risultante sarà maggiore della somma delle entropie dei singoli sistemi. Andando ad applicare quanto detto alla scatola, risulta chiaro il perché della configurazione "c": se le molecole fossero distinguibili l'una dall'altra, avremmo ben 6 modi distinti di disporle per ottenere il caso "c", mentre negli altri casi le combinazioni possibili sono molte meno. Il caso "c" è dunque quello ad entropia maggiore e quindi quello verso cui molto probabilmente evolverà il sistema.

In generale, il secondo principio della termodinamica implica che i sistemi tendono a raggiungere, nel tempo, stati sempre più uniformi ed omogenei. In tali stati finali, allo stesso aspetto macroscopico del sistema corrispondono molte più configurazioni microscopiche rispetto allo stato iniziale.

Il concetto di entropia assume un significato leggermente diverso nella teoria dell'informazione. Immaginiamo di dover inviare un messaggio in codice Morse e che per farlo siano necessari ad esempio 30 simboli tra punti e linee. Se il messaggio fosse composto a caso e non da un telegrafista, i possibili messaggi risultanti sarebbero 2^{30}, pochi dei quali fornirebbero un qualcosa di intellegibile. Possiamo in ogni caso associare un'entropia al sistema in esame: se prendiamo come criterio riconoscibile il numero medio di simboli necessari a codificarlo correttamente, otteniamo come valore per l'entropia 30, che è il logaritmo in base 2 del numero di possibili configurazioni.

Nell'esempio è stato impiegato il codice Morse, che è un *codice binario*. Si è in presenza di un codice binario ogni volta che per comporre un messaggio possono essere impiegati solo due simboli (il punto e la linea nell'esempio, 0 e 1 in informa-

tica). Un linguaggio basato su un codice simile è molto più semplice rispetto, ad esempio, all'alfabeto. Ogni simbolo impiegato nella sequenza costituente il messaggio dà un'informazione minima. Se fosse impiegato l'alfabeto, ogni lettera porterebbe in sé già un grande quantitativo di informazione: quella che la distingue dalle altre lettere. Il codice binario fornisce quindi l'unità minima e irriducibile dell'informazione: il *bit*. Il bit può assumere solo il valore di uno dei due simboli impiegati dal codice binario e quindi un messaggio di qualsiasi tipo sarà composto da una sequenza ordinata di bit.

L'informazione, intesa in senso fisico, è materia (particelle del Modello Standard, fotoni, etc.) localizzata in un volume ben definito.

Tornando all'esempio della scatola, possiamo esprimere la posizione di ciascuna molecola in questo modo: se si trova a sinistra del divisorio, bit con valore 1, se si trova a destra, bit con valore 0. Così facendo associamo ad ogni configurazione una sequenza di 0 e 1 ed il criterio riconoscibile, necessario per definire l'entropia, diventa il numero di bit con cui descriviamo la configurazione, ovvero 4.

Generalizzando, se si adotta come criterio riconoscibile il numero di bit necessari n per codificare in modo univoco una configurazione di un sistema, allora il numero di configurazioni possibili sarà 2^n e l'entropia sarà, per definizione: $S = \log_2 2^n = n$. In sostanza, S rappresenta l'informazione "nascosta" nel sistema, perché tiene conto delle parti di esso che sarebbero troppo piccole o numerose per essere distinguibili alla scala a cui si sta osservando il sistema. Chiaramente, se la configurazione possibile è unica, allora l'entropia sarà zero.

In quest'ottica si inserisce il nesso tra entropia e temperatura. Quando si fornisce calore ad un sistema, se ne aumenta la temperatura e di conseguenza l'agitazione termica. Questo comporta una perdita dell'informazione contenuta nel sistema, perché aumenterà il numero delle possibili disposizioni molecolari per uno stesso stato e quindi anche il grado di incertezza con cui lo si descrive. Questo vuol dire che fornire calore ad un sistema ne aumenta l'entropia. Inoltre, l'energia "organizzata" (cinetica, chimica, etc.) tende a degradarsi in calore. È quanto previsto dal secondo principio della termodinamica: il sistema tende ad evolvere verso stati a maggior entropia. Per questo motivo non vedremo mai una macchina arrugginita tornare ad essere nuova fiammante o un uovo rotto tornare integro: andrebbe contro il secondo principio!

Andiamo ora a considerare i buchi neri: la loro caratteristica principale è ingerire tutto ciò che attraversa il loro orizzonte degli eventi. Questo comprende ovviamente anche corpi dotati di entropia. Verrebbe quindi spontaneo concludere che gettare della materia con una certa quantità di entropia in un buco nero consenti-

rebbe di diminuire spontaneamente l'entropia totale della materia che si trova all'esterno dell'orizzonte degli eventi. Anche se si considerassero come un unico sistema il buco nero e ciò che lo circonda, l'entropia della materia interna al buco nero non sarebbe misurabile da un osservatore esterno. Quello che abbiamo ottenuto sembra una chiara violazione del secondo principio della termodinamica. La questione potrebbe essere facilmente risolta se esistesse un modo per l'osservatore esterno di avere delle indicazioni sull'entropia all'interno del buco nero.

Nella precedente sezione si era messo in evidenza che il raggio del buco nero dipende dalla sua massa, che aumenta ogni volta che una porzione di materia oltrepassa l'orizzonte degli eventi. Si ha in pratica che ad un aumento della massa del buco nero corrisponde un aumento della superficie del suo orizzonte. Questo consente di stabilire un legame tra la superficie dell'orizzonte e la quantità di materia dotata di entropia che viene inghiottita. Inoltre, si salva il secondo principio: la superficie di un buco nero può solo aumentare, così come la sua massa, perché nulla può sfuggire alla sua attrazione gravitazionale. Allo stesso modo anche l'entropia complessiva del buco nero e della materia che lo circonda tenderà ad aumentare.

Il problema, a questo punto, diventa stabilire di quanto cambia il raggio di un buco nero se vi si lascia cadere un singolo bit di informazione. Per simulare il singolo bit si può ricorrere ad un fotone che abbia la posizione il più indeterminata possibile: imponiamo che la sua lunghezza d'onda sia tanto grande da coprire l'intero buco nero. In questo modo, l'esistenza del fotone avrà un unico bit di informazione: fotone all'interno o all'esterno dell'orizzonte.

Il calcolo fu svolto dal fisico Jacob Bekenstein e ciò che ottenne fu che aggiungere un bit di informazione fa crescere la superficie dell'orizzonte di un buco nero di circa una lunghezza di Planck al quadrato ($\sim 10^{-70} m^2$).

Immaginando di costruire il buco nero un bit alla volta, quello che si ottiene è che l'area del suo orizzonte è proporzionale al numero di bit "ingeriti", e quindi alla sua entropia. In base a questo ragionamento, l'orizzonte degli eventi sarebbe una sorta di superficie fittamente ricoperta di bit materiali, in netto contrasto con la sua definizione iniziale di "punto di non ritorno". È questa la contraddizione che sta alla base della Guerra dei buchi neri.

Stephen Hawking vide nel lavoro di Bekenstein una terribile pecca: ogni corpo dotato di entropia ha anche associata una temperatura, perciò anche i buchi neri dovevano avere una temperatura. Tuttavia, un corpo con una particolare temperatura deve emettere radiazione con un certo tasso. Come può un buco nero emettere radiazione, se niente può sfuggire, per definizione, alla sua attrazione gravitazionale?

Hawking sfruttò la matematica della teoria quantistica dei campi per rispondere a questa domanda. La teoria quantistica dei campi prevede che il campo elettromagnetico possa avere "tremori quantistici" anche in assenza di perturbazioni da parte di particelle cariche. Queste fluttuazioni quantistiche sono una conseguenza del principio di indeterminazione di Heisenberg.

Consideriamo un gas in un recipiente che viene raffreddato per sottrazione di calore: la temperatura minima che può raggiungere è lo zero assoluto (–273.15°C). Gli atomi, avendo perso tutta la loro energia, dovrebbero essere perfettamente fermi. Non si sta tenendo conto, però del principio di indeterminazione: indipendentemente dal grado di incertezza sulla posizione degli atomi Δx, la velocità sarebbe identicamente nulla per tutte gli atomi e Δv sarebbe zero. Questo viola il principio di Heisenberg, in base a cui deve verificarsi $4\pi m\Delta x\Delta v > h$, con m massa dell'atomo e $h = 6.626 \cdot 10^{-34}$Js costante di Planck.

Dato che Δx non può essere infinita (il gas è contenuto in un recipiente), si avrà una sorta di moto fluttuante residuo chiamato *moto di punto zero* o *quantum jitter*. A tutti gli oscillatori è associata un'energia, che nel caso dei tremori quantistici può anche essere molto elevata ma "non è disponibile", perché il sistema è già in una condizione di minimo assoluto di energia. Sono moti molto diversi dai tremori termici, che invece sono causati da un eccesso di energia ed hanno effetti molto più evidenti e violenti.

La teoria quantistica dei campi suggerisce un modo per visualizzare le due tipologie di fluttuazioni. Le *fluttuazioni termiche* sono dovute alla presenza di *fotoni reali* che trasferiscono energia sotto forma di calore. Le fluttuazioni quantistiche invece sono dovute a *coppie di fotoni virtuali* che vengono create e rapidamente riassorbite dal vuoto. Se lo spazio fosse allo zero assoluto, sarebbero questi gli unici moti presenti.

I due tipi di fluttuazione in circostanze normali non si mescolerebbero data la loro diversa natura ma, dato che l'orizzonte degli eventi di un buco nero è un luogo assolutamente anomalo, i suoi effetti sui tremori sono molto interessanti.

Tornando al paradosso di Agatha e Andrew: Agatha si trova in viaggio verso la singolarità del buco nero (in un ambiente allo zero assoluto) ed è circondata da coppie di fotoni virtuali di cui non ha la minima percezione perché non le trasmettono calore. Dal punto di vista di Andrew invece la presenza dei tremori quantistici è molto rilevante. Alcune coppie potrebbero essere in parte dentro e in parte fuori dall'orizzonte degli eventi. Andrew però vede solo il fotone singolo rimasto fuori e non può capire che appartiene ad una coppia. Per lui è un fotone termico reale. Andrew osserverà l'inesorabile caduta di Agatha verso una zona incredibilmente

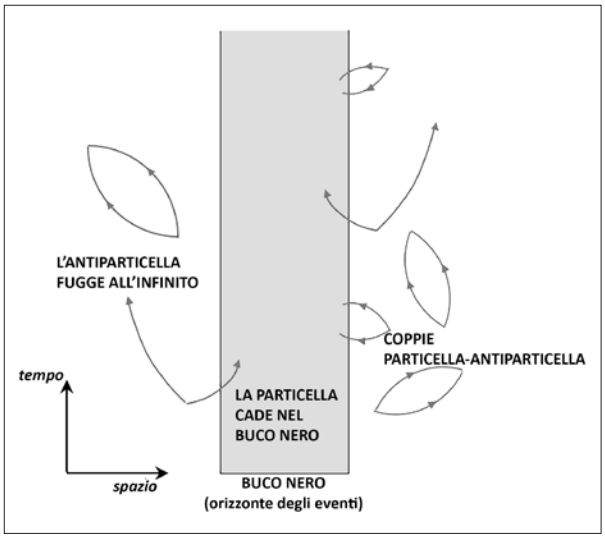


FIG. 4

calda e la vedrà vaporizzarsi. Questo è chiaramente un paradosso: come è possibile che Agatha attraversi incolume l'orizzonte degli eventi nel proprio sistema di riferimento se Andrew la vede disintegrarsi?

L'esempio mette bene in evidenza quello che è stato definito il "paradosso dell'informazione". La Relatività generale e la meccanica quantistica prevedono due diversi comportamenti per la materia in prossimità del punto di non ritorno. Secondo la prima, l'informazione attraversa l'orizzonte degli eventi rimanendo inalterata. Nel caso quantistico invece viene distrutta prima ancora di attraversare l'orizzonte: attorno al buco nero l'interazione della materia con i fotoni termici rimescola i bit alla rinfusa, per poi riemetterli sottoforma di fotoni o altre particelle, come se fossero i prodotti di un'evaporazione. La contraddizione che emerge dalle due versioni è evidente.

Hawking calcolò che la perturbazione delle fluttuazioni del vuoto causata dalla presenza del buco nero fa sì che vengano emessi veri e propri fotoni termici[1], come se il buco nero fosse un corpo nero[2] caldo. Questi fotoni si chiamano *radiazione di Hawking*. A partire da questo, Hawking ricavò matematicamente che l'entropia di un buco nero è un quarto dell'area dell'orizzonte misurata in unità di Planck e che la sua temperatura segue una legge del tipo:

[1] S.W. Hawking, *Particle creation by black holes*, in *Commun. math. Phys.*, 43, 1975, pp. 199-220.
[2] Un corpo nero è un oggetto che assorbe completamente la luce e la irraggia, ma non la riflette.

$$T = \frac{hc^3}{16\pi^2 GMk} \tag{1}$$

dove c è la velocità della luce, G la costante di gravitazione universale, h la costante di Planck, M la massa del buco nero e k la costante di Boltzmann ($\sim 1.38 \cdot 10^{-23}$ J/K).

A mano a mano che viene emessa radiazione di Hawking, il buco nero evapora. I fotoni espulsi hanno infatti una loro energia e di conseguenza, per la relazione $E = mc^2$, la massa del buco nero dovrà diminuire. Applicando allora la relazione (1) si otterrà che il buco nero man mano che perde massa e si contrae si scalda fino a temperature elevatissime, comparabili con quella che potrebbe aver caratterizzato il Big Bang. Bisogna sottolineare che il tasso con cui viene emessa radiazione di Hawking è incredibilmente basso. Consideriamo un buco nero di massa pari a quella del Sole ($\sim 1.989 \cdot 10^{30}$ kg): la sua temperatura sarebbe, in base alla relazione (1), circa 10^{-7} K, molto vicina allo zero assoluto. A causa di questa temperatura così bassa, il buco nero impiegherebbe per evaporare completamente un tempo dell'ordine di 10^{66} anni, di gran lunga superiore all'età del nostro Universo. Rimane tuttavia un'incognita: cosa accade all'informazione che era stata stipata nel buco nero?

LE SOLUZIONI PROPOSTE: IL CONVEGNO DI SANTA BARBARA

Nel 1976 Hawking pubblicò un articolo[3] in cui affermava che i bit di informazione caduti in un buco nero non sarebbero stati semplicemente rimescolati fino a rendere inutilizzabile l'informazione da essi rappresentata. Una volta raggiunta la singolarità, non avrebbero avuto alcun modo di essere recuperati e con l'evaporazione del buco nero sarebbero andati definitivamente distrutti.

Un'utile analogia può essere quella della cassaforte: immaginiamo di aver chiuso dei preziosi documenti in una cassaforte e di averne perduto la chiave; fintanto che la cassaforte resta dove l'abbiamo lasciata non ci sono problemi, sappiamo dove si trovano i documenti anche se non possiamo consultarli. Se invece viene distrutta, non ci sarà verso di recuperarli.

Gli effetti di una perdita di informazione estremamente concentrata come quella contenuta in un buco nero sarebbero molto preoccupanti. Perdere informazione equivale a generare entropia e di conseguenza calore. Hawking aveva postulato, nel-

[3] S.W. Hawking, *Breakdown of predictability in gravitational collapse*, in *Phys. Rev. D*, 14, 1976, pp. 2460-2473.

la sua teoria, anche l'esistenza di buchi neri *virtuali* nel vuoto con un ciclo di vita tale da passare inosservati. Il loro effetto sarebbe quello di cancellare l'informazione anche in assenza di buchi neri reali nelle vicinanze. Leonard Susskind calcolò che l'esistenza di questi buchi virtuali avrebbe riscaldato il vuoto fino a miliardi di miliardi di gradi in una frazione di secondo.

Quello che preoccupava di più Susskind e 't Hooft era che nella teoria di Hawking si contraddiceva il principio di conservazione dell'informazione, uno dei capi saldi della meccanica quantistica. Il principio di conservazione è legato alla reversibilità delle leggi fisiche: se si conosce perfettamente la struttura microscopica di un dato sistema e lo si sottopone ad una trasformazione, in via teorica è sempre possibile ricostruire a partire dallo stato finale del sistema quello iniziale. La meccanica quantistica, nonostante sia intrinsecamente probabilistica, ha una struttura matematica molto sottile per cui mantiene questa proprietà.

L'approccio di 't Hooft alla Guerra dei buchi neri si basò proprio sulla reversibilità. In fisica delle particelle, l'ambito di ricerca di 't Hooft, si è soliti rappresentare le collisioni tra particelle con un'astrazione matematica: la *matrice S*. Questa matrice è una sorta di tabella che lega le particelle della collisione con tutti i possibili prodotti. Gli elementi della tabella sono delle probabilità (*ampiezze di probabilità*, per la precisione) e riassumono la dinamica della collisione. La particolarità della matrice S è la reversibilità: si tratta infatti di una matrice invertibile ma terribilmente complicata da scrivere, perché deve tener conto di tutti i possibili risultati della collisione. Pensiamo ad una bomba innescata da un elettrone: il risultato della collisione saranno migliaia di frantumi e per poter scrivere la matrice S occorrerebbe associare a ciascun frammento una direzione e una velocità e poi associare un'ampiezza di probabilità ad ogni possibile configurazione finale. L'aspetto fondamentale della matrice S è che ci consente non solo di predire lo stato finale di un sistema conoscendo quello iniziale, possiamo anche ricostruire il processo inverso grazie alla proprietà di invertibilità. Secondo 't Hooft la formazione e l'evaporazione di un buco nero sono solo un caso molto intricato di collisione di particelle e quindi sostanzialmente rappresentabile con una complicatissima matrice S. Il problema che si presentò fu capire quali fossero gli oggetti a cui applicare la matrice S e da quali leggi fosse governata la loro interazione. Occorreva, in sostanza, trovare l'origine microscopica dell'entropia del buco nero. Non bisogna dimenticare infatti che nella scorsa sezione, per associare un'entropia al buco nero, si era considerata solo la relazione tra entropia della materia caduta oltre il punto di non ritorno e variazione del raggio del buco nero.

Hawking rispose alle argomentazioni di 't Hooft proponendo un'altra matrice:

la *matrice* $. Secondo lui una matrice invertibile come la matrice S non era adatta per descrivere l'evoluzione dei buchi neri. Dato che l'unica quantità in grado di uscire dal buco nero è la radiazione di Hawking, la matrice $ venne costruita in modo tale da fornire invariabilmente come risultato la radiazione di Hawking, anche nel caso in cui fosse stata essa stessa lo stato di partenza.

La svolta decisiva arrivò nel 1993 al convegno di Santa Barbara. Erano presenti, tra altri eminenti fisici, i tre protagonisti della Guerra dei buchi neri. L'idea avanguardistica, tuttavia, questa volta arrivò da Susskind.

Susskind era convinto che l'approccio corretto fosse quello di 't Hooft e si era dedicato quindi alla ricerca di una possibile origine microscopica dell'entropia dei buchi neri. Una delle conclusioni a cui era giunto riguardava l'orizzonte degli eventi. Nella sezione precedente si era detto che la materia, per un osservatore esterno, doveva scaldarsi, vaporizzarsi ed essere riemessa sottoforma di radiazione di Hawking prima di raggiungere l'orizzonte. Per fornire un'interpretazione coerente di questo fenomeno, Susskind ipotizzò l'esistenza di uno strato surriscaldato appena al di sopra dell'orizzonte e costituito da oggetti microscopici, probabilmente con dimensioni non superiori alla lunghezza di Planck (10^{-35} m). Lo strato venne poi denominato *orizzonte allargato* ed è ad oggi uno dei concetti standard della fisica dei buchi neri.

La presenza di un orizzonte allargato chiarisce anche il perché dell'evaporazione dei buchi neri: gli atomi dell'orizzonte sono molto energetici e in caso di collisioni particolarmente violente possono essere espulsi dalla prossimità dell'orizzonte verso lo spazio più esterno.

Introdurre l'orizzonte allargato fu la rampa di lancio per ciò che sconvolse l'auditorium di Santa Barbara: il *principio di complementarietà dei buchi neri*.

Il termine "complementarietà" non venne scelto casualmente: la risposta di Susskind al paradosso dell'informazione consisteva nell'affermare che, considerando sempre l'esempio di Andrew ed Agatha, *entrambe* le versioni riportate dai due osservatori sono vere.

Una contraddizione è tale solo quando è possibile confrontare i dati raccolti. Prendiamo ad esempio due termometri che devono misurare la temperatura di una stessa pentola d'acqua. Se i due termometri riportano valori diversi, risulta istintivo concludere che uno dei due termometri deve essere calibrato male, la contraddizione tra i risultati della misurazione è evidente. Nel caso dei buchi neri, però, non è possibile confrontare i dati raccolti da un osservatore esterno con quelli dell'osservatore che ha oltrepassato il punto di non ritorno. Dato che niente può uscire dal buco nero, l'unico modo che l'osservatore esterno avrebbe per confrontarsi col suo

collega sarebbe oltrepassare di persona l'orizzonte, ma così facendo perderebbe il suo ruolo di osservatore esterno. Susskind aveva intuito che non sarebbe stato possibile ideare un esperimento in cui coesistessero sia l'osservazione dall'esterno sia l'attraversamento del punto di non ritorno. La conclusione fu che le due situazioni dovevano essere perfettamente complementari, proprio come lo sono il comportamento ondulatorio e quello particellare della luce.

Al convegno Susskind utilizzò il principio di complementarietà per spiegare la sua teoria sul destino dell'informazione caduta nel buco nero. Per l'osservatore esterno l'informazione che si avvicina all'orizzonte si comporta in modo simile ad una goccia di inchiostro che cade in una bacinella piena d'acqua. La goccia viene assorbita e rimescolata fino ad essere irriconoscibile, ma se l'acqua comincia ad evaporare trasporterà con sé anche le molecole di inchiostro.

Prima dell'intervento di Susskind, le opinioni dei presenti in merito al paradosso dell'informazione potevano essere così suddivise:

1. *opinione Hawking*: l'informazione che cade in un buco nero è definitivamente perduta;
2. *opinione 't Hooft-Susskind*: l'informazione filtra all'esterno attraverso i fotoni e le altre particelle che costituiscono la radiazione di Hawking;
3. l'informazione rimane confinata in un residuo di dimensioni dell'ordine della scala di Planck;
4. altre teorie.

La teoria del residuo di informazione si basa sul fatto che, durante l'evaporazione, il buco nero arriva a raggiungere temperature altissime e quindi negli ultimi stadi le particelle emesse avrebbero un'energia elevatissima, molto al di sopra di quella di cui abbiamo esperienza. Quando il raggio del buco nero raggiunge la lunghezza di Planck, l'evaporazione potrebbe anche arrestarsi, perché la sua massa corrisponderebbe alla massa di Planck ($2{,}17645 \cdot 10^{-8}$ kg). Rimarrebbe quindi un residuo con intrappolata al suo interno tutta l'informazione del buco nero. Sarebbe una sorta di particella capace di nascondere una quantità pressoché infinita di informazione. Un oggetto simile avrebbe entropia infinita e la sua esistenza sarebbe un disastro termodinamico. Questa ipotesi non riscosse infatti un grande successo.

La voce "altre teorie" comprendeva l'idea dei "baby universi". In base a questa teoria, durante l'evaporazione potrebbe staccarsi una porzione di spazio-tempo dal buco nero, dando vita a dei nuovi universi autosufficienti e sconnessi rispetto a noi. L'informazione rimarrebbe intrappolata in questo nuovo spazio e al termine dell'evaporazione sarebbe definitivamente irrecuperabile. L'idea in sé per sé non è poi così assurda, dato che il nostro stesso universo è in espansione. Non costituisce però

una soluzione al paradosso, dato che l'informazione sarebbe comunque inosservabile nel nostro sistema di riferimento.

Al termine della conferenza venne effettuato un piccolo sondaggio su quale delle 3 teorie fosse quella più plausibile e, con somma soddisfazione di Susskind, la vincitrice fu proprio l'opzione 2.

Negli anni successivi alla conferenza di Santa Barbara 't Hooft e Susskind continuarono a cercare di dimostrare la loro opinione sul destino dell'informazione. La dimostrazione arrivò grazie al principio olografico, secondo il quale tutto ciò che è contenuto in una data regione spaziale può essere descritto da bit di informazione confinati sul bordo della regione stessa, proprio come i pixel di un ologramma. Nella dimostrazione si assume tuttavia la validità della teoria delle stringhe, che come è stato detto nell'introduzione non ha ancora ricevuto le necessarie conferme sperimentali. Possiamo dire quindi che il paradosso dell'informazione è stato risolto da Susskind e 't Hooft ma solo in un mondo descritto dalla teoria delle stringhe.

Hawking invece rimase della propria opinione dopo la conferenza di Santa Barbara e continuò a pubblicare lavori molto validi sull'argomento, ma in seguito ai lavori di Susskind e 't Hooft perse una scommessa col fisico John Preskill. La scommessa riguardava il fatto che l'informazione potesse sfuggire ai buchi neri, cosa che era stata dimostrata, anche se mediante la teoria delle stringhe. Preskill ottenne come pegno un'enciclopedia sul baseball.

BIBLIOGRAFIA ESSENZIALE

Hawking, S.W., *Dal Big Bang ai buchi neri. Breve storia del tempo*, Milano, Rizzoli, 2000.
Susskind, L., *La guerra dei buchi neri*, Milano, Adelphi, 2009.

WHAT CAN ECONOMISTS LEARN FROM MACHINE LEARNING?

KEVIN MICHAEL FRICK

Il *machine learning* gode di notevole popolarità in econometria perché può essere usato come strumento per predire con precisione il comportamento umano. Questo articolo analizza alcuni case study di interesse, spiegando i processi coinvolti da un punto di vista ingegneristico e informatico.
Si introduce per prima cosa una "definizione operativa del *machine learning*" e viene chiarita la posizione del machine learning all'interno dell'ambito dell'intelligenza artificiale. Viene poi esposta una review dell'utilizzo di modelli di machine learning per l'elaborazione e l'interpretazione di big data, accompagnata da appunti sul funzionamento degli approcci di modellazione utilizzati.

Machine learning is becoming increasingly popular in econometrics as it can be used as a tool to make accurate predictions of human behavior. This paper analyzes a select few notable case studies, explaining the processes involved from a computer engineering point of view. An "operational definition of machine learning" is introduced first, then the position of machine learning as a subset of research on artificial intelligence is clarified. An overview of how machine learning can be used to process and interpret big data is then given and accompanied by introductory paragraphs detailing the inner workings of relevant modeling approaches.

PREFAZIONE

L'economia è una disciplina sociale ambiziosa, che si pone l'obiettivo di prevedere il comportamento umano, le scelte individuali e collettive, in tutti quei contesti in cui le persone si trovano a dover fronteggiare un trade-off tra costi e benefici.

Questa capacità di prevedere con la maggior precisione possibile come le persone si comporteranno in un determinato contesto, o come reagiranno al mutamento di alcune circostanze – ad esempio, un aumento dei prezzi, una riduzione delle imposte, o una variazione del tasso di inflazione – non è un fine in sé, ma un mezzo. È lo strumento necessario al decisore sociale – tipicamente un manager, o un politico – per valutare quali siano le strategie più efficaci per massimizzare la propria "funzione obiettivo", utilizzando gli strumenti e muovendo le leve a propria disposizione.

La "funzione obiettivo" che il decisore sociale persegue, in sé e per sé, non è oggetto di analisi economica. Per l'economista si tratta di un elemento "esogeno". A seconda dell'orientamento politico, un amministratore pubblico può desiderare di massimizzare il benessere della collettività considerata nella sua interezza, oppure focalizzarsi su alcune sue componenti: tutelare i lavoratori, proteggere le fasce più deboli, oppure promuovere l'efficienza e la produttività delle imprese, stimolare l'innovazione, la competitività dell'industria nazionale, o le esportazioni. Un amministratore privato, tipicamente, avrà il compito di massimizzare i profitti e la redditività dell'impresa che dirige.

L'economista è un tecnico, il cui compito è stabilire come le risorse a disposizione possano essere usate nel modo più efficiente possibile per perseguire un obiettivo dato. Il problema è che i meccanismi che trasformano gli input (le risorse) in output (i risultati) ruotano su perni e ingranaggi indisciplinati, il cui funzionamento raramente segue regole precise. Questi perni e ingranaggi sono le persone che compongono le società umane, con le loro peculiarità e idiosincrasie determinate dalle caratteristiche fisiche individuali, dalle vicende personali e dal contesto culturale in cui gli uomini crescono e vivono. Dunque, la previsione dell'economista – tipicamente sintetizzata da un modello matematico che rappresenta in modo schematico e semplificato una serie di relazioni causa-effetto – non può che essere imprecisa e necessariamente circostanziata. Inevitabilmente soggetta ad eccezioni. Le persone non si comportano *tutte* e *sempre* nello stesso modo. Le loro decisioni individuali sono spesso determinate da elementi secondari, difficilmente osservabili e controllabili: caratteristiche quasi impercettibili del contesto di scelta che possono indurre la medesima persona ad agire in modo diverso anche in circostanze apparentemen-

te molto simili, e che a volte inducono persone diverse a compiere scelte opposte, a fronte dello stesso problema decisionale.

Nell'ultimo decennio, si è aperto per gli economisti e per gli scienziati sociali in generale, un panorama completamente nuovo. La disponibilità di dati estremamente disaggregati su aspetti anche molto privati ed intimi del comportamento individuale di milioni di persone ha reso teoricamente possibile una indagine più dettagliata, quasi microscopica, delle determinanti delle decisioni umane. E tuttavia questa mole straordinaria di dati rimane inaccessibile, inutilizzabile in assenza di strumenti e metodologie adeguate. Non si può mettere il vino nuovo in otri vecchi, altrimenti il vino nuovo spacca gli otri: le metodologie di elaborazione ed analisi dei dati disponibili fino alla fine del ventesimo secolo non sono in grado di gestire e contenere la mole sempre crescente di informazione. È in questo contesto che si sono sviluppati gli algoritmi di *machine learning*, che sono in grado di ridurre la complessità delle dinamiche umane, sintetizzando ed evidenziando le relazioni causali predominanti, in modo da renderle comprensibili alla mente umana.

In questo saggio Kevin Michael Frick illustra le potenzialità di questo strumento per l'analisi economica del comportamento umano, attraverso tre casi di studio che permettono al lettore di comprendere la versatilità delle metodologie descritte, ma anche di cogliere l'essenza dei diversi algoritmi di machine learning, presentati in modo tecnicamente accurato, ma non tecnicistico.

Il primo studio riguarda le scelte in condizioni di incertezza, cioè le decisioni che gli individui devono prendere in circostanze in cui le conseguenze delle proprie azioni sono in parte determinate da elementi stocastici incontrollabili. Poiché l'incertezza è condizione caratterizzante della vita umana e ne permea quasi tutti gli ambiti, è evidente che capire quali considerazioni guidino le persone in questi contesti è fondamentale, specialmente per gli economisti. Il modello più diffuso in economia per prevedere le scelte in condizioni di incertezza è quello dell'Utilità Attesa. Questo studio illustra l'applicazione di un algoritmo di *machine learning* a un dataset di decisioni compiute da individui in un esperimento on-line, in cui il contesto di scelta è rigorosamente controllato in modo esogeno. I risultati mostrano che l'algoritmo riproduce il modello standard dell'Utilità Attesa nei contesti di scelta in cui i soggetti – pur non potendo prevedere esattamente le conseguenze delle proprie decisioni – sono comunque in grado di associare una probabilità precisa ad ogni conseguenza possibile. E tuttavia, nei contesti caratterizzati da ambiguità, quando cioè anche le probabilità associate a ciascuna conseguenza sono esse stesse incerte, il modello di Utilità Attesa si rivela inadeguato a predire le scelte umane perché non tiene in considerazione la tendenza delle persone ad evitare que-

sto tipo di "ambiguità" (*ambiguity aversion*); qui l'algoritmo di *machine learning* produce un modello di decisione più accurato rispetto a quello tradizionalmente adottato dagli economisti.

Il secondo studio riguarda le decisioni in contesti strategici, cioè in quelle circostanze in cui le conseguenze delle azioni di un individuo dipendono in modo fondamentale anche da ciò che fanno altri agenti coinvolti nello stesso problema. In genere, in economia lo studio di questo tipo di decisioni si basa su modelli derivati dalla "teoria dei giochi". Per poter formulare delle predizioni precise sul comportamento umano, la teoria dei giochi si basa su alcune ipotesi piuttosto forti sulla razionalità degli individui, e sul fatto che questa razionalità sia conoscenza comune: per valutare le possibili conseguenza delle proprie azioni, ciascun individuo assume che anche gli altri si comporteranno sempre in modo razionale, senza cioè commettere errori sistematici, e che gli altri sappiano che lui è razionale, e così via, *ad infinitum*. Gli economisti comportamentali, negli ultimi decenni, hanno evidenziato come queste ipotesi siano poco credibili dal punto di vista psicologico e spesso portino a modelli che commettono sistematici errori di previsione, ed hanno formulato modelli di decisione alternativi, basati su ipotesi meno restrittive. Uno di questi modelli è quello di "level-k thinking" considerato in questo secondo studio. Qui gli algoritmi di *machine learning* vengono applicati a un set di dati, anche in questo caso generati tramite un esperimento in laboratorio che ha coinvolto come partecipanti dei volontari umani. I risultati rivelano come l'algoritmo permetta di formulare un modello di previsione del comportamento che di fatto è un ibrido tra il modello game-teoretico standard e quello di level-k, che risulta più flessibile e più accurato dal punto di vista predittivo, rispetto a ciascuno dei due modelli originari.

Il terzo studio mostra infine una applicazione di *machine learning* alle decisioni giudiziarie. Affronta quindi un problema più pratico, meno astratto rispetto a quelli affrontati nei due casi-studio precedenti, e si basa sull'utilizzo di dati raccolti sul campo, anziché in laboratorio. La decisione considerata qui è quella del giudice che, all'inizio di un processo, deve stabilire se l'imputato debba rimanere in custodia cautelare o possa essere rilasciato in attesa del processo. Usando un dataset di oltre 750 mila casi di arresto avvenuti a New York, i ricercatori individuano quali caratteristiche osservabili dell'imputato devono essere tenute in considerazione per prevedere la probabilità che questi non si presenti in giudizio, qualora venga lasciato a piede libero. I risultati evidenziano come i giudici tendano a compiere errori sistematici, e suggeriscono che le tecniche di Machine Learning possano essere un supporto utile alle decisioni umane anche in questi contesti, e che possano anche contribuire a ridurre le distorsioni sistematicamente introdotte dalle preferenze

personali dei decisori, che possono risultare, ad esempio, in forme di discriminazione implicita, e potenzialmente involontaria.

Questo saggio evidenzia le potenzialità delle tecniche di *machine learning* applicate alle scienze sociali, ma ne lascia trasparire anche i rischi. La disponibilità di dati sempre più precisi e particolareggiati sulle decisioni umane rendono queste decisioni più prevedibili, e per questo anche più facilmente manipolabili. Queste considerazioni aprono il campo a riflessioni importanti, in ambito politico, sociologico e filosofico, che esulano dallo scopo di questo lavoro ma che sarà comunque necessario affrontare, non solo in ambito accademico, nel prossimo futuro.

Maria Bigoni
Professoressa ordinaria di Economia Politica
Dipartimento di Scienze Economiche
Università di Bologna

HUMAN LEARNING AND MACHINE LEARNING

Humans, like the majority of animals, are able to learn from their experiences. Just like a dog learns it shouldn't eat food that makes it feel sick, people continuously learn throughout their lives: first in school, then at work, and repeatedly during their social, private and public activities. Learning is useful because it doesn't only allow to better react to events that already happened but enables one to draw general rules to govern their behavior when presented with unfamiliar situations. The inability to learn is why machines, however computationally powerful, have historically had trouble carrying out tasks that are trivial to humans. For example, identifying the location of a photographer when they take a picture with their phone involves a trivial look-up of GPS coordinates, but understanding whether the subject of the photo is a bird or a mountain was still a daunting task as of the last decade, requiring very advanced algorithms that were heavy on computational resources.

In recent years, astounding progress has been made and most of us carry a smartphone in our pockets that is able to identify the subjects of pictures with better accuracy than its owner (Karpathy 2014). Machine learning is what enabled researchers to reach this kind of results.

Machine learning is a subfield of research on artificial intelligence which is deeply intertwined with the other branches of AI, but unlike "traditional" research which led, for example, to the Deep Blue chess-playing computer (commonly known as *symbolic artificial intelligence*, see Haugeland 1989), machine learning does not use predetermined rules to decide which action to take or classification to make and instead analyzes a data set in order to develop a series of rules on its own that are then applied to different data sets. This makes machine learning much more interesting to economists since, compared to symbolic artificial intelligence, it can be more easily applied to a very specific subproblem, whereas symbolic artificial intelligence tries to reproduce "general" intelligence, similar to what humans have. The difference between machine learning and human learning boils down to the kind of learning and application of knowledge that takes place: machines are quickly becoming much better than humans at drawing conclusions based on past experiences (i.e. data) and creating models that allow for understanding and linking past and future experiences, even when presented with unfamiliar situations; on the other hand, machine learning algorithms are for the most part still only able to efficiently solve some specific problem, and are incapable of creative and associative thinking, as described in Rudin 2017 – at least, for now.

Scholars from different fields of study often have differing views on what machine learning is. Computer scientists and mathematicians might be considered those having the most "correct" view of what the practices, uses and scopes of machine learning are, but applied disciplines often differ from pure studies. Susan Athey, professor at Stanford Graduate School of Business and consulting researcher to Microsoft, provides an "operational definition of machine learning" in her 2018 paper "The impact of machine learning on economics", defining machine learning as a "field that develops algorithms designed to be applied to data sets". This definition is useful as a starting point to explore what machine learning can be used for. As of 2019, some of its main applications are pattern recognition (e.g. classifying pictures), reinforcement learning (e.g. discovering which characteristics better "fit" an environment), natural language interpretation (e.g. translating ambiguous and context-dependent sentences). Moreover, applications of machine learning to policy analysis problems (e.g. estimating the impact of a public transport cost cut on air pollution levels) are particularly interesting to economists.

Athey's definition, however, is lacking descriptiveness about *how* machine learning accomplishes these results. While a detailed description of the most common machine learning algorithms is outside the scope of this paper, mentioned applications will be supplemented by technical explanations of the methods used.

The so-called "computer revolution" sparked another revolution in statistics: never before had data been so easily available and in similar quantities. As early as 2007 94% of the global information storage capacity was digital and the capacity of digital media alone was more than 10 times the global information capacity in 1986 (Hilbert and López 2011). As of 2014, Facebook's Hive data warehouse alone stores around 4000 new terabytes of data *every day* (Wiener and Bronson 2014) which in 1986 was about one tenth of the *global* information capacity. Spreadsheets are clearly unfit to handle such a large amount of data, but even traditional relational databases such as MySQL, conceived for digital data storage, slow down to a crawl when processing more than a few million rows of data. Let's put processing issues aside for a moment. The issue which still stands is that similar quantities of data are impossible for a human to interpret and understand: we need to draw a correlation between tens of thousands of predictors and one observed phenomenon.

MACHINE LEARNING AND PREDICTION OF HUMAN BEHAVIOR

Does machine learning provide the answer? Many machine learning algorithms are able to not only discover a wide array of nonlinear relationships in big data but to

actively strive to reduce the complexity of the model they are generating in order to make it easier for a human being to understand these relationships. Machine learning models, however, have their limits too. As both statisticians and economists know, correlation does not imply causation. Causation, however, is what economists are most interested about: gauging policy effects or evaluating market fluctuations require causal inference. Research on machine learning has, for the most part, focused on prediction on historic data sets. Predictive models like this do not necessarily help causal inference, but can be useful in estimating the impact of a policy as they can predict what *would have happened* without intervening. Big data and the use of machine learning for their analysis have thus, right now, a position of importance in every economist's toolbox, in view of both the predictive and explanatory power of machine learning and the relatively low entry barrier: the advantages in data engineering have made it possible to design, train and test simple machine learning models on an ordinary laptop. More advanced models require a more powerful desktop computer with a graphics processing unit (GPU), but that's still a common sight in many households. Of course, enormous data centers filled with server racks are still a necessity for the kind of data processing that takes place at multinational companies such as Facebook or Google. The variety of devices that can be used to carry out research and development inside the field of machine learning allows for understanding just how much the field varies in scale and scope.

One of the applications of machine learning that is enjoying huge popularity today is using such algorithms as a tool for predicting human behavior when confronted with different situations. The reasons why it's useful to apply machine learning to this task are easy to infer: big samples of data are much easier to come by than big samples of people, as well as faster to study and free of consequences on real society. Moreover, machine learning allows researchers to experiment on a variety of samples from around the world since, thanks to the Open Data phenomenon, many public and private entities are releasing data on their operations that is constantly being updated and added upon. Researchers from the University of California, Berkeley (Kosinski, Stillwell, and Graepel 2013) used common statistical methods such as logistic regression to predict, with surprising accuracy (more than 85% in all cases), sexual orientation, ethnicity and position on the political spectrum of volunteers based on Facebook likes, demographic profiles and results of psychometric tests. Except for psychometric tests, these kind of data are made publicly available by default when somebody fills them into their Facebook profile.

Since the dawn of capitalism – or arguably even before, as demonstrated by massive economic surveys such as the eleventh-century Domesday Book – rulers, analysts

and policymakers alike have been interested in evaluating how economic and societal processes evolve over time and how people's actions are influenced by these processes. Mathematical models, as with many complex processes, are invaluably helpful to this aim. Over the years, increasingly complex economic models which involve human behavior have been developed by researchers in the fields of classical and neo-classical economics, game theory and behavioral economics. As with all mathematical models, assumptions are invariably required: one possibility frequently invoked by classical economics is schematizing human behavior by assuming that people always think rationally and are invariably self-interested, that is, their only aim is maximizing some kind of index or function, be it gross domestic product or subjective happiness. The incompleteness of this assumption is immediately apparent. The field of behavioral economics concerns itself, among other matters, with the evaluation of the extent of this incompleteness – that is, the evaluation of the amount of rationality that people exhibit – and the concoction of models able to provide a broader and more complete understanding of human behavior.

One might wonder how well does machine learning fare against similar, traditional economic models. Machine learning models outperform traditional ones and can *learn* some "irrational" human behavior. For example, when given choice between two lotteries, one with a fixed 50% chance of winning and the other with a chance of winning that is 50% on average, but varies between 0% and 100% between plays, most people prefer the first one because they dislike uncertainty over the winning probability. Machine learning models can learn similar patterns in human behavior and even re-discover traditional economic models devised to account for precisely this behavior. Machine learning models can also be used to compensate for the absence of a control group: Hal R. Varian, chief economist at Google, argues that predictions for "what would have happened without intervening" (e.g. changing a policy) made on the basis of an applied machine learning model can be even *better* than a control group because the model can take into account spurious variables (e.g. the effects of differing weather between two cities on sales) that a control group cannot.

CASE STUDY: LEARNING OF RISK- AND AMBIGUITY-AVERSE BEHAVIOR

Alexander Peysakhovich and Jeffrey Naecker, respectively senior research scientist at Facebook AI and economist at Google, published a comparison between traditional economic models in the domain of risk (where the outcome of an

event is uncertain, but the observer has full information about its probability distribution) and in the domain of ambiguity (where the outcome of an event is uncertain and the observer can only estimate it based on given data). The paper builds its dataset using Amazon MTurk, a crowdsourcing platform that allows researchers to hire human workers to perform certain tasks: for example, in this case, rating how likely they would be to play certain games of chance. MTurk is widely used in economic literature and there is substantial evidence that MTurk data sets are just as representative, if not more so (Paolacci and Chandler 2014), than traditional data sets. Participants were given instructions about the experiment, which required them to enter a numerical value that represented their "willingness to play" a certain lottery; a lottery was defined as an urn containing some red, green and blue balls, totaling 100. Each color had an associated monetary prize. Participants were then asked to complete a comprehension quiz and data from those who answered incorrectly were not considered in the final sample. This led to a final sample size of 315 people.

This sample was then split into a training set (70%) and a test set (30%) and model parameters were estimated using data from the training set while accuracy was gauged on the test set. Splitting data into a training and a test set in order to counteract in-sample over-fitting is standard practice in machine learning literature and will be mentioned several times throughout this paper.

The paper shows that in the domain of risk machine learning algorithms are able to rediscover the expected utility model with probability weighting (EUP) and perform just as well as the model itself in terms of average squared error, which was the metric used to gauge prediction accuracy throughout the paper. In the domain of ambiguity, however, regularized regression carried out through machine learning not only outperforms traditional economic models, but is also able to *learn* human ambiguity-aversion. The example proposed in the paper is that of two lotteries, with the same payoff, one having 50% odds of winning and the other having uniformly drawn odds in [0, 1] with an average of 50%: humans prefer the first and machine learning is able to predict this behavior. This also highlights how machine learning is not only useful as a prediction tool but can also be used to advance and improve current economic models, for example in this case by taking into account ambiguity aversion. Another point heavily driven home throughout the paper is the method used to evaluate machine learning algorithms for *out-of-sample* accuracy instead of *in-sample* fit: traditional statistical evaluation of models tends to err on the side of overfitting. However, training the machine learning model on only part of the available data allows for testing

its accuracy on the remainder, thus heavily penalizing models that can have high in-sample accuracy but are terrible predictors when used on different data from the ones it was trained on – like real world data.

LASSO AND RIDGE REGRESSION

A problem which requires predicting a real-valued outcome based on a set of explanatory variables (called features in machine learning) is known as a regression problem. Lasso and ridge regression are two subsets of regularized regression. The principle of regularized regression is similar to ordinary least squares (OLS) regression in that it assumes that it's possible to predict a vector of variables **y** as a linear function of *predictor variables* **X**, and express it in the form $\mathbf{y} = \mathbf{X}\beta$. In the one-dimensional case, that is if there is only one predictor variable and one predicted variable, OLS regression assumes there is a positive or negative linear relationship between the predictor and predicted variables which can be modeled as a line on the Cartesian plane. The objective of linear regression is to compute an estimate **b** of the vector β, which is known as the *weight vector* and corresponds to a single number in the one-dimensional case, that satisfies some criteria that define the "best fit" and usually involve the minimization of some *loss function*, or *cost function* which measures "by how much" the model is wrong. In the case of single-variable OLS regression, the criterium being used is the minimization of the *sum of squared residuals* (SSR) which is calculated as the sum of the squared difference between each known data point y_i and its estimate $\hat{y}_i = x_i b$, that is $SSR = \Sigma_i^n (y_i - bx_i)^2$.

Regularized regression attempts to minimize the complexity of the model by introducing a cost for non-zero weights. In a model with P predictor variables whose associated weights are β_p, regularized regression is carried out by imposing a penalty term of the form $\Sigma_p^P [(1-\alpha)|\beta_p| + \alpha|\beta_p|^2]$ and then trying to minimize the sum of SSR and this term. This method is called *elastic net regression*, of which lasso (*least absolute shrinkage and selection operator*) regression is a special case where $\alpha = 0$ and ridge regression is a special case where $\alpha = 1$. These methods attempt to reduce the number of non-zero regression coefficients (Varian 2014) in order to reduce complexity, leading to a model with less over-fit (and thus best out-of-sample performance) that is more easily interpreted by a human. In the paper being discussed, lasso and ridge regression are used to reduce over-fitting but are also a necessity, as one of the model used in the paper, where every individual is given their own risk-aversion coefficient, would try to

estimate about 55000 parameters which is way more than the number of data rows (about 2100).

An interesting remark can be made about the performance of regularized regression. Computer scientists usually evaluate performance using big O notation, which can be interpreted as an upper bound for the number of mathematical operations required to run the algorithm being analyzed until termination. The computational complexity of lasso and ridge regression, given a data set of size n and a model with m features the computational complexity is $O(m^3 + m^2n)$ (Efron et al. 2004) for both models: this means that regularized regression becomes slower and heavier very quickly when increasing the number of parameters of the model, but scales much better with data set size. This is definitely a plus for a model to be used in big data analysis.

GAME THEORY

The field of game theory has a wide array of applications, from the military to sales to agriculture. One of the most common models in game theory is the normal form game, or matrix game. In a two-player matrix game, the row and the column player are both given the same matrix that outlines the expected payoffs (which can be negative) for both players' choices. The two players then make their choices at the same time, or otherwise without knowing the other player's choice before making theirs.

A classical example is the prisoner's dilemma: two purported criminals are separated and asked to confess to a crime.

Row prisoner, column prisoner	**Cooperate**	**Defect**
Cooperate	1 year, 1 year	10 years, free
Defect	free, 10 years	5 years, 5 years

TABLE 1 Prisoner's dilemma payoff matrix.

If they both confess (defect) they both go to jail for a long sentence; if they both stay silent (cooperate) they both go to jail for a short sentence; if one confesses and the other doesn't, the defector goes free while the other person goes to jail and serves a very long sentence. This results in the matrix in table 1.

The prisoner's dilemma is useful to explain the concept of a Nash equilibrium. In this case, loosely speaking, the Nash equilibrium is the outcome resulting from

both prisoners looking at the payoff matrix and noticing that, whatever the choice of the other player, it's always advantageous to defect: going free is better than serving one year, and serving 5 years is better than serving 10. The Nash equilibrium is not necessarily the best overall choice: if both players cooperated, acting *irrationally*, they both would get a lower sentence than the one they'd get by acting *rationally* and ending up on the Nash equilibrium. In game theory terms, the Nash equilibrium is not necessarily *Pareto-optimal*. A strategy is Pareto-optimal or Pareto-dominant when there are no strategies that result in a better outcome for at least one player, with no other players ending up worse off than before: the strategy's payoffs are said to *Pareto-dominate* the payoffs in every other strategy. Pareto-dominant Nash equilibriums (PDNEs) do exist, but it's not a given that a Nash equilibrium will be Pareto-dominant.

While standard game-theoretical models rely on the assumption of common knowledge of rationality (everyone is fully rational and everyone knows it), in the last decades behavioral economists have developed alternative models that relax this assumption for the sake of psychological realism and predictive accuracy. Level-k play (Nagel 1995) is one of these models. In level-k play, a player is said to be level-0 if they choose their behavior ignoring the information they have on the other players and act randomly. A level-1 player assumes the other players are level-0 and chooses the play that maximizes expected payoff under this hypothesis. Reasoning inductively, for any value of k, level-k players are those who assume their opponents are level-$(k-1)$ and act accordingly.

A textbook example of level-k reasoning are the possible strategies for the Keynesian beauty contest, in which players are asked to choose a number and rewarded according to how close the number is to a certain fraction of the average of the other players' choices. If participants are asked to choose a number between 1 and 100, and rewarded if their number is close to half the average of the other players' choices, level-0 players will pick such a number at random from a uniform distribution; level-1 players will know that the average of a uniform distribution between 0 and 100 is 50 and will therefore pick 25 as their guess; level-2 players will assume everyone else is picking 25 and will therefore pick 13, and so on and so forth.

Level-1(α) play (Fudenberg and Liang 2019) is a variation of level-1 play which accounts for human risk aversion by choosing the action that corresponds to level-1 play on a game whose payoffs are a real-valued root of those of the original game: for every payoff u, this kind of play considers a payoff $f(u) = u^{\alpha}$, $\alpha < 1$. When the game has actions whose payoffs are close to those of the level-1 action but have lower variation, this kind of play achieves better accuracy than standard level-1.

A Keynesian beauty contest with the rules stated above can converge to the Nash equilibrium too. Players will note that numbers over 50 cannot be half of any other number under 100, so it does not make sense to choose a number greater than 50. If a player assumes that all others noticed this, they will conclude that it does not make sense to choose a number greater than 25. Reasoning iteratively, the Nash equilibrium will be reached when all players choose 1 as their number.

CASE STUDY: INITIAL PLAY ON MATRIX GAMES

If one player of any of these games could predict the other player's action with reasonably more certainty than random guessing, they could have a better chance at making the best choice they can to get the highest reward. Drew Fudenberg and Annie Liang, professors of Economics respectively at MIT and the University of Pennsylvania, used machine learning algorithms in their paper "Predicting and understanding initial play" to predict which of various standard econometric models such as the Nash equilibrium and level-1(α) play best fits any given game, then using that model to predict the row player's initial play. The games participants were asked to play were generated in different ways: for the initial data set, 86 games were selected from six game theory papers; afterwards, the authors noticed that level-1(α) play was a very good predictor of human behavior, achieving 89% accuracy, when ran on a different set of games with randomly-generated payoffs. Therefore, they thought that it would be more efficient to focus on games where level-1(α) was *not* a good predictor. To this end, the paper details how these games were generated: a rule was first trained to predict how often level-1(α) play would be preferred by human players, then random games were generated and those where more than 50% of players would follow this model were discarded until a data set of 200 games had been generated.

After generating these games, the authors used MTurk to assign 40 players to each game and analyzed the data, confirming their prediction that level-1(α) play would not do a good job of predicting initial play in such games.

On their first test run the paper finds that a bagged decision tree, a machine learning algorithm described in the following paragraph, is able to discover the same human risk-averse pattern covered in the previous case study by predicting that a play with similar expected payoff but lower variation is preferred to one with higher variation, even if the first one is the level-1 action . This was already known in game theory literature and had driven the development of the level-1(α) model, which ta-

kes into account ambiguity-aversion. The paper continues by using machine learning to generate games where level-1(α) is not accurate in order to uncover further hidden patterns. This results in games where choices that are part of a Pareto-dominant Nash equilibrium are more frequently chosen. Training a decision tree on a data set with randomly generated games leads to a tree that splits games into four classes, uses PDNE on two and level-1(α) on the other two and is able to predict initial play with 79% accuracy and 69% completeness, which is better than both models separately.

DECISION TREES

A problem which requires predicting a true/false outcome based on a set of features is known as a classification problem. The goal of *decision tree building* algorithms is to construct a tree that leads to a decision about how to classify the observation. Decision trees can also be applied to regression problems, in which the outcome is a continuous variable.

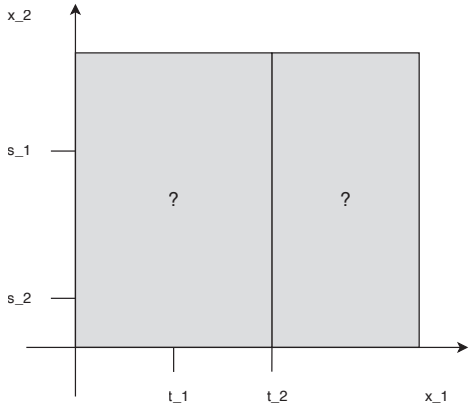
A graphical explanation of how decision trees are built is best applied to the case of binary trees (i.e. every non-leaf node of the tree has two children) and two features. Consider the data set as a scatter plot of points in the variables x_1 and x_2 and a continuous response y. The first decision partitions the space into two regions according to the mean value of y. The second decision partitions the two regions again, according to the same rule. One or more of these regions are finally partitioned again and then some stopping rule terminates the algorithm. This results in the partition plot in fig. 1c and the decision tree in fig. 1d. Decision trees are usually represented with the root at the top to present the choices that are made one after the other in the standard Western top-down reading order.

The restriction to binary trees is usually preferred because of computational efficiency considerations and this case can, of course, be generalized to an arbitrary number of predictors. A decision tree is easily interpreted by a human and the most and least important factors in deciding whether to predict "true" or "false" can be discerned by simply reading which predictors are or aren't used by the decision tree. Moreover, tree models are able to uncover more subtle relationships between data: a simple regression might find a linear or polynomial relationship between a variable and a feature, while a tree could show that only extremely high or low values of that feature alter the variable and that the feature has no influence when its value is in the middle of its range. Tree models, however, tend to over-fit and grow too much trying to cover every variation on the training data set. To counteract this

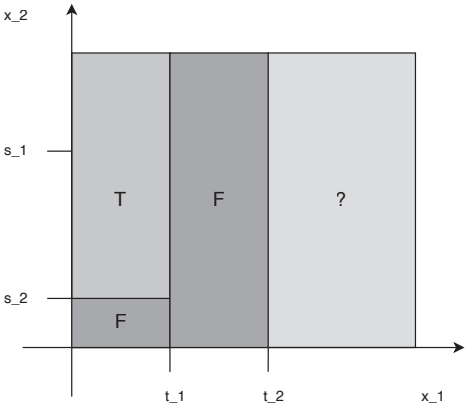
tendency, trees are usually *pruned* by imposing a cost for complexity, which in turn is usually defined as the number of terminal nodes (leaves).

Bagged decision tree building is one way to deal with over-fitting by repeatedly choosing with replacement a (not necessarily smaller) sample from a data set, a process known as bootstrapping. A decision tree is then built from each subsample and the resulting predictions are finally averaged.

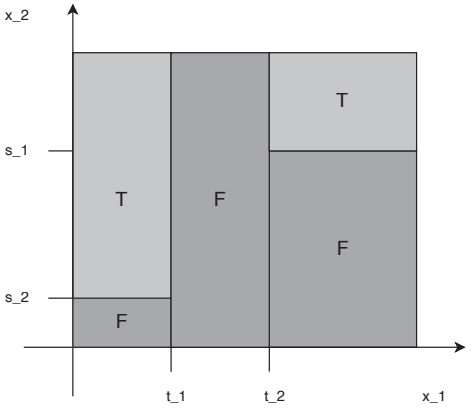
Random forest building is a more complex machine learning algorithm which uses multiple, unpruned trees that are restricted to use a randomly chosen subset of the full set of predictors. To classify an observation, a majority vote is used in the case of true/false classification and an average in the case of continuous va-



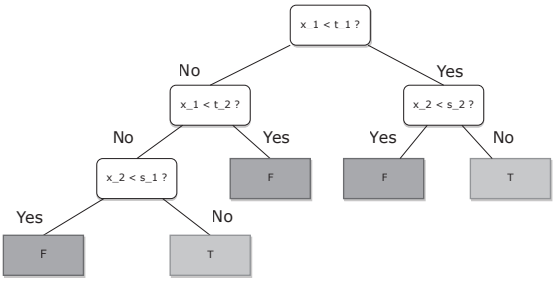
(a) First scatter plot partition.



(c) Third partition.



(b) Second partition.



(d) Resulting decision tree.

riables. Random forests, unlike trees, are harder to interpret as the number of trees is usually in the thousands. They can, however, still offer useful insight about which variables contribute the most to prediction accuracy.

Boosting is a method of dealing with over-fitting which involves repeatedly calling the classification function, with a data set which is gradually altered to give greater weight to misclassified data points, then choosing the classification via averaging or a majority vote. Gradient boosting accomplishes this by computing an estimator to be added to the loss function which is used to penalize misclassifications.

CASE STUDY: BAIL DECISIONS

Both computer scientists and economists are used to "abstracting away" the complexity of reality by devising and studying mathematical models that allow for isolating variables of interest. Models of human behavior, however, are not some kind of abstraction that crumbles when taken out of a lab; on the contrary, they have immediate applications to issues of interest to policymakers (Kleinberg, Ludwig, et al. 2015).

For example, every day, judges all over the world are requested to make what amounts to a very specific and very important prediction of human behavior: should they release the defendants that stand before them, will those defendants flee the country, commit other crimes, or just go on with their lives and wait for their trial, appearing in court when called upon? This isn't a matter of crime and punishment; in fact, the trial hasn't been carried out yet and police investigations might not even be over. The judge only knows the defendant's criminal history and some personal details, so any kind of correlation is not immediately obvious and, more importantly, there is no need to establish a causal relationship – since there either isn't one, or it's obvious – but only to predict somebody's behavior and act accordingly.

Researchers at Cornell University tackled precisely this case (Kleinberg, Lakkaraju, et al. 2017). The authors obtained a data set of 758027 arrests made in New York City between 2008 and 2013 for which judges had to decide whether to release or jail while waiting for the trial. They then used gradient-boosted decision trees to predict how high the risk is that the defendant will fail to appear in court (FTA) and analyze the results in order to devise a rule that can help judges reduce FTA rates making use of their algorithm. Decision trees' hyperparameters

such as tree depth are selected using five-fold cross-validation. k-fold cross validation is another common practice in machine learning: the data set is divided into k subsets, known as "folds", then parameters (or, in this case, hyperparameters) are estimated using $k - 1$ folds as training data and the remaining fold as test data. The process is repeated k times and results are then averaged.

The authors' decision trees are then trained only using data that is strictly related to the case, such as criminal records and previous FTAs, and the only unrelated feature is age at the time of arrest. Two glaring issues are evident from these two statements: first, we can only know FTAs from released defendants, but we can't know what a jailed defendant would have done if released; second, the devised models only use administrative data but the judge does not, which might lead to training – and therefore prediction – issues. The authors use gang tattoos as an example: if many people who have gang tattoos are young, and judges tend to always jail people with gang tattoos as they are considered high-risk, depending on how the loss function is constructed decision trees might erroneously decide to jail all young people or, conversely, to release them even though a judge would jail them, because they don't know about gang tattoos.

These two issues are related since, if we assume that judges take into account variables that are unobserved by the algorithm, we can't assume FTA rates of released defendants to be indicative of anything concerning jailed defendants with similar observed predictor variables. This is known as the *selective labels problem* and is tackled by exploiting its one-sidedness: it's easy to know what would have happened if a released defendant had been jailed. Since what the authors are interested in is *influencing a decision*, one way judges could be instructed to use their algorithm is to jail those that they would have released but are predicted to have a high risk of FTA: the data show that high-risk defendants are released much more often that they would if judges acted "rationally" and jailed everybody with a high predicted FTA risk. These people could in principle be low-risk, and the judges might realize this and release them, but relating observed FTA rates to predicted risk shows that defendants that are predicted to be high-risk do have a high observed FTA rate.

This kind of analysis does not take into account the cost implied by jailing defendants and only influences decisions on the basis of FTA risk. However, trying to draw a correlation between judge jailing rate and defendant characteristics does not lead to any kind of consistent finding (in statistical terms, drawing a histogram of p-values of F -test statistics does not show any unusual mass at low p-vaues). Therefore, if the average judge were to use the algorithm this way, it

would be possible to maintain the same FTA rate jailing only 48.2% as many people, or get FTA reductions that are 75.8% larger by jailing more people. By assuming unobservable variables have no effect, the algorithm could reduce FTAs by 24.7% without incrementing jailing rate, or reduce the detention rate by 41.9% without incrementing FTA rate.

These results seem to suggest that judges are making mistakes in their predictions, or decisions. Therefore, the authors then move on to try and explain what is causing these mispredictions. The first interesting finding is that judges with higher jailing rate are detaining more low-risk people, so a judge being "stricter" does not necessarily mean they are "better". Continuing along this line, the authors notice that judges struggle much more with high-risk cases, treating them as if they were low-risk and heavily weighing their decision based on the current offense which caused the defendant to be arrested. One possible explanation for both issues is that judges select on variables that are not observed by the algorithm, and righfully so: drawing from behavioral science literature, the author hypothesize that human judges overweigh interpersonal information, such as the degree of eye contact being made. Still, the authors do not manage to fully explain the source of judicial error: unobservable variables can only explain about a quarter of the difference between the judges' release decisions and the arguably better ones made by the algorithm.

ECONOMETRICS AND MACHINE LEARNING

Machine learning is not simply a faster way to analyze data or an evolution in spreadsheet technology. The fact that machine learning algorithms are able to re-discover traditional econometric models means that knowledge of these algorithms is now as valuable as the models themselves in an economist's set of tools. Beyond that, machine learning can help econometrists combine existing models or develop new ones in order to be able to both make better predictions and understand why those predictions work.

Moreover, machine learning is not only useful as a model but also as a way to identify causal relations, which play a fundamental role in policy applications, with more certainty thanks to samples that are more representative or novel ways to set up a study. This does not mean that the field of econometrics will be absorbed into that of machine learning but rather that any new research should take into account the possibilities offered by these tools which can help econometrists get their results not only faster but also more reliably.

REFERENCES

Athey, Susan (2018). "The impact of machine learning on economics". In: *The economics of artificial intelligence: An agenda*. University of Chicago Press.

Efron, Bradley et al. (2004). "Least angle regression". In: *The Annals of statistics* 32.2, pp. 407– 499.

Fudenberg, Drew and Annie Liang (2019). "Predicting and understanding initial play". In: Haugeland, John (1989). *Artificial intelligence: The very idea*. MIT press.

Hilbert, Martin and Priscila L ́opez (2011). "The world's technological capacity to store, communicate, and compute information". In: *science* 332.6025, pp. 60–65.

Karpathy, Andrej (2014). *What I learned from competing against a ConvNet on ImageNet*. url: https://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/.

Kleinberg, Jon, Himabindu Lakkaraju, et al. (Aug. 2017). "Human Decisions and Machine Predictions". In: *The Quarterly Journal of Economics* 133.1, pp. 237–293. issn: 0033-5533. doi: 10.1093/qje/qjx032. eprint: https://academic.oup.com/qje/article-pdf/133/1/237/ 30636517/qjx032.pdf. url: https://doi.org/10.1093/qje/qjx032.

Kleinberg, Jon, Jens Ludwig, et al. (2015). "Prediction policy problems". In: *American Economic Review* 105.5, pp. 491–95.

Kosinski, Michal, David Stillwell, and Thore Graepel (2013). "Private traits and attributes are predictable from digital records of human behavior". In: *Proceedings of the National Academy of Sciences* 110.15, pp. 5802–5805. issn: 0027-8424. doi: 10.1073/pnas.1218772110. eprint: https://www.pnas.org/content/110/15/5802.full.pdf. url: https://www.pnas.org/ content/110/15/5802.

Mullainathan, Sendhil and Jann Spiess (2017). "Machine learning: an applied econometric approach". In: *Journal of Economic Perspectives* 31.2, pp. 87–106.

Nagel, Rosemarie (1995). "Unraveling in guessing games: An experimental study". In: *The American Economic Review* 85.5, pp. 1313–1326.

Paolacci, Gabriele and Jesse Chandler (2014). "Inside the Turk: Understanding Mechanical Turk as a participant pool". In: *Current Directions in Psychological Science* 23.3, pp. 184–188.

Peysakhovich, Alexander and Jeffrey Naecker (2017). "Using methods from machine learning to evaluate behavioral models of choice under risk and ambiguity". In: *Journal of Economic Behavior & Organization* 133, pp. 373–384.

Rudin, Peter (2017). *Thoughts on Human Learning vs. Machine Learning*. url: https://singularity2030. ch/thoughts-on-human-learning-vs-machine-learning/.

Thrun, Sebastian and Chris Anderson (2017). *What AI is – and isn't*. url: https://www.ted. com/talks/sebastian_thrun_and_chris_anderson_the_new_generation_of_computers_ is_programming_itself.

Varian, Hal R. (2014). "Big data: New tricks for econometrics". In: *Journal of Economic Perspectives* 28.2, pp. 3–28.
Wiener, Janet and Nathan Bronson (2014). "Facebook's top open data problems". In: url: https://research.facebook.com/blog/facebook-s-top-open-data-problems/.

Finito di stampare nel mese di novembre 2020
per i tipi di Bononia University Press

Il nostro mondo racchiude un sottile ossimoro: se da un lato appare sempre più semplice e soddisfa istantaneamente i nostri bisogni elementari grazie a tecnologie sempre più avanzate, dall'altro diventa sempre più complesso e richiede un articolato sforzo concettuale per poterlo comprendere a fondo. Poco ci è richiesto per subire il divenire del mondo; molto per poterlo capire.

Come ci prepariamo quindi ad affrontare il mondo? La scuola ci propone percorsi specializzati, ma proprio per questo incompleti, perché trascurano concetti elementari che ogni cittadino dovrebbe conoscere. Le astrazioni stenografiche sono proprio questi concetti chiave, utili a colmare questo gap di conoscenza e indispensabili nella formazione culturale di chiunque voglia essere parte attiva della società contemporanea.

Il volume presenta cinque nuove astrazioni stenografiche: cinque appassionanti racconti di scienza, scritti da alcuni fra i migliori membri del Collegio Superiore dell'Università di Bologna, che illustrano alcune idee fondamentali per comprendere il mondo intorno a noi, aiutandoci a diventare cittadini consapevoli, più colti e più preparati.

Astrazioni Stenografiche è una collana di divulgazione scientifica creata sotto la supervisione della Professoressa Beatrice Fraboni, direttrice del Collegio Superiore, e curata dal Professor Matteo Cerri.

ISBN 978-88-6923-676-1

€ 15,00